



About this release

Contents

Front Matter
Contributors
Preface

CHAPTER 1 - The Systemic Response to Injury
CHAPTER 2 - Fluid and Electrolyte Management
CHAPTER 3 - Hemostasis, Surgical Bleeding, and Transfusion
CHAPTER 4 - Shock
CHAPTER 5 - Surgical Infections
CHAPTER 6 - Trauma
CHAPTER 7 - Burns
CHAPTER 8 - Wound Care and Wound Healing
CHAPTER 9 - Oncology
CHAPTER 10 - Transplantation
CHAPTER 11 - Surgical Complications
CHAPTER 12 - Physiologic Monitoring of the Surgical Patient

CHAPTER 13 - Skin and Subcutaneous Tissue
CHAPTER 14 - Breast
CHAPTER 15 - Tumors of the Head and Neck
CHAPTER 16 - Chest Wall, Pleura, Lung, and Mediastinum
CHAPTER 17 - Congenital Heart Disease
CHAPTER 18 - Acquired Heart Disease
CHAPTER 19 - Thoracic Aneurysms and Aortic Dissection
CHAPTER 20 - Arterial Disease
CHAPTER 21 - Venous and Lymphatic Disease
CHAPTER 22 - Manifestations of Gastrointestinal Disease
CHAPTER 23 - Esophagus and Diaphragmatic Hernia
CHAPTER 24 - Stomach
CHAPTER 25 - Small Intestine
CHAPTER 26 - Colon, Rectum, and Anus
CHAPTER 27 - The Appendix
CHAPTER 28 - Liver
CHAPTER 29 - Gallbladder and Extrahepatic Biliary System
CHAPTER 30 - Pancreas
CHAPTER 31 - Spleen
CHAPTER 32 - Intraabdominal Infections
CHAPTER 33 - Abdominal Wall, Omentum, Mesentery, and Retroperitoneum
CHAPTER 34 - Abdominal Wall Hernias
CHAPTER 35 - Pituitary and Adrenal
CHAPTER 36 - Thyroid and Parathyroid
CHAPTER 37 - Pediatric Surgery
CHAPTER 38 - Urology
CHAPTER 39 - Gynecology
CHAPTER 40 - Neurosurgery
CHAPTER 41 - Orthopaedics
CHAPTER 42 - Surgery of the Hand
CHAPTER 43 - Plastic and Reconstructive Surgery
CHAPTER 44 - Minimally Invasive Surgery

Version 1.1

Compiled to iSilo format from Principles of Surgery CD-ROM by **snickers**

brought to you by **PalmWarez**

[Go back to book](#)

PRINCIPLES OF SURGERY
SEVENTH EDITION
Editor-in-Chief

Seymour I. Schwartz, M.D.

Distinguished Alumni Professor and Chair
Department of Surgery
University of Rochester Medical Center
Rochester, New York

Associate Editors

G. Tom Shires, M.D.

Professor of Surgery
Director, Trauma Institute
University of Nevada School of Medicine
Las Vegas, Nevada

Frank C. Spencer, M.D.

Professor and Chairman
Department of Surgery
New York University Medical Center
New York, New York

John M. Daly, M.D

Lewis Atterbury Stimson Professor and Chairman
Department of Surgery
Cornell University Medical College
New York, New York

Josef E. Fischer, M.D.

Christian R. Holmes Professor and Chairman
Department of Surgery
University of Cincinnati College of Medicine
Cincinnati, Ohio

Aubrey C. Galloway, M.D.

Professor of Surgery
Director of Surgical Research
New York University Medical Center
New York, New York

NOTICE

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The editors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possiblity of human error or changes in medical sciences, neither the editors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from use of such information. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this book is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

Copyright Page: Print Version
Principles of Surgery, 7/e

Copyright © 1999, 1994, 1989, 1984, 1979, 1974, 1969

The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.
1234567890 DOWDOW 998
This book was set in Times Roman by York Graphic Services, Inc. The editors were Marty Wonsiewicz and Peter McCurdy; the production supervisor was Rich Ruzycka; the cover designer was Ed Schultheis. Barb Littlewood prepared the index. R.R. Donnelley & Sons, Co. was printer and binder. This book is printed on acid-free paper. The illustration in the upper left of the book's cover is derived from the painting Limb Transplantation Miracle by Saints Cosmos and Damian. Used by permission of the Wuerttembergische Landesmuseum, Stuttgart, Germany. The image on the lower left of the cover is used courtesy of the Yale Historical Library. The figure on the lower right is from Dis Is das Buch der Chirurgia by Hieronymous Brunschwig. Courtesy of the National Library of Medicine, Bethesda, Maryland. The illustration on the back cover and on the part opening pages is from An Explanation of the Fashion and Use of Three and Fifty Instruments of Chirurgery—Gathered out of Ambrosius Pareus. Library of Congress Cataloging-in-Publication Data Principles of surgery/editors, Seymour I. Schwartz . . . [et al.]. — 7th ed. p. cm. Includes bibliographical references and index. ISBN 0-07-054256-2 1. Surgery. I. Schwartz, Seymour I. [DNLM: 1. Surgery. 2. Surgical Procedures, Operative. WO 100P957 1999] RD31.P88 1999 617—dc21 DNLM/DLC 98-28061 for Library of Congress CIP

Contributors

James T. Adams, MD (Chapter 33)

Professor of Surgery
University of Rochester Medical Center

R. Peter Altman, MD (Chapter 37)

Rudolph N. Schullinger Professor of Surgery (Pediatric Surgery)
Columbia Presbyterian Medical Center

Kathryn D. Anderson, MD (Chapter 37)

Professor of Surgery (Pediatric Surgery)
University of Southern California, Los Angeles

Richard V. Anderson, MD (Chapter 18)

Department of Surgery
New York University Medical Center

Michael Artman, MD (Chapter 17)

Professor of Pediatrics, Physiology, and Neuroscience
New York University Medical Center

Stanley W. Ashley, MD (Chapter 24)

Associate Professor of Surgery
Harvard Medical School

Annabel Barber, MD (Chapters 2 and 4)

Associate Professor of Surgery
University of Nevada School of Medicine

Philip S. Barie, MD (Chapter 32)

Associate Professor of Surgery
Cornell University Medical College

Monica Bertagnolli, MD (Chapter 9)

Assistant Professor of Surgery
Cornell University Medical Center

Elisa H. Birnbaum, MD (Chapter 26)

Assistant Professor of Surgery
Washington University School of Medicine

Kirby I. Bland, MD (Chapter 14)

J. Murray Beardsley Professor and Chairman
Brown University School of Medicine

Michael F. Boland, MD (Chapter 40)

Department of Neurosurgery
Missouri Baptist Medical Center

Jon M. Burch, MD (Chapter 6)

Professor of Surgery
University of Colorado Health Sciences Center

Steve E. Calvano, PhD (Chapter 1)

Associate Professor of Surgery
University of Medicine and Dentistry of New Jersey—Robert Wood Johnson Medical School

William T. Chance, MD (Chapter 22)

Associate Professor of Surgery
University of Cincinnati Medical Center

Joseph M. Civetta, MD (Chapter 12)

Professor and Head, Department of Surgery

University of Connecticut Health Sciences Center

Orlo H. Clark, MD (Chapter 36)

Professor and Vice Chair of Surgery

University of California, San Francisco

I. Kelman Cohen, MD (Chapter 8)

Professor of Surgery (Division of Plastic and Reconstructive Surgery)

Medical College of Virginia

Virginia Commonwealth University

John J. Coleman III, MD (Chapter 15)

Professor of Surgery (Plastic Surgery)

Indiana University School of Medicine

Stephen B. Colvin, MD (Chapters 17 and 19)

Chief of Cardiothoracic Surgery

New York University Medical Center

Edward M. Copeland III, MD (Chapter 14)

The Edward R. Woodward Professor and Chairman of Surgery

University of Florida College of Medicine

William T. Couldwell, MD, PhD (Chapter 35)

Professor and Chairman of Neurological Surgery

New York Medical College

Mary C. Crossland, BSN, RN (Chapter 8)

Director, Wound Healing Center

Columbia Retreat Hospital

Anthony M. D'Alessandro, MD (Chapter 10)

Associate Professor of Surgery

University of Wisconsin Medical School

John M. Daly, MD (Chapters 9, 24, and 33)

Lewis Atterbury Stimson Professor and Chairman of Surgery

Cornell University Medical College

Jerome J. DeCosse, MD (Chapter 9)

Lewis Thomas University Professor

Cornell University Medical College

Mark H. Deierhoi, MD (Chapter 10)

Professor of Surgery

University of Alabama at Birmingham

Tom R. DeMeester, MD (Chapter 23)

Professor and Chairman of Surgery

University of Southern California School of Medicine, Los Angeles

Robert F. Diegelmann, PhD (Chapter 8)

Professor of Surgery (Plastic Surgery)

Medical College of Virginia

Virginia Commonwealth University

B. Mark Evers, MD (Chapter 25)

Chela and Jimmy Storm Distinguished Professor of Surgery

University of Texas Medical Branch

Denis Evoy, MD (Chapter 24)

Department of Surgery

Cork University Hospital, Ireland

Gary A. Fantini, MD (Chapter 33)

Associate Professor of Surgery
Cornell University Medical College

David R. Farley, MD (Chapter 36)

Assistant Professor of Surgery
Mayo Medical School

Elliott Fegelman, MD (Chapter 11)

Assistant Professor of Surgery
University of Cincinnati Medical Center

Josef E. Fischer, MD (Chapters 11, 22, and 33)

Christian R. Holmes Professor and Chairman of Surgery
University of Cincinnati Medical Center

James W. Fleshman, MD (Chapter 26)

Associate Professor of Surgery
Washington University School of Medicine

Reginald J. Franciose, MD (Chapter 6)

Assistant Professor of Surgery
University of Colorado Health Sciences Center

Robert D. Fry, MD (Chapter 26)

Professor of Colon and Rectal Surgery
Thomas Jefferson University

Aubrey C. Galloway, MD (Chapters 17, 18 and 19)

Professor of Surgery
New York University Medical Center

Robert J. Ginsberg, MD (Chapter 16)

Professor of Surgery
Cornell University Medical
Memorial Sloan-Kettering Cancer Center

Martin F. Graham, MD (Chapter 8)

Professor of Pediatrics, Biochemistry, and Molecular Biophysics
Medical College of Virginia
Virginia Commonwealth University

Richard M. Green, MD (Chapters 20 and 21)

Associate Professor of Surgery
University of Rochester Medical Center

Eugene A. Grossi, MD (Chapter 18)

Associate Professor of Surgery
New York University Medical Center

Philip C. Guzzetta, MD (Chapter 37)

Professor and Chairman of Pediatric Surgery
University of Texas Southwestern Medical Center

David M. Heimbach, MD (Chapter 7)

Professor of Surgery
University of Washington School of Medicine

Julian T. Hoff, MD (Chapter 40)

Professor of Neurologic Surgery
University of Michigan

Richard J. Howard, MD (Chapter 5)

Robert H. and Kathleen M. Axline Professor of Surgery
University of Florida College of Medicine

John G. Hunter, MD (Chapter 44)

Professor of Surgery
Emory University

William W. Hurd, MD (Chapter 39)

Associate Professor of Obstetrics and Gynecology
Indiana University School of Medicine

Jay Johannigman, MD (Chapter 11)

Assistant Professor of Surgery
University of Cincinnati Medical Center

M. J. Jurkiewicz, MD (Chapter 43)

Professor of Surgery (Plastic Surgery), Emeritus
Emory University School of Medicine

Munci Kalayoglu, MD (Chapter 10)

Professor of Surgery and Pediatrics
University of Wisconsin Medical School

Allan D. Kirk, MD, PhD (Chapter 10)

Senior Scientist
Naval Medical Research Institute

Orlando C. Kirton, MD (Chapter 12)

Associate Professor of Surgery
University of Miami School of Medicine

Stuart J. Knechtle, MD (Chapter 10)

Associate Professor of Surgery
University of Wisconsin Medical School

Ira J. Kodner, MD (Chapter 26)

Professor of Surgery
Washington University School of Medicine

Rosemary A. Kozar, MD (Chapter 27)

Assistant Professor of Surgery
Allegheny University of the Health Sciences

Edward Lin, DO (Chapter 1)

Department of Surgery
New York Hospital Medical Center

Stephen R. Lowry, MD (Chapters 1 and 2)

Professor and Chairman of Surgery
University of Medicine and Dentistry of New Jersey— Robert Wood Johnson Medical School

Frederick Luchette, MD (Chapter 22)

Associate Professor of Surgery
University of Cincinnati Medical Center

Stephen J. Mathes, MD (Chapter 13)

Professor of Surgery (Plastic Surgery)
University of California, San Francisco

John D. McConnell, MD (Chapter 38)

Professor and Chairman of Urology
University of Texas Southwestern Medical Center

Jeffrey S. Miller, MD (Chapter 19)

Department of Surgery
New York University Medical Center

Ernest E. Moore, MD (Chapter 6)

Professor of Surgery

University of Colorado Health Sciences Center

Donald L. Morton, MD (Chapter 9)

Medical Director and Surgeon-in-Chief

John Wayne Cancer Institute at Saint John's Health Center, Santa Monica, CA

Professor Emeritus

University of California, Los Angeles School of Medicine

Kurt D. Newman, MD (Chapter 37)

Professor of Surgery and Pediatrics

George Washington University School of Medicine

Jeffrey A. Norton, MD (Chapter 35)

Professor and Vice Chairman of Surgery

University of California, San Francisco School of Medicine

Michael S. Nussbaum, MD (Chapter 22)

Associate Professor of Surgery

University of Cincinnati Medical Center

Jon S. Odorico, MD (Chapter 10)

Assistant Professor of Surgery

University of Wisconsin Medical School

Kenneth Ouriel, MD (Chapters 20 and 21)

Associate Professor of Surgery and Radiology

University of Rochester Medical Center

Margaret S. Pearle, MD (Chapter 38)

Assistant Professor of Urology

University of Texas Southwestern Medical Center

Clayton A. Peimer, MD (Chapter 42)

Professor of Orthopaedic Surgery

School of Medicine and Biomedical Sciences

State University of New York, Buffalo

Jeffrey H. Peters, MD (Chapter 23)

Associate Professor of Surgery

University of California, Los Angeles School of Medicine

Paul C. Peters, MD (Chapter 38)

Professor Emeritus of Urology

University of Texas Southwestern Medical Center

Thomas E. Read, MD (Chapter 26)

Assistant Professor of Surgery

Washington University School of Medicine

Howard A. Reber, MD (Chapter 30)

Professor of Surgery

University of California, Los Angeles School of Medicine

Bruce A. Reitz, MD (Chapter 10)

The Norman E. Shumway Professor and Chairman of

Cardiothoracic Surgery

Stanford University School of Medicine

Robert E. Rogers, MD (Chapter 39)

Emeritus Professor of Obstetrics and Gynecology

Indiana University School of Medicine

Randy N. Rosier, MD (Chapter 41)

Professor of Orthopaedics

University of Rochester Medical Center

Joel J. Roslyn, MD (Chapter 27)

Alma Dea Morani Professor and Chairman of Surgery

Allegheny University of the Health Sciences

Valerie W. Rusch, MD (Chapter 16)

Professor of Surgery

Cornell University Medical College

Memorial Sloan-Kettering Cancer Center

Gregory P. Sadler, MD (Chapter 36)

Department of Surgery

Oxford University, England

Jay J. Schnitzer, MD, PhD (Chapter 37)

Assistant Professor of Surgery

Harvard Medical School

Seymour I. Schwartz, MD (Chapters 3, 28, 29 and 31)

Distinguished Alumni Professor and Chair of Surgery

University of Rochester Medical Center

G. Tom Shires, MD (Chapters 2 and 4)

Professor of Surgery

University of Nevada School of Medicine

G. Thomas Shires III, MD (Chapters 2 and 4)

Associate Professor of Surgery

University of Texas Southwestern Medical Center

Marie F. Simard, MD (Chapter 35)

Research Assistant Professor of Neuroendocrinology

New York Medical College

Hans W. Sollinger, MD (Chapter 10)

Folker O. Belzer Professor of Surgery

University of Wisconsin Medical School

Joseph Solomkin, MD (Chapter 32)

Professor of Surgery, Pharmacology and Cell Biophysics

University of Cincinnati Medical Center

Frank C. Spencer, MD (Chapter 19)

Professor and Chairman of Surgery

New York University Medical Center

Mark R. Sultan, MD (Chapter 15)

Assistant Professor of Surgery

Columbia-Presbyterian Medical Center

Gregory P. Sutton, MD (Chapter 39)

Mary Fendrich Hulman Professor of Gynecologic Oncology

Indiana University School of Medicine

James C. Thompson, MD (Chapter 25)

Ashbel Smith Professor of Surgery

University of Texas Medical Branch

Courtney M. Townsend, Jr., MD (Chapter 25)

John Woods Harris Distinguished Professor of Surgery

University of Texas Medical Branch

Jon A. van Heerden, MD (Chapter 36)

Fred C. Andersen Professor of Surgery

Mayo Medical School

Albert J. Varon, MD (Chapter 12)

Professor of Anesthesiology and Surgery

University of Miami School of Medicine

Michael P. Vezeridis, MD (Chapter 14)

Professor of Surgery

Brown University School of Medicine

George M. Wantz, MD (Chapter 34)

Clinical Professor of Surgery

Cornell University Medical College

Glenn D. Warden, MD (Chapter 7)

Professor of Surgery

University of Cincinnati Medical Center

Martin H. Weiss, MD (Chapter 35)

Professor and Chairman of Neurosurgery

University of Southern California School of Medicine, Los Angeles

Michael A. West, MD, PhD (Chapter 32)

Associate Professor of Surgery

University of Minnesota

Dietmar W. Wittmann, MD, PhD (Chapter 32)

Professor of Surgery

Medical College of Wisconsin

Robert J. Wood, MD (Chapter 43)

Associate Professor of Surgery (Plastic Surgery)

Emory University School of Medicine

Isaac L. Wornum, III, MD (Chapter 8)

Associate Professor of Surgery (Plastic Surgery)

Medical College of Virginia

Virginia Commonwealth University

Dorne R. Yager, PhD (Chapter 8)

Assistant Professor of Surgery (Plastic and Reconstructive Surgery), Microbiology, and Immunology

Medical College of Virginia

Virginia Commonwealth University

David M. Young, MD (Chapter 13)

Assistant Professor-in-Residence

University of California at San Francisco School of Medicine

David D. Yuh, MD (Chapter 10)

Department of Cardiothoracic Surgery

Stanford University School of Medicine

Preface

The Seventh Edition of ***Principles of Surgery*** completes our participation in the surgical education of an entire generation of medical students and surgical residents throughout the world. We also were pleased to have played a role in the continuing education of practicing surgeons.

Many have regarded teachers as the noblest of people; others attach that designation to healers. As one who has been privileged to serve in both roles, namely, to have provided a vehicle for educating those who will perpetuate the healing profession in the realm of Surgery, and, at the same time, to have had the opportunity to participate in relieving patients of their disease, I consider myself twice blessed.

As a surgeon I have been satisfied by successes in patient care. As a teacher, I have been literally rewarded by expressions of appreciation by our readers who have indicated that we have enhanced their education.

Thirty-two years have passed since we accepted the publisher's and our own self-generated challenge to develop a "new and modern" textbook of Surgery. The favorable reception and the text's longevity suggest that we have succeeded. As the landmark of the Seventh Edition is completed, the frustrations and toils are erased and what remains is an immeasurable sense of gratification.

Seymour I. Schwartz, M.D.
June, 1998

ACKNOWLEDGMENT

We are particularly appreciative of the efforts of Andrea Weinstein, who had an integral role in each of the processes throughout the development of this edition. John Guardiano also contributed significantly to the technical editing of the manuscript.

Preface to the First Edition

The raison d'être for a new textbook in a discipline which has been served by standard works for many years was the Editorial Board's initial conviction that a distinct need for a modern approach in the dissemination of surgical knowledge existed. As incoming chapters were reviewed, both the need and satisfaction became increasingly apparent and, at the completion, we felt a sense of excitement at having the opportunity to contribute to the education of modern and future students concerned with the care of surgical patients.

The recent explosion of factual knowledge has emphasized the need for a presentation which would provide the student an opportunity to assimilate pertinent facts in a logical fashion. This would then permit correlation, synthesis of concepts, and eventual extrapolation to specific situations. The physiologic bases for diseases are therefore emphasized and the manifestations and diagnostic studies are considered as a reflection of pathophysiology. Therapy then becomes logical in this schema and the necessity to regurgitate facts is minimized. In appreciation of the impact which Harrison's PRINCIPLES OF INTERNAL MEDICINE has had, the clinical manifestations of the disease processes are considered in detail for each area. Since the operative procedure represents the one element in the therapeutic armamentarium unique to the surgeon, the indications, important technical considerations, and complications receive appropriate emphasis. While we appreciate that a textbook cannot hope to incorporate an atlas of surgical procedures, we have provided the student a single book which will satisfy the sequential demands in the care and considerations of surgical patients.

The ultimate goal of the Editorial Board has been to collate a book which is deserving of the adjective "modern." We have therefore selected as authors dynamic and active contributors to their particular fields. The au courant concept is hopefully apparent throughout the entire work and is exemplified by appropriate emphasis on diseases of modern surgical interest, such as trauma, transplantation, and the recently appreciated importance of rehabilitation. Cardiovascular surgery is presented in keeping with the exponential strides recently achieved.

There are two major subdivisions to the text. In the first twelve chapters, subjects that transcend several organ systems are presented. The second portion of the book represents a consideration of specific organ systems and surgical specialties.

Throughout the text, the authors have addressed themselves to a sophisticated audience, regarding the medical student as a graduate student, incorporating material generally sought after by the surgeon in training and presenting information appropriate for the continuing education of the practicing surgeon. The need for a text such as we have envisioned is great and the goal admittedly high. It is our hope that this effort fulfills the expressed demands.

Seymour I. Schwartz, M.D.

CHAPTER 1 - The Systemic Response to Injury

Edward Lin
Stephen F. Lowry
Steve E. Calvano

INTRODUCTION

The host response to injury—surgical, traumatic, or infectious—is characterized by various endocrine, metabolic, and immunologic alterations. If the inciting injury is minor and of limited duration, wound healing and restoration of metabolic and immune homeostasis readily occurs. More significant insults lead to further deterioration of the host regulatory processes, which, without appropriate intervention, often precludes full restoration of cellular and organ function or results in death. The spectrum of cellular metabolic and immunologic dysfunction resulting from injury suggests a complex mechanism for identifying and initially quantifying the injurious event. This initial response is inherently inflammatory, inciting the activation of cellular processes designed to restore or maintain function in tissues while also promoting the eradication or repair of dysfunctional cells. These dynamic processes imply the existence of antiinflammatory or counterregulatory processes that promote the restoration of homeostasis (Fig. 1-1). A discussion of the response to injury must account for the collective dynamics of neuroendocrine, immunologic, and metabolic alterations characteristic of the injured patient. This chapter discusses concepts related to macroendocrine and microendocrine contributions to the basic metabolic and immunological consequences of injury and also the current concepts of metabolism and nutritional support for the surgical patient as a practical and readily applicable adjunct for the provision of essential substrates. The dynamics of hormonal and immunologic influences on the metabolic and substrate requirements of the injured patient are emphasized.

ENDOCRINE RESPONSE TO INJURY

Overview of Hormone-Mediated Response

The classic response to injury comprises multiple axes. These hormone response pathways are activated by (1) mediators released by the injured tissue, (2) neural and nociceptive input originating from the site of injury, or (3) baroreceptor stimulation from intravascular volume depletion. The hormones released in response to these activating stimuli may be divided into those primarily under hypothalamopituitary control and those primarily under autonomic nervous system control (Table 1-1). The interaction between these origins form the basis of the hypothalamic-pituitary axis, which represents a series of signaling and feedback loops regulating the endocrine response to injury (Fig. 1-2).

Hormone-Mediated Receptor Activity

Hormones may be classified according to chemical structure and to the mechanisms by which they elicit biologic effects (Tables 1-2 and 1-3). Central to the hormone-mediated response at the cellular level is the hormone (ligand)–receptor interaction and subsequent postreceptor activity. Most macroendocrine hormone receptors can be categorized on the basis of their mechanisms of signal transduction into three major types: (1) receptor kinases with ligands such as insulin and insulinlike growth factors; (2) guanine nucleotide–binding or G protein–coupled receptors that are activated by peptide hormones, neurotransmitters, and prostaglandins; and (3) ligand-gated ion channels that permit ion transport upon ligand-receptor binding (Fig. 1-3).

Hormone-Mediated Intracellular Pathways

One of the most common intracellular second messengers by which hormones exert their effects is the modulation of cyclic adenosine monophosphate (cAMP). Receptor occupation by stimulatory hormones induces a cell membrane alteration that activates the enzyme adenylate cyclase. Adenylate cyclase catalyzes the conversion of adenosine triphosphate (ATP) to cAMP, which activates various intracellular protein kinases. Substances that decrease cAMP generally exert an influence opposite to those observed for substances that increase cAMP. Increases in intracellular cAMP are associated with functional lymphocyte responses that generally are immunosuppressive. In T lymphocytes, agents that increase cAMP levels diminish proliferation, lymphokine production, and cytotoxic functions. Plasma cell production of immunoglobulins is markedly attenuated. Neutrophils manifest decreased chemotaxis and reduced production of superoxides, H₂O₂, and lysosomal enzymes. Basophils or mast cells demonstrate a decreased release of histamine. Many prolonged hormone-mediated responses to injury increase intracellular cAMP levels through a direct action on membrane receptors or by increasing the sensitivity of leukocytes to substances that directly increase cAMP. Hormonal actions are further mediated by intracellular receptors. These intracellular receptors have binding affinities for the hormone and for the targeted gene sequence on the DNA. These intracellular receptors may be located within the cytosol or may already be localized in the nucleus, bound to the DNA. The classic example of a cytosolic hormonal receptor is glucocorticoid receptor. Intracellular glucocorticoid receptors are maintained in an active state by linking to the stress-induced protein, heat-shock protein (HSP). When the hormone ligand binds to the receptor, the dissociation of HSP from the receptor activates the receptor-ligand complex and is transported to the nucleus (Fig. 1-4).

Hormones Under Anterior Pituitary Regulation

Corticotropin-Releasing Hormone

Pain, fear, anxiety, or emotional arousal generate neural signals to the paraventricular nucleus of the hypothalamus, stimulating the synthesis of corticotropin-releasing hormone (CRH), which is then delivered by way of the hypothalamic-hypophyseal portal circulation to the anterior pituitary (Fig. 1-5). Proinflammatory cytokines and arginine vasopressin (AVP) also can induce CRH synthesis and release. In the anterior pituitary, CRH serves as the major stimulant of adrenocorticotrophic hormone (ACTH, adrenocorticotropin) production and release (Fig. 1-6). This is accomplished by CRH-mediated activation of adenylate cyclase in the ACTH-producing corticotrophs, which increases intracellular cAMP levels and activates the pathway leading to increased ACTH production. CRH, formerly referred to as corticotropin-releasing factor, has induced, in laboratory animals, hyperdynamic cardiovascular and catecholamine release characteristic of the sympathetic stress response. CRH secretion can be activated by angiotensin II, neuropeptide Y (NPY), serotonin, acetylcholine, interleukin-1, and interleukin-6. The release of CRH can be inhibited by g- aminobutyric acid (GABA), substance P, atrial natriuretic peptide (ANP), endogenous opioids, and l-arginine. Receptors for CRH in primates have been described not only in the pituitary gland and the central nervous system but also in the renal medulla, in the marginal zone and red pulp of the spleen, and in the sympathetic ganglia. Circulating glucocorticoids serve as potent negative feedback signals to the hypothalamus and have demonstrated in animal models an ability to reduce CRH mRNA transcription. Conversely, adrenalectomized animals demonstrate elevated CRH mRNA transcriptional activities that are reversed with exogenous administration of dexamethasone or prednisolone. CRH-binding proteins (CRH-BP) synthesized by the liver also serve as regulators of CRH activity. These collectively demonstrate endogenous pathways that may potentially regulate or preclude excessive CRH-mediated responses to injury. Injured tissues also produce CRH that may contribute locally to the inflammatory response. Experimental studies suggest a role for CRH in preventing vascular leakage in injured or inflamed tissues, although the implications have not been identified.

Adrenocorticotrophic Hormone

ACTH is synthesized, stored, and released by the anterior pituitary upon CRH stimulation. ACTH is a 39–amino acid peptide that is synthesized as a larger precursor complex known as proopiomelanocortin (POMC). POMC is cleaved within the cytosol to the components a-melanocyte stimulating hormone (a-MSH), b-lipotropin, the endogenous opioid b-endorphin, and ACTH. In the nonstressed healthy human being, ACTH release is regulated by circadian signals; the greatest elevation of ACTH occurs late at night and lasts until just before sunrise. This pattern is dramatically altered or obliterated in injured subjects. Most injury is characterized by elevations in CRH and ACTH that are proportional to the severity of injury. While pain and anxiety are prominent mediators of ACTH release in the conscious injured patient, other ACTH-promoting mediators may become relatively more active in the injured patient. These include vasopressin, angiotensin II, cholecystokinin, vasoactive intestinal polypeptide (VIP), catecholamines, oxytocin, and proinflammatory cytokines. Within the zona fasciculata of the adrenal gland, ACTH signaling activates intracellular adenylate cyclase, the cAMP-dependent protein kinase pathway, and mitochondrial cytochrome P-450 system. This chain of activities leads to increased glucocorticoid production via desmolase-catalyzed side-chain cleavage of cholesterol (Fig. 1-7). Conditions of excess ACTH stimulation result in adrenal cortical hypertrophy.

Cortisol/Glucocorticoids

Cortisol is the major glucocorticoid in human beings and is essential for survival after significant physiological stress. Cortisol levels in response to injury are not under the influence of normal diurnal variations and can remain persistently elevated, depending on the type of systemic stress. Burn patients have demonstrated elevated circulating cortisol levels for up to 4 weeks, and soft-tissue injury and hemorrhage may sustain elevated cortisol levels for up to a week. Circulating cortisol rapidly returns to normal levels on restoration of blood volume after hemorrhage. Conversely, adequate cortisol levels after mild hemorrhage is a prerequisite for timely restitution of blood volume in experimental animals. Coexisting systemic stress, such as infections, also can prolong the elevated cortisol levels after injury. Cortisol is a major effector of host metabolism. It potentiates the actions of glucagon and epinephrine, leading to hyperglycemia in the host. In the liver, cortisol stimulates the enzymatic activities favoring gluconeogenesis, including induction of phosphoenol pyruvate carboxykinase and transaminases. Peripherally, it decreases insulin-binding to insulin receptors in muscles and adipose tissue. In skeletal muscle, cortisol induces proteolysis and augments the release of lactate. The release of available lactate and amino acids has the net effect of shifting substrates for hepatic gluconeogenesis. Cortisol also stimulates lipolysis and inhibits glucose uptake by adipose tissues. It increases the lipolytic activities of ACTH, growth hormones, glucagon, and epinephrine. The resulting rises in plasma free fatty acids, triglycerides, and glycerol from adipose tissue mobilization serve as available energy sources and additional substrates for hepatic gluconeogenesis. About 10 percent of plasma cortisol is present in the free, biologically active form. The remaining 90 percent is bound to corticosteroid-binding globulin (CBG) and albumin. With injury, total plasma cortisol concentrations increase, but CBG and albumin levels decrease by as much as 50 percent. This can lead to an increase in free cortisol level of as much as ten times the normal level. Acute adrenal insufficiency is a life-threatening complication most commonly associated with adrenal suppression from the use of exogenous glucocorticoids with consequent atrophy of the adrenal glands. These patients present with weakness, nausea, vomiting, fever, and hypotension. Objective findings include hypoglycemia from decreased gluconeogenesis, hyponatremia from impaired renal tubular sodium resorption, and hyperkalemia from diminished kaliuresis. Although hyponatremia and hyperkalemia generally are a result of insufficient mineralocorticoid (aldosterone) activity, the loss of cortisol activity also contributes to electrolyte abnormalities. Glucocorticoids have long been used as effective immunosuppressive agents. Administration of glucocorticoids can induce rapid lymphopenia, monocytopenia, eosinopenia, and neutrophilia. Immunologic changes include thymic involution, depressed cell-mediated immune responses reflected by decreases in T killer and natural killer functions, T lymphocyte blastogenesis, mixed lymphocyte responsiveness, graft-versus-host reactions, and delayed hypersensitivity responses. With glucocorticoid administration, monocytes lose the capacity for intracellular killing, but they appear to maintain normal chemotactic and phagocytic properties. Neutrophil function is affected by glucocorticoid treatment in terms of intracellular superoxide reactivity and depressed chemotaxis. Phagocytosis of polymorphonuclear leukocytes (PMNs) remains unchanged. Glucocorticoids are omnibus inhibitors of immunocyte proinflammatory cytokine synthesis and secretion. This glucocorticoid-induced down-regulation of cytokine stimulation serves an important negative regulatory function in the inflammatory response to injury.

Macrophage Inhibitory Factor

Initially identified as a T lymphocyte–derived inhibitor of macrophage migration, macrophage inhibitory factor (MIF) is a glucocorticoid antagonist produced by the anterior pituitary.

This hormone can potentially reverse the immunosuppressive effects of glucocorticoids systemically via anterior pituitary secretion and at local sites of inflammation where MIF is produced by T lymphocytes. In experiments in which anti-MIF antibodies were administered to endotoxemic mice, survival increased presumably because glucocorticoid antiinflammatory effects were not counterregulated by MIF.

Thyrotropin-Releasing Hormone and Thyroid Stimulating Hormone

Thyrotropin-releasing hormone (TRH) serves as the primary stimulant for the synthesis, storage, and release of *thyroid-stimulating hormone (TSH)* in the anterior pituitary. TSH in turn stimulates thyroxine (T₄) production from the thyroid gland. T₄ is converted to triiodothyronine (T₃) by peripheral tissues. T₃ is more potent than T₄, but both are transported intracellularly by cytosolic receptors, which then bind DNA to mediate the transcription of multiple protein products. Free forms of T₄ and T₃ in the circulation can inhibit the hypothalamic release of TRH and pituitary release of TSH via negative feedback loops. TRH and estrogen stimulate TSH release by the pituitary, and T₃, T₄, corticosteroids, growth hormones, somatostatin, and fasting inhibit TSH release.

Thyroid hormones (thyronines), when elevated above normal levels, exert various influences on cellular metabolism and function. Thyronines enhance membrane transport of glucose and increase glucose oxidation. These hormones increase the formation and storage of fat when carbohydrate intake is excessive, but this process decreases during starvation. The increase in cellular metabolism from excess thyroid hormone production leads to proportional elevations in overall oxygen consumption as well as heat production. Although T₃ levels are frequently decreased after injury, there is no compensatory rise in TSH release. After major injury, reduced available T₃ and circulating TSH levels are observed and peripheral conversion of T₄ to T₃ is impaired. This impaired conversion may be explained in part by the inhibitory effects of cortisol and an increased conversion of T₄ to the biologically inactive molecule known as reverse T₃ (rT₃). Proinflammatory cytokines also may also contribute to this effect. Elevated rT₃, but reduced T₄ and T₃, is an observation characteristic of acute injury or trauma, referred to as *euthyroid sick syndrome* or *nonthyroidal illness*. Experimental mild endotoxemia in otherwise healthy human subjects has shown that thyroid hormone alterations in systemic inflammation is not mediated by endogenous IL-1.

While total T₄ (protein bound and free) levels may be reduced after injury, free T₄ concentrations remain relatively constant. In severely injured or critically ill patients, a reduced free T₄ concentration has been predictive of high mortality (Fig. 1-8).

Lymphoid cells have high-affinity nuclear and cytoplasmic binding sites for thyronines. One consequence of exposure to thyronines is an increase in the uptake of amino acids and glucose into the cell. Whether this is a direct effect of thyroid hormones or a secondary effect of increased cellular metabolism is unknown. Leukocyte metabolism measured by oxygen consumption is increased in hyperthyroid individuals and subjects to whom thyroid hormones have been administered. Animal studies have demonstrated that surgically or chemically induced thyroid hormone depletion significantly decreases cellular and humoral immunity. Conversely, thyroid hormone repletion is associated with enhancement of both types of immunity. Human monocytes, natural killer cells, and activated B lymphocytes express receptors for TSH. Exposure of B cells to TSH in vitro induces a moderate increase in immunoglobulin secretion.

Growth Hormones

Hypothalamic *growth hormone releasing hormone (GHRH)* travels through the hypothalamic-hypophyseal portal circulation to the anterior pituitary and stimulates the release of *growth hormone (GH)* in a pulsatile fashion mostly during the sleeping hours. In addition to GHRH, GH release is influenced by autonomic stimulation, thyroxine, AVP, ACTH, α -melanocyte stimulating hormone, glucagon, and sex hormones. Other stimuli for GH release include physical exercise, sleep, stress, hypovolemia, fasting hypoglycemia, decreased circulating fatty acids, and increased amino acid levels. Conditions that inhibit GH release include hyperglycemia, hypertriglyceridemia, somatostatin, beta-adrenergic stimulation, and cortisol.

The role of GH during stress is to promote protein synthesis while enhancing the mobilization of fat stores. Fat mobilization occurs by direct stimulation in conjunction with potentiation of adrenergic lipolytic effects on adipose stores. In the liver, hepatic ketogenesis also is promoted by GH. GH inhibits insulin release and decreases glucose oxidation, leading to elevated glucose levels.

The protein synthesis properties of GH after injury are partially mediated by the secondary release of *insulinlike growth factor-1 (IGF-1)*. This hormone, which circulates predominantly in bound form with several binding proteins, promotes amino acid incorporation and cellular proliferation and attenuates proteolysis in skeletal muscle and in the liver. IGFs, formerly referred to as somatomedins, are mediators of hepatic protein synthesis and glycogenesis. In the adipose tissue, IGF increases glucose uptake and lipid synthesis. In skeletal muscles, it increases glucose uptake and protein synthesis. IGF also has a role in skeletal growth by promoting the incorporation of sulfate and proteoglycans into cartilage. In vitro studies using proteoglycan synthesis as a marker for IGF-1 activity have demonstrated that interleukin-1 α , tumor necrosis factor- α (TNF- α), and interleukin-6 can inhibit the effects of IGF-1.

There is a rise in circulating GH levels after injury, major surgery, and anesthesia. The associated decrease in protein synthesis and observed negative nitrogen balance is attributed to a reduction in IGF-1 levels. GH administration has improved the clinical course of pediatric burn patients. Its use in injured adult patients is unproven. The liver is the predominant source of IGF-1 and preexisting hepatic dysfunction may contribute to the negative nitrogen balance after injury. IGF-binding proteins also are produced within the liver and are necessary for effective binding of IGF to the cell. IGF has the potential for attenuating the catabolic effects after surgical insults.

Leukocytes express high-affinity surface receptors for GH. GH and IGF-1 are immunostimulatory and promote tissue proliferation. In vitro, GH augments the proliferation of T lymphocytes to mitogens and the cytotoxicity of T killer cells to allogenic stimuli. Macrophages also respond to GH with a modest respiratory burst. GH-deficient mice manifest immune deficiencies that can be partially reversed by the administration of GH. GH-deficient human beings do not demonstrate any significant immunologic abnormalities. Normal subjects given intravenous GH demonstrate no significant immunological changes except for neutrophilia. GH has immunomodulating effects, but the relevance of this influence remains to be determined.

Somatostatin

Somatostatin is a 14-amino acid polypeptide produced by various cell types, including gastric antrum cells and pancreatic islet D cells. It is a potent inhibitor of GH, TSH, renin, insulin, and glucagon release. The role of somatostatin in the response to injury is unclear, but it may regulate excessive nutrient absorption and the activities of GH and IGF during convalescence from injury.

Gonadotrophins and Sex Hormones

Luteinizing-hormone releasing hormone (LHRH) or gonadotropin-releasing hormone (GnRH) is released from the hypothalamus and stimulates follicle-stimulating hormone (FSH) and *luteinizing hormone (LH)* release from the anterior pituitary. The release of these hormones can be effectively blocked by CRH, prolactin, estrogen, progestins, and androgens. The most relevant clinical correlation is seen after injury, stress, or severe illness, when LH and FSH release is suppressed. The reduction in LH and FSH consequently reduces estrogen and androgen secretion. This is attributed to the inhibitory activities of CRH on LH and FSH release and accounts for the menstrual irregularity and decreased libido reported after surgical stress and other injuries.

Estrogens inhibit cell-mediated immunity, natural killer cell activity, and neutrophil function, but are stimulatory for antibody-mediated immunity. Conditions associated with high-estrogen levels appear to predispose the patient to increased infectious complications. Androgens appear to be predominantly immunosuppressive. Castration is associated with enhanced immune function that can be reversed by exogenous androgens.

Prolactin

The hypothalamus suppresses prolactin secretion from the anterior pituitary by the activities of LHRH/GnRH and dopamine. Stimulants for prolactin release are CRH, TRH, GHRH, serotonin, and vasoactive intestinal polypeptide (VIP).

Elevated prolactin levels after injury have been reported in adults, whereas reduced levels are noted in children. The hyperprolactinemia also may account for the amenorrhea frequently seen in women after injury or major operations.

Like growth hormone, prolactin has immunostimulatory properties. Chemically induced inhibition of prolactin in animals has demonstrated increased susceptibility to infection, decreased lymphocyte proliferation, decreased interleukin-2 production and receptor expression, decreased interferon- γ production, and macrophage dysfunction. Exogenous administration of prolactin reversed these effects. There is increasing evidence that prolactin also is synthesized and secreted by T lymphocytes and may function in an autocrine or paracrine fashion.

Endogenous Opioids

Elevated endogenous opioids are measurable after major operations or insults to the patient (Table 1-4). The b-endorphins have a role in attenuating pain perception, and they are capable of inducing hypotension through a serotonin-mediated pathway. Conversely, the enkephalins produce hypertension. In the gastrointestinal tract, the activation of opioid receptors reduces peristaltic activity and suppresses fluid secretion.

The role of endogenous opioids in glucose metabolism is complex. While b-endorphins and morphine induce hyperglycemia, they also increase insulin and glucagon release by the pancreas. In animal models, endogenous opioids, such as dynorphins, have demonstrated a paracrine role in modulation of vasopressin and oxytocin secretion. Studies demonstrating the presence of opioid receptors in the adrenal medulla also suggest a role in regulating catecholamine release.

Certain immune cells also release endorphins that share an antinociceptive role in modulating the response of local sensory neurons to noxious stimuli. Endorphins also influence the immune system by increasing natural killer cell cytotoxicity and T cell blastogenesis. Interleukin-1 activates the release of POMC from the pituitary gland. Endogenous opioids (endorphin and enkephalin) and exogenous opiates both mediate their effects through mammalian delta, kappa, and mu receptors. Opioids compromise the natural (innate) and specific (adaptive) immune system. They inhibit the proliferation and differentiation of lymphocytes and monocytes/macrophages. The immunoinhibitory effects of exogenous opiates can be dose dependent and sensitive to the state of activation of the immune system.

Hormones Under Posterior Pituitary Regulation

Arginine Vasopressin

Vasopressin or arginine vasopressin (AVP) (or antidiuretic hormone, ADH) is synthesized in the anterior hypothalamus and transported by axoplasmic flow to the posterior pituitary for storage. The major stimulus for AVP release is elevated plasma osmolality, which is detected by sodium-sensitive hypothalamic osmoreceptors. There is evidence of extracerebral osmoreceptors for AVP release in the liver or the portal circulation. AVP release is enhanced by beta-adrenergic agonists, angiotensin II stimulation, opioids, anesthetic agents, pain, and elevated glucose concentrations. Changes in effective circulating volume by as little as 10 percent can be sensed by baroreceptors, left atrial stretch receptors, and chemoreceptors, leading to AVP release. Release is inhibited by alpha-adrenergic agonists, and atrial natriuretic peptide (ANP).

In the kidney, AVP promotes reabsorption of water from the distal tubules and collecting ducts. Peripherally, AVP mediates vasoconstriction. This effect in the splanchnic circulation may cause the trauma-induced ischemia/reperfusion phenomenon that precedes gut barrier impairment. AVP, on a molar basis, is more potent than glucagon in stimulating hepatic glycogenolysis and gluconeogenesis. The resulting hyperglycemia increases the osmotic effect that contributes to the restoration of effective circulating volume. Elevated AVP secretion is another characteristic of trauma, hemorrhage, open-heart surgery, and other major operations. This elevated level typically persists for 1 week after the insult.

The syndrome of inappropriate antidiuretic hormone secretion (SIADH) refers to the excessive vasopressin release that is manifested by low urine output, highly concentrated urine, and dilutional hyponatremia. This diagnosis can be made only if the patient is euvolemic. Once normal volume is established, a plasma osmolality below 275 mOsm/kg H₂O and a urine osmolality above 100 mOsm/kg H₂O is indicative of SIADH. SIADH is commonly seen in patients with head trauma and burns.

In the absence of AVP, central diabetes insipidus occurs and there is voluminous output of dilute urine. Frequently seen in comatose patients, the polyuria in untreated diabetes insipidus can precipitate a state of hypernatremia and hypovolemic shock. Attempts at reversal should include free water and exogenous vasopressin (desmopressin).

Oxytocin

Oxytocin and AVP are the only known hormones secreted by the posterior pituitary. They share structural similarities, but the role of oxytocin in the injury response is unknown. In human beings, the only consistent stimulus for secretion of oxytocin is suckling or other nipple stimulation in lactating women. This stimulates contraction of lactating mammary glands and induces uterine contractions in parturition. There is no recognized stimulus for oxytocin release, nor are there any known functions in men.

Hormones of the Autonomic System

Catecholamines

Catecholamines exert significant influence in the physiologic response to stress and injury. The hypermetabolic state observed after severe injury has been attributed to activation of the adrenergic system. Both of the major catecholamines, norepinephrine and epinephrine, are increased in plasma after injury, with average elevations of three to four times above baseline immediately after injury, reaching their peak in 24 to 48 h before returning to baseline levels. The patterns of norepinephrine and epinephrine appearance parallel each other after injury. Most of the norepinephrine in plasma results from synaptic leakage during sympathetic nervous system activity, while virtually all plasma epinephrine derives from the secretions of chromaffin cells of the adrenal medulla.

Catecholamines exert metabolic, hormonal, and hemodynamic influences on diverse cell populations. In the liver, epinephrine promotes glycogenolysis, gluconeogenesis, lipolysis, and ketogenesis. It causes decreased insulin secretion but increased glucagon secretion. Peripherally, epinephrine increases lipolysis in adipose tissues and inhibits insulin-facilitated glucose uptake by skeletal muscle. These collectively promote the often evident stress-induced hyperglycemia, not unlike the effects of cortisol on blood glucose. Catecholamines also increase the secretion of thyroid and parathyroid hormones, T₄ and T₃, and renin, but inhibit the release of aldosterone.

Catecholamines exert discernible influences on immune function, e.g., epinephrine occupation of beta receptors on leukocytes increases intracellular cAMP. This ultimately decreases immune responsiveness in lymphocytes. Like cortisol, epinephrine enhances leukocyte demargination with resultant neutrophilia and lymphocytosis. Epinephrine also lowers the ratio of CD4 to CD8 T lymphocytes. Immunologic tissue such as the spleen, thymus, and lymph nodes possess extensive adrenergic innervation. Chemical sympathectomy of peripheral nerves has been demonstrated to augment antibody response after immunization with a specific antigen. It also reverses the depressed mitogenic response of splenocytes preincubated with endotoxin. Normal volunteers infused with epinephrine exhibit depressed mitogen-induced T lymphocyte proliferation.

Aldosterone

The mineralocorticoid aldosterone is synthesized, stored, and released in the adrenal zona glomerulosa. Its release may be induced by angiotensin II, hyperkalemia, and the pituitary hormone known as aldosterone stimulating factor (ASF), but ACTH is the most potent stimulus for aldosterone release in the injured patient.

The major function of aldosterone is to maintain intravascular volume by conserving sodium and eliminating potassium and hydrogen ions. While the major effect is exerted in the kidneys, this hormone also is active in the intestines, salivary glands, sweat glands, vascular endothelium, and the brain. In the early distal convoluted tubule, aldosterone increases sodium and chloride reabsorption and excretion of hydrogen ions. In the late distal convoluted tubule, further sodium reabsorption takes place while potassium ions are excreted. Vasopressin also acts in concert with aldosterone to increase osmotic water flux into the tubules.

Patients with aldosterone deficiency develop hypotension and hyperkalemia, whereas patients with aldosterone excess develop edema, hypertension, hypokalemia, and metabolic alkalosis. After injury, ACTH stimulates a brief burst of aldosterone release. Angiotensin II induces a protracted aldosterone release that persists well after ACTH returns to baseline. As with cortisol, normal aldosterone release also is influenced by the circadian cycle, though this effect is lost in the injured patient.

Renin-Angiotensin

Renin is synthesized and stored primarily within the renal juxtaglomerular apparatus near the afferent arteriole. The juxtaglomerular apparatus comprises the juxtaglomerular neurogenic receptor, the juxtaglomerular cell, and the macula densa. Renin initially exists in an inactive form as prorenin. The activation of renin and its release is mediated by ACTH, AVP, glucagon, prostaglandins, potassium, magnesium, and calcium. The juxtaglomerular cells are baroreceptors that respond to a decrease in blood pressure by increasing renin secretion. The macula densa detects changes in chloride concentration in the renal tubules.

Angiotensinogen is a protein primarily synthesized by the liver but also identified in the kidney. Renin catalyzes the conversion of angiotensinogen to angiotensin I within the kidney. Angiotensin I remains physiologically inactive until it is converted in the pulmonary circulation to angiotensin II by angiotensin-converting enzyme present on endothelial surfaces.

Angiotensin II is a potent vasoconstrictor that also stimulates aldosterone and vasopressin synthesis. It also is capable of regulating thirst. Angiotensin II stimulates heart rate and myocardial contractility. It also potentiates the release of epinephrine by the adrenal medulla, increases CRH release, and activates the sympathetic nervous system. It can induce glycogenolysis and gluconeogenesis. The renin-angiotensin system participates in the response to injury by maintaining volume homeostasis.

Insulin

Insulin is derived from pancreatic beta islet cells and released upon stimulation by certain substrates, autonomic neural input, and other hormones. In normal metabolism, glucose is the major stimulant of insulin secretion. Other substrate stimulants include amino acids, free fatty acids, and ketone bodies. Hormonal and neural influences during stress alter this response. Epinephrine and sympathetic stimulation inhibit insulin release. Other factors that diminish insulin release include glucagon, somatostatin, gastrointestinal hormones, b-endorphins, and interleukin-1. Peripherally, cortisol, estrogen, and progesterone interfere with glucose uptake. The net result of impaired insulin production and function after injury is stress-induced hyperglycemia, which is in keeping with the general catabolic state.

Insulin exerts a global anabolic effect; it promotes hepatic glycogenesis and glycolysis, glucose transport into cells, adipose tissue lipogenesis, and protein synthesis. In the injured patient, a biphasic pattern of insulin release is observed. The first phase occurs within a few hours after injury and is manifested as a relative suppression of insulin release, reflecting the influence of catecholamines and sympathetic stimulation. The later phase is characterized by a return to normal or excessive insulin production but with persistent hyperglycemia, demonstrating a peripheral resistance to insulin. The ratio of insulin to glucose (not their individual values) is used as a predictor of mortality and survival.

Activated lymphocytes express receptors for insulin. Insulin enhances T lymphocyte proliferation and cytotoxicity. Mouse spleen cells transiently exposed to a mitogen can continue to proliferate and maintain cytotoxicity if insulin is added to the medium. Institution of insulin therapy to newly diagnosed diabetics is associated with increased B and T lymphocyte populations.

Glucagon

Glucagon is a product of pancreatic alpha islet cells. As with insulin, the release of glucagon also is mediated by its substrates, autonomic neural input, and other hormones. Whereas insulin is an anabolic hormone, glucagon serves more of a catabolic role. The primary stimulants of glucagon secretion are plasma glucose concentrations and exercise. Glucagon stimulates hepatic glycogenolysis and gluconeogenesis, which under basal conditions account for approximately 75 percent of the glucose produced by the liver. In contrast to insulin, glucagon promotes hepatic ketogenesis and lipolysis in adipose tissue. The release of glucagon after injury is initially decreased, but returns to normal 12 h later. By 24 h, glucagon levels are supranormal and can persist for up to 3 days.

IMMUNE RESPONSE TO INJURY

While the classic neuroendocrine response to injury has been extensively investigated, many characteristics of the inflammatory response associated with injury remain unexplained. Even after the normalization of macroendocrine hormone function after the primary injury, the persistence of systemic inflammation, the progression of organ dysfunction, and even late mortality indicate the presence of other potent mediators influencing the injury response. These mediators usually are small proteins or lipids that are synthesized and secreted by immunocytes. These micromolecules, collectively referred to as cytokines, are indispensable in tissue healing and in the immune response generated against microbial invasions. As mounting evidence suggests, the activities of these cytokine mediators are integrally related to classic hormone function and metabolic responses to injury.

Cytokine-Mediated Response

Patients with injuries or infections exhibit hemodynamic, metabolic, and immune responses partially orchestrated by endogenous cytokines. Unlike classical hormonal mediators such as catecholamines and glucocorticoids, which are produced by specialized tissues and exert their influence predominantly by endocrine routes, cytokines are produced by diverse cell types at the site of injury and by systemic immune cells (Table 1-5). Cytokine activity is primarily exerted locally via cell-to-cell interaction (paracrine).

Cytokines are small polypeptides or glycoproteins that exert their influence at very low concentrations. In their monomeric form, most are less than 30 kilodaltons (kD). In their biologically active form, some of these cytokines function as oligomers (e.g., trimeric tumor necrosis factor- α) with higher molecular weights. Most cytokines also differ from classical hormones in that they are not stored as preformed molecules. Their relatively rapid appearance after injury reflects active gene transcription and translation by the injured or stimulated cell.

Cytokines exert their influence by binding to specific cell receptors and activating intracellular signaling pathways leading to modulation of gene transcription. By this mechanism, cytokines influence immune cell production, differentiation, proliferation, and survival. These mediators also regulate the production and actions of other cytokines, which may either potentiate (proinflammatory) or attenuate (antiinflammatory) the inflammatory response. The capacity of cytokines to activate diverse cell types and to incite equally diverse responses underscores the pleiotropism of these inflammatory mediators (Table 1-6). There is also a marked degree of overlapping activity among different cytokines.

Cytokines are effector molecules that direct the inflammatory response to infections (bacterial, viral, and fungal) and injury and actively promote wound healing. These responses are manifested by fever, leukocytosis, and alterations in respiratory and heart rates. It is the exaggerated, acute production of proinflammatory cytokines that is responsible for the hemodynamic instability characteristic of septic shock. The chronic and excessive production of these cytokines is partly responsible for the metabolic derangements of the injured patients, such as debilitating muscle wasting and cachexia. Preexisting cytokine production can contribute to end-organ injury leading to multiple organ failure and late mortality in severely injured or infected patients. The presence of antiinflammatory cytokines may serve to attenuate some of these exaggerated responses. The excessive release of antiinflammatory cytokines may render the patient immunocompromised and increase susceptibility to infections.

Understanding of the pathophysiology of inflammatory cytokine mediators has been derived largely from patients with endotoxemia or sepsis. Inflammatory mediator responses to infections and traumatic injury are not dissimilar, particularly in the temporal sequence of cytokine expression. The cytokine response evidenced by fever, leukocytosis, hyperventilation, and tachycardia commonly seen in injury is referred to as systemic inflammatory response syndrome (SIRS) and is not necessarily the result of an identifiable infectious process. Central to the insult suffered by the host and the subsequent inflammatory response is the activity of the host's immunocyte population, circulating and tissue-fixed. Discussions of the inflammatory response should not be dissociated from these cellular entities.

The cytokine cascade activated in response to injury consists of a complex network with diverse effects on all aspects of physiological regulatory mechanisms. Cytokines are pivotal determinants of the host response after injury and a proper perspective of their immunobiologic sequelae can have important applications in the comprehensive care of the surgical patient. The number of cytokines identified has expanded to nearly 30, but their functions and elicited responses, particularly in injury, are incompletely understood largely because of the pleiotropic, redundant, and mutual interactions among these mediators. The cytokines described here represent a limited list of better-characterized mediators related to injury and the inflammatory response.

Tumor Necrosis Factor- α

The inflammatory response to severe cross-sectional tissue injury or infectious agents evokes a complex cascade of proinflammatory cytokines. Among these, *tumor necrosis factor- α* (TNF- α) is the earliest and one of the most potent mediators of the subsequent host response. The sources of TNF- α synthesis include monocytes/macrophages and T cells, which are abundant in the peritoneum and splanchnic tissues. Kupffer cells represent the single largest concentrated population of macrophages in the human body. Surgical

or traumatic injuries to the viscera may have profound influences on the generation of inflammatory mediators and homeostatic responses such as acute phase protein production (Fig. 1-9).

The release of TNF- α in response to acute injury is rapid and short-lived. Experiments simulating an acute inflammatory response by means of endotoxin challenge in human subjects have demonstrated a monophasic *tumor necrosis factor (TNF)* appearance curve, peaking at approximately 90 min and followed by a return to undetectable levels within 4 h (Fig. 1-10). Even with a half-life of 15 to 18 min, the brief appearance of TNF can induce marked metabolic and hemodynamic changes and activate cytokines distally in the cascade. The abbreviated production of TNF implies the presence of effective endogenous modulators, which serve to prevent any propagation of unregulated TNF- α activity. This has been proved because several natural mechanisms that antagonize TNF production or activity have been identified. Endogenous inhibitors in the form of cleaved extracellular domains of the transmembrane TNF receptors (*soluble TNF receptors, sTNFRs*) are readily detectable in the circulation. These receptors may serve a protective role by competitively sequestering excess circulating TNF, but are probably only capable of doing so against low levels of TNF activity and for brief periods.

TNF- α also is a major cytokine related to muscle catabolism and cachexia during stress. Amino acids are mobilized from skeletal muscles and shunted toward the hepatic circulation as fuel substrates. Studies have demonstrated that TNF- α -induced muscle catabolism occurs through a ubiquitin-proteasome proteolytic pathway with increased expression of the ubiquitin gene.

Other associated functions of TNF- α include coagulation activation and promoting the release of prostaglandin E₂ (PGE₂), platelet-activating factor (PAF), glucocorticoids, and eicosanoids.

Interleukin-1

TNF- α also induces the biosynthesis and release of interleukin-1 (IL-1) from macrophages and endothelial cells. There are two known proinflammatory species of IL-1, IL-1 α and IL-1 β . IL-1 α is predominantly cell membrane-associated and exerts its influence via cellular contacts. The more detectable form released in the circulation is IL-1 β , which is produced in greater quantities than IL-1 α and capable of inducing the characteristic systemic derangements after injury. The potency and effects of IL-1 reflect those of TNF- α , eliciting similar physiologic and metabolic alterations. At high doses of IL-1 and TNF- α , these cytokines independently initiate a state of hemodynamic decompensation. At low doses, they can produce the same response only if administered simultaneously. These observations emphasize the synergism of TNF- α and IL-1 in eliciting some proinflammatory responses. The half-life of IL-1 is approximately 6 min, which, along with its primary role as a local inflammatory mediator, makes its detectability in acute injury or illness even less likely than that of TNF- α .

Among its effects, IL-1 induces the classic inflammatory febrile response to injury by stimulating local prostaglandin activity in the anterior hypothalamus. Associated with the hypothalamic activity is the induction of anorexia by an IL-1 effect on the satiety center. This cytokine also augments T cell proliferation by enhancing the production of IL-2 and also may influence skeletal muscle proteolysis, characteristic of cachexia. Attenuated pain perception after surgery can be mediated by IL-1 by promoting the release of β -endorphins from the pituitary gland and increasing the number of central opioid-like receptors. Like TNF, IL-1 is a potent stimulant for ACTH and glucocorticoid release via its actions on the hypothalamus and pituitary gland.

A non-agonist IL-1 species, known as *IL-1 receptor antagonists (IL-1ra)*, also is released during injury. This molecule effectively competes for binding to IL-1 receptors yet exacts no overt signal transduction. IL-1ra, which is often detectable during inflammation or injury, serves as a potent regulator of IL-1 activity.

Distal cytokine mediators, released as part of the inflammatory cascade initiated by TNF- α and IL-1, include IL-2, IL-4, IL-6, IL-8, *granulocyte/macrophage colony-stimulating factor (GM-CSF)*, and *interferon- γ (IFN- γ)*.

Interleukin-2

Although necessary as an inflammatory mediator in promoting T lymphocyte proliferation, immunoglobulin production, and gut barrier integrity, IL-2 has not been readily detectable in the circulation during acute injury. Similar to IL-1, its short half-life of less than 10 min adds to the difficulty in detecting it after injury. IL-2 secretion by lymphocytes is impaired after acute injury and several disease states, notably cancer and acquired immunodeficiency syndrome (AIDS). Perioperative blood transfusions also are associated with reduced IL-2 production. Attenuated IL-2 expression contributes to the transient immunocompromised state of the surgical patient. A low point in gut barrier IL-2 activity resulting from injuries can predispose the patient to enteric organism activation of the inflammatory cytokine cascade. There is evidence for accelerated lymphocyte programmed cell death (apoptosis), in association with diminished IL-2 activity, mediated by the proapoptotic Fas/CD95 cell receptor in the early postoperative period. The combined diminution of lymphocyte survival and IL-2 activity may contribute to the immunocompromised phenotype of the injured patient.

Studies have demonstrated a population density shift from type 1 T helper cells (T_H1, cell-mediated and opsonizing antibody immune responses, including IL-2, IL-12, and IFN- γ production) to type 2 T helper cells (T_H2, IgE antibody-mediated immune response, including IL-4, IL-6, IL-10, IL-13 production) after surgical stress. The immune activities of the T_H2 response usually are less effective against microorganisms, and they accentuate the risks for postoperative infections (Fig. 1-11).

Interleukin-4

IL-4 is a glycoprotein molecule, produced by activated T_H2 cells, with diverse biologic effects on hemopoietic cells, including induction of B lymphocyte proliferation. Of particular importance in its role in antibody-mediated immunity is the capacity to enhance macrophage major histocompatibility complex class II (HLA-DR and HLA-DP) expression and adhesion molecules, making them efficient antigen-presenting cells. IL-4 also induces class switching in differentiating B lymphocytes to produce predominantly IgG₄ and IgE, which are important immunoglobulins in allergic and antihelminthic responses.

As a potent antiinflammatory cytokine, IL-4 down-regulates several functions associated with activated human macrophages, namely, the effects of IL-1 β , TNF- α , IL-6, IL-8, and superoxide production. These antiinflammatory effects of IL-4 are not seen in resting monocytes. The importance of this cytokine is the capacity to down-regulate the response of inflammatory macrophages exposed to stimuli such as bacterial endotoxin or proinflammatory cytokines. IL-4 can induce programmed cell death in inflammatory macrophages, but this effect is abrogated by IFN- γ . IL-4 and IFN- γ antagonize one another's effects on B cells. IL-4 also appears to increase macrophage susceptibility to the antiinflammatory effects of glucocorticoids. IL-13 may share several properties with IL-4.

Interleukin-6

Because of the elevated blood levels of IL-6 often observed during acute injury or stress, it is used frequently as an indicator of the systemic inflammatory response and a predictor of preoperative morbidity. TNF- α and IL-1 are major inducers of IL-6. IL-6 can be produced by virtually all cell types, including the intestines. After injury, IL-6 levels in the circulation are detectable by 60 min, peak between 4 to 6 h, and can persist for as long as 10 days. The relatively long half-life partially explains its ease of detectability. IL-6 levels appear to be proportional to the extent of tissue injury during an operation rather than the duration of the surgical procedure itself.

Evidence suggests a complex role for IL-6 in mediating proinflammatory and antiinflammatory activities. IL-1 and IL-6 are important mediators of the hepatic acute-phase protein response during injury and appear to enhance C-reactive protein, fibrinogen, haptoglobin, amyloid A, α -1-antitrypsin, and complement production (see Fig. 1-9). IL-6 not only induces PMN activation during injury and inflammation but also may delay the phagocytic disposal of senescent or dysfunctional PMNs during injury. The persistence of inflammatory PMNs after injury might explain the injurious effects on distant tissues, such as the pulmonary or renal system.

IL-6 mediates the antiinflammatory pathway during injury through different mechanisms. It is capable of attenuating TNF and IL-1 activity while promoting the release of sTNFRs and IL-1ra. Prolonged and persistent expression of IL-6 is associated with immunosuppression and postoperative infectious morbidity. Elevated IL-6 levels postoperatively can impair glutaminase activity, causing a reduction in plasma glutamine.

Interleukin-8

The appearance of IL-8 activity is temporally associated with IL-6 after injury and has been proposed as an additional biomarker for the risk of multiple organ failure.

IL-8 does not produce the hemodynamic instability characteristic of TNF- α and IL-1 but rather serves as a PMN activator and potent chemoattractant. IL-8 is being established as a major contributor to organ injury such as the acute lung injury.

Interleukin-10

IL-10 is an important endogenous regulatory mediator during the inflammatory response; it acts primarily by modulating TNF- α activity. Its appearance in the circulation during endotoxemia closely follows the appearance TNF- α . Supporting experiments have demonstrated that neutralization of IL-10 during endotoxemia increases monocyte TNF- α production and mortality, but restitution of IL-10 reduces TNF- α levels and the associated deleterious effects. IL-10 may have additional protective roles after injury-induced inflammation by promoting IL-1ra and sTNFR production. In animal experiments, the sustained systemic production of IL-10 during septic peritonitis modulates the systemic inflammatory response. Murine experiments have demonstrated rapid induction of IL-10 messenger RNA (mRNA) activity after cecal ligation and puncture, and higher mortality when this activity is blocked with anti-IL-10. This immunomodulatory effect also may abrogate the proinflammatory response necessary for local clearance of invading organisms.

Interleukin-12

The capacity of IL-12 to promote the differentiation of T_H1 cells and the production of IFN- γ makes it a pivotal molecule in cell-mediated immunity after injury or infection. In mice with fecal peritonitis, survival increases with IL-12 administration. IL-12 also is implicated in preventing programmed cell death (apoptosis) in certain T lymphocyte populations after their activation.

Interleukin-13

IL-13 is a pleiotropic cytokine that shares many of the properties of IL-4 as well as a modest amino acid sequence (about 30 percent). IL-13 is produced during T_H2 responses. IL-4 and IL-13 modulate macrophage function, but IL-13 has no identifiable effect on T lymphocytes and only has influence on subpopulations of B lymphocytes. IL-4 and IL-13 receptors share a common signaling component. IL-13 can up-regulate macrophage major histocompatibility complex class I and II antigens and other surface antigens, such as CD23. IL-13 can inhibit nitric oxide production and the expression of proinflammatory cytokines, and it can enhance the production of IL-1ra. The net effect of IL-13, along with IL-4 and IL-10, is antiinflammatory.

Interferon- γ

Much of IL-12 biology is mediated through the production and activities of IFN- γ . Human T helper (T_H) cells activated by the bacterial antigens IL-2 or IL-12 readily produce IFN- γ .

Conversely, IFN- γ can induce the production of IL-2 and IL-12 by T helper cells. With its release from activated T cells, IFN- γ is detectable in vivo by 6 h and has a half-life of approximately 30 min. IFN- γ levels peak at 48 to 72 h and may persist for 7 to 8 days. Injured tissues, such as operative wounds, also demonstrate the presence of IFN- γ production 5 to 7 days after injury. Natural killer cells are potent inducers of IL-12 production.

IFN- γ has important roles in activating circulating and tissue macrophages. Alveolar macrophage activation mediated by IFN- γ may induce acute lung inflammation after major surgery or trauma.

Granulocyte/Macrophage-Colony Stimulating Factor

Granulocyte/macrophage-colony stimulating factor (GM-CSF) production is induced by IL-2 and endotoxin. In vitro studies have demonstrated a prominent role for GM-CSF in delaying apoptosis of macrophages and PMNs. This growth factor is effective in promoting the maturation and recruitment of functional leukocytes necessary for normal inflammatory cytokine response, and potentially in wound healing. The mechanisms may be the result of the suppression of IL-10 production. Results of perioperative GM-CSF administration in patients undergoing major oncologic procedures have demonstrated augmentation of neutrophil numbers and function.

Regulation of Inflammatory Cell Death

Programmed Cell Death

During systemic inflammation, the response mounted by the host to injury and infection manifests the collective activities of circulating and tissue-fixed immunocytes and endothelial cell populations. In the normal host, programmed cell death (apoptosis) is the principal mechanism by which senescent or dysfunctional cells, including macrophages and PMNs, are systematically disposed of without activating other immunocytes or the release of proinflammatory contents. The signals inducing normal apoptosis differ from cell to cell but most likely converge at a common final pathway. These signals arise from the extracellular environment and may include hormonal and paracrine activities (Fig. 1-12). The inflammatory milieu disrupts the normal apoptotic machinery in dysfunctional or aging cells, consequently delaying the disposal of activated macrophages and PMNs. Several proinflammatory cytokines delay the normal temporal sequence of macrophage and PMN apoptosis in vitro. These include TNF, IL-1, IL-3, IL-6, GM-CSF, granulocyte colony-stimulating factor (G-CSF), and IFN-g. By contrast, IL-4 and IL-10 accelerate apoptosis in activated monocytes. The prolonged survival of inflammatory immunocytes may perpetuate and augment the inflammatory response to injury and infection, precipitating multiple organ failure and eventual death in severely injured and critically ill patients.

TNF Receptor-Mediated Programmed Cell Death

Tumor necrosis factor receptors (TNFRs) belong to a superfamily of approximately 15 transmembrane proteins that are present on virtually all cells, including immunocytes. Members of this family, which are linked by their conserved extracellular sequences, also include lymphotoxin-b receptor, Fas/CD95 (APO-1), nerve growth factor receptors (NGFR), CD27, CD30, OX40, 4-1BB, DR3 (WSL-1, APO-3, TRAMP), and DR4 (APO-2). Activation of these receptors induce specific cell responses that may include initiation of programmed cell death. There are two specific transmembrane TNFRs (type I, p55; and type II, p75), but they have distinct intracellular domains. The p55 TNFR induces apoptosis, cytotoxicity, expression of adhesion molecules on endothelial cells, and activation of the sphingomyelin pathway and nuclear factor-kappa B (NF-kB). The p75 TNFR induces proliferation of T cells, fibroblasts, natural killer cells, and proinflammatory cytokine release. The p55 TNFR has the dominant role in triggering apoptosis, but the concurrent participation of type I and type II TNFRs is necessary for initiating this process. The participation of both receptors is required, because activated intracellular p75 TNFR-related protein transducers are shared by the p55 TNFR signaling complex. During sepsis and experimental endotoxemia, down-regulation of macrophage and PMN TNFR activity is observed. This attenuation in TNFR activity may delay apoptosis of inflammatory macrophages and PMNs, prolonging the inflammatory response (Fig. 1-13). The p55 TNFR and Fas receptors exhibit similar cytoplasmic sequence motifs, known as the "death domain." These death domains interact with other intracellular proteins to propagate downstream signaling for programmed cell death. While TNFR I is found on virtually all cell types, Fas (CD95, APO-1) expression in murine models is predominantly expressed in the liver, lung, heart, intestine, skin, and lymphocytes. In human beings, Fas expression also is tissue-specific. When triggered by its specific ligand, FasL, Fas induces its only known function, which is to initiate apoptosis. Mutations of Fas or FasL are implicated as a cause for lymphoproliferative disorders and delayed disposal of inflammatory macrophages.

Fas/CD95 Receptor-Mediated Programmed Cell Death

The only known role of the Fas receptor is to initiate programmed cell death. Because of the intracellular homology of Fas to p55 TNFR, they both induce apoptosis via similar mechanisms, but Fas-mediated apoptosis occurs with greater speed (within hours) than that mediated by p55 TNFR. This may indicate a more direct (i.e., less complex) pathway for Fas-mediated apoptosis than that of the TNFR type I pathway. While the induction of apoptosis via Fas/FasL cross-linking in activated immunocytes may be advantageous during systemic inflammation, this activity at the tissue level may be detrimental to the host. Fas-mediated activity in the liver during inflammation may precipitate or exacerbate ongoing hepatic injury. There is hepatic parenchymal up-regulation of Fas receptors with acute toxic injury simultaneously enhanced by FasL expression of infiltrating lymphocytes. Studies also suggest a role for Fas/FasL interaction in thyroid gland destruction and thyroiditis. Therapeutic strategies derived from Fas/FasL interaction requires selectivity to minimize inadvertent organ injury.

Immunocyte Receptor Activity in Inflammation

Membrane TNFR

In human endotoxemia, TNFR expression in macrophages and PMNs is down-regulated. In macrophages, the decrease in surface TNFR reaches a low point 2h after endotoxin infusion and recovers to normal levels in 6h (Fig. 1-14). This receptor recovery continues to supranormal levels at 24 h. PMNs exhibit a more sustained decrease in surface TNFRs under the same conditions. Among signs such as fevers, leukocytosis, and chills, macrophage TNFR expression pattern is the most sensitive correlate identified for human response to endotoxin exposure. There is a reduction of cell-surface TNFRs in septic patients. Nonsurviving patients with severe sepsis have an immediate reduction in cell-surface TNFR expression, while surviving patients have almost normal receptor levels from the outset. TNFR expression can potentially be used as an indicator of outcome in patients with severe sepsis.

Soluble TNFR

Soluble TNFRs, proteolytically cleaved extracellular domains of membrane-associated TNFRs, also are elevated in patients with severe sepsis. sTNFRs retain their affinity for TNF and therefore compete with the cellular receptors for the binding of free TNF. This represents a counterregulatory response to excessive systemic TNF activity. In contrast to macrophage membrane TNFRs, nonsurviving septic patients demonstrate a significant elevation only in the p55 sTNFR compared to surviving patients. Cell-associated TNFR expression is more reliable than sTNFR as an early predictor of risk and outcome in human sepsis (Fig. 1-15).

Hormones and Cytokine Interactions

Cortisol/Glucocorticoids

Hypercortisolemia differentially influences leukocyte counts and cytokine expression in a temporal fashion. Glucocorticoid administration immediately before or concomitantly with endotoxin infusion in healthy human beings is able to attenuate the symptoms (e.g., fever, tachycardia), catecholamine response, and acute phase response, but it increases IL-10 release. Increased IL-10 release may contribute to the acute antiinflammatory effect of glucocorticoids. Hypercortisolemia induced by 6 h or more of glucocorticoid administration before endotoxin infusion does not attenuate the responses that are seen from endotoxin infusion alone. Infusion of cortisol for more than 12 h before endotoxin infusion increases TNF and IL-6 release. This may explain the varied systemic responses to infection in critically ill or severely injured patients who have associated hypercortisolemia. Such responses are influenced by antecedent events that alter the hormonal milieu. Glucocorticoids also can influence the regulation of T lymphocyte proliferation or programmed cell death, as demonstrated by in vitro dexamethasone-induced apoptosis of human T lymphocytes. CD8+ T cells are more sensitive to glucocorticoid-induced apoptosis than CD4+ cells. Glucocorticoid-induced apoptosis of T lymphocytes requires elevations of intracellular cAMP. IL-2, IL-4, and IL-10 protects these T lymphocytes from glucocorticoid-induced apoptosis. The proinflammatory cytokine IL-1 and, probably, TNF and IL-6 can activate the hypothalamus-pituitary-adrenal axis and induce the release of CRH and ACTH, leading to increased circulatory glucocorticoid levels. Glucocorticoids, in turn, inhibit endotoxin-induced production of TNF at the level of mRNA translation. Dexamethasone also inhibits neutrophil apoptosis and prolongs their functional responsiveness. This can be detrimental to the patient because the delay in clearance from tissues may perpetuate the injurious effects of activated neutrophils.

Catecholamines

Catecholamines inhibit endotoxin-induced macrophage production of TNF-a in vitro and in human whole blood ex vivo. In normal human subjects, short-term preexposure to epinephrine effectively inhibits endotoxin-induced TNF production. Concurrently, this short-term preexposure to epinephrine increases the production of the antiinflammatory cytokine IL-10. Longer preexposure, 24 h, had a less pronounced antiinflammatory effect. Endogenous epinephrine or exogenous administration as a component of sepsis treatment may serve to limit excessive proinflammatory effects of the cytokine network during the early phase of systemic infection. Epinephrine attenuates endotoxin-induced down-regulation of TNFR expression on human monocytes in vivo, an effect that is beta-receptor mediated and cAMP dependent. The use of catecholamines in treatment may have the potential for influencing immune cell function.

OTHER MEDIATORS OF INJURY RESPONSE

Endothelial Cell Mediators

Endothelial Cell Function

In addition to modulating coagulation and vasomotor activities, mediators elaborated by the vascular endothelium in response to injury are well-documented contributors to the inflammatory process. In a paracrine fashion, local mediators such as TNF-a, IL-1, endotoxin, thrombin, histamine, and IFN-g are capable of stimulating or activating the endothelial cell during local tissue injury. In response, the endothelial cell releases several mediators, including IL-1, platelet-activating factor (PAF), prostaglandins (PGI₂ and PGE₂), GM-CSF, growth factors, endothelin, nitric oxide, and small amounts of thromboxane A₂ (TxA₂). Activated endothelial cells also release collagenases capable of autodigesting their own basement membranes. This permits neovascularization and vascular remodeling at sites of injury in order to facilitate adequate oxygen supply and immunocyte transport. Angiotensin-converting enzymes (ACE) convert angiotensin I to angiotensin II on the surface of endothelial cells, making it a potent regulator of vascular tone. Endothelial cell mediators can modulate cardiovascular and renal function and influence the hypothalamus-pituitary-adrenal axis (Fig. 1-16). The activated endothelial cell up-regulates its expression of leukocyte adhesion receptor molecules such as E-selectin (formerly referred to as endothelial-leukocyte adhesion molecule-1, ELAM-1), P-selectin, and intercellular adhesion molecules (ICAM-1, ICAM-2). The adhesion of leukocytes and platelets to the endothelial surface occurs early in the endothelial-derived inflammatory process. In cultured endothelial cells, basal expression of E-selectin during inflammation requires the stimulation of TNF-a and IL-1. Within 1 h after treatment with either of these two cytokines, mRNA activity for E-selectin is detectable. The expression of E-selectin on endothelial cell surfaces is maximal at 4 to 6 h. Recovery from the inflammatory process also is characterized by internalization of these adhesion molecules within the endothelial cell. Neutrophil adhesion to the endothelium during injury has important clinical implications for increasing vascular permeability and passage of leukocytes into injured tissues. These are important in the etiology of conditions such as acute lung and ischemia-reperfusion injuries. In the nonstressed state, the endothelium possesses little capacity to recognize and bind circulating leukocytes. Local injuries and inflammatory mediator stimulation promote the margination of circulating PMNs to the endothelial surfaces. These marginated PMNs are deformable and travel along the endothelial surfaces at markedly reduced velocities, which is referred to as *rolling*. Rolling represents a process of transient attachment and detachment between receptors of PMNs and the endothelium. The subsequent development of stronger receptor adhesions, PMN activation by the endothelial mediators, and release of PMN proteinases at endothelial junctions precedes the migration of PMNs out of the vascular compartment, a process referred to as *diapedesis*. Although necessary for local tissue inflammation and eradication of microbes, activated PMNs and the subsequent release of inflammatory mediators and reactive oxygen metabolites are implicated in capillary leakage, acute lung injury, and postischemic injury (Fig. 1-17). The release of mediators by the endothelium and their subsequent influence on neighboring and distant tissues ascribes endocrine properties to endothelial cells during injury. The ability to attract leukocytes and produce inflammatory mediators makes endothelial cells important participants in the immune response to injury.

Endothelium-Derived Nitric Oxide

Endothelium-derived nitric oxide or relaxing factor (EDNO or EDRF) can be released in response to acetylcholine stimulation, hypoxia, endotoxin, cellular injury, or mechanical

shear stress from circulating blood. Its vasodilatory activity has been demonstrated in large (conduit) arteries and in resistance vessels of most mammalian species, including human beings. Induction of vascular smooth muscle relaxation by EDNO requires the activation of soluble guanylate cyclase and an increase in cytosolic cyclic guanosine monophosphate (cGMP) within the myocytes. Methylene blue inhibits guanylate cyclase, prevents the production of cGMP, and inhibits vascular relaxation. cGMP also is present in platelets and can be activated by EDNO. Increased cGMP in platelets is associated with reduced adhesion and aggregation. EDNO induces vasodilation and platelet deactivation (Fig. 1-18). EDNO also mediates protein synthesis in hepatocytes and electron transport in hepatocyte mitochondria. It is a readily diffusible substance with a half-life of a few seconds. EDNO spontaneously decomposes into nitrate and nitrite.

EDNO is formed from oxidation of L-arginine, a process catalyzed by nitric oxide synthase (NO-synthase). Cofactors of NO-synthase activity include calmodulin, ionized calcium, and NADPH. In addition to the endothelium, this enzymatic activity also is present in PMNs, macrophages, renal cells, Kupffer cells, and cerebellar neurons.

In normal vasculature, experiments blocking EDNO activity induce a state of vasoconstriction that is readily reversed with L-arginine administration. This demonstrates that the vasculature is in a constant state of vasodilation because of the continuous basal release of EDNO. Endogenous inhibitors of EDNO have been identified that are autoregulators of endothelial tone.

Elevations of EDNO in septic shock and trauma, as measured by its nitrite and nitrate metabolites, are evidenced in association with low systemic vascular resistance and elevated endotoxin levels.

Prostacyclin

Prostacyclin (PGI₂) is an important endothelium-derived vasodilator synthesized in response to vascular shear stress and hypoxia. It has functions similar to those of EDNO.

Prostacyclin is derived from arachidonic acid and causes relaxation and platelet deactivation by increasing cAMP. It has been used to reduce pulmonary hypertension, particularly in pediatric patients.

Endothelins

Endothelins (ET) are elaborated by vascular endothelial cells in response to injury, thrombin, transforming growth factor- β (TGF- β), IL-1, angiotensin II, arginine vasopressin, catecholamines, and anoxia. Structurally formed from a 38–amino acid precursor molecule, ET is a 21–amino acid peptide with potent vasoconstrictor properties. Among the peptides in this family (ET-1, ET-2, ET-3), endothelial cells produce only ET-1. ET-1 is the most biologically active and potent vasoconstrictor known, estimated to be ten times more potent than angiotensin II. Three endothelin receptors, referred to as ET_A, ET_B, ET_C, function by the G protein–coupled receptor mechanism. ET_B receptors are linked to the formation of EDNO and PGI₂, which are negative feedback mechanisms. This may explain the transient vasodilation obtained with low-dose administration of ET-1 and the need for EDNO and ET to maintain physiologic tone in vascular smooth muscles. The vasoconstrictor activity of ET can be reversed by the administration of acetylcholine, which stimulates EDNO production. Increased serum levels of ET are correlated with the severity of injury after major trauma, major surgical procedures, and in cardiogenic or septic shock.

Platelet-Activating Factor

Another endothelium-derived product is PAF, a phospholipid constituent of cell membranes that can be induced by TNF, IL-1, AVP, and angiotensin II. This potent inflammatory mediator stimulates production of TxA₂ through the cyclooxygenase pathway and promotes platelet aggregation. TxA₂ is also a potent vasoconstrictor. Experimentally, PAF has increased glucagon and catecholamine activity. It can induce hypotension, increase vascular permeability, hemoconcentration, pulmonary hypertension, bronchoconstriction, primed PMN activity, eosinophil chemotaxis/degranulation, and thrombocytopenia. It induces a general leukocytopenia by way of margination. Administration of antagonists to PAF in experimental human endotoxemia demonstrates partial attenuation of symptoms such as myalgias and rigors, but these inhibitors are ineffective in reversing hemodynamic derangements.

PAF alters the shape of endothelial cells, causing them to contract and increase permeability. In cultured endothelial cells, cell contraction permits the passage of macromolecules, such as albumin, across cell junctions. PAF is a chemotactant for leukocyte adherence to the vascular wall and facilitates migration out of the vascular compartment. The disparity between PAF-induced vascular permeability and PAF-induced vasoconstriction is most likely the result of differential receptor types and affinity found in different vascular segments. Other cells that secrete PAF include macrophages, PMNs, basophils, mast cells, and eosinophils.

Atrial Natriuretic Peptides

Atrial natriuretic peptides (ANPs) are peptides released by the central nervous system and by specialized endothelium found in atrial tissues in response to wall tension. ANPs are potent inhibitors of aldosterone secretion and prevent reabsorption of sodium. In rats, the myocardial endothelium-derived nitric oxide (EDNO) inhibits the release of ANP, while ET-1 is a potent secretagogue of ANP. The role of ANP in human response to injury is unknown.

Intracellular Mediators

Heat-Shock Proteins

In addition to heat stimulation, stimuli such as hypoxia, trauma, heavy metals, local trauma, and hemorrhage induce the production of intracellular heat-shock proteins (HSPs).

These proteins are presumed to protect cells from the deleterious effects of traumatic stress. HSPs function intracellularly in the assembly, disassembly, stability, and transport of proteins. The classic example of HSP activity is the intracellular transport of steroid molecules. The formation of HSPs require gene induction by the heat-shock transcription factor (HSF). Gene expression occurs in parallel with hormonal activities of the hypothalamus-pituitary-adrenal axis. This response may be ACTH-sensitive, and the production may decline with age. Although HSPs are important intracellular effectors, their relevance in the human response to injury can only be inferred from animal data.

Reactive Oxygen Metabolites

Reactive oxygen metabolites (ROMs) are short-lived, highly reactive molecular oxygen species with an unpaired outer orbit. They cause tissue injury by peroxidation of cell membrane unsaturated fatty acids.

ROMs are produced by complex processes that involve anaerobic glucose oxidation coupled with the reduction of oxygen to superoxide anion. Superoxide anion is a potent ROM but can be metabolized to other reactive species such as hydrogen peroxide and hydroxyl radical. Cells are not immune to damage by their own ROMs but are generally protected by oxygen scavengers that include glutathione and catalases. In ischemic tissues, the intracellular mechanisms for production of ROMs become fully activated but are nonfunctional because of a lack of oxygen supply. With restoration of blood flow and oxygen supply, large quantities of ROMs are produced that induce reperfusion injury.

In response to a stimulus, activated leukocytes are potent generators of reactive oxygen metabolites. ROMs also can induce apoptosis. Studies using T lymphocytes have demonstrated a major apoptotic mechanism mediated by depletion of intracellular glutathione or ROM scavenger. The proapoptotic Fas/CD95 receptor activation is implicated in depleting GSH with resultant intracellular ROM accumulation and cell death. Repletion of GSH in these cells can reverse these effects.

Other Inflammatory Mediators

Eicosanoids

The eicosanoid class of mediators, which encompasses prostaglandins (PG), thromboxanes (Tx), leukotrienes (LT), hydroxyeicosatetraenoic acids (HETE), and lipoxins (Lx), are oxidation derivatives of the membrane phospholipid, arachidonic acid (eicosatetraenoic acid). They are secreted by virtually all nucleated cells except lymphocytes. The synthesis of arachidonic acid from phospholipids requires enzymatic activation of phospholipase A₂ (Fig. 1-19). The cyclooxygenase and the lipoxygenase pathways are two major routes by which arachidonic acid is oxygenated. Most eicosanoids generated from the cyclooxygenase pathway are given the subscript designation of 2 (e.g., TxA₂), while products of the lipoxygenase pathway are designated 4 (e.g., LTE₄). These subscripts indicate the number of carbon double bonds present in the side chains. Products of the cyclooxygenase pathway include all of the prostaglandins and thromboxanes. The formation of prostacyclin (PGI₂) requires further enzymatic activity by prostacyclin synthetase, and the formation of TxA₂ requires the activity of thromboxane synthetase. The lipoxygenase pathway generates the leukotrienes and HETE.

Initial phospholipase A₂ activation can be achieved by compounds such as epinephrine, angiotensin II, bradykinin, histamine, and thrombin. Conversely, phospholipase A₂ can be inhibited by lipocortin, which is induced by cortisol. The synthesis of prostaglandins and thromboxanes also are inhibited by nonsteroidal antiinflammatory drugs and salicylates, which are cyclooxygenase inhibitors.

Eicosanoids are not stored in cells but are synthesized rapidly upon stimulation by hypoxic and ischemic injury, direct tissue injury, endotoxin, norepinephrine, AVP, angiotensin II, bradykinin, serotonin, acetylcholine, and histamine. Many of these stimuli also induce a second cyclooxygenase enzyme, referred to as COX-2, that enhances the production of arachidonic acid metabolites. COX-2 activity can be inhibited by glucocorticoids, which provide specific inhibition of cyclooxygenase metabolites, as opposed to lipocortin, which inhibits production of arachidonic acid metabolites. The products of arachidonic acid metabolism are functionally cell/tissue specific. Vascular endothelium primarily synthesizes PGI₂, which causes vasodilation and platelet deactivation. Thromboxane synthetase converts platelet prostaglandins to TxA₂, a potent vasoconstrictor and platelet aggregator. Macrophages are capable of synthesizing cyclooxygenase and lipoxygenase products.

Second messengers mediate much of eicosanoid activity. For example, PGE compounds, in a manner similar to ACTH, TSH, and LH, inhibit AVP activity and hormone-stimulated lipolysis by activating adenylate cyclase activity and generating intracellular cAMP. Thromboxane and leukotrienes have opposite effects from PGE by increasing intracellular free calcium via the phosphatidylinositol pathway.

Eicosanoids have diverse effects systemically on endocrine and immune function, neurotransmission, and vasomotor regulation (Table 1-7). Eicosanoids are major components of the inflammatory response in injured tissue, characterized by vascular permeability, leukocyte migration, and vasodilation. Collectively, their deleterious effects are implicated in acute lung injury, pancreatitis, and renal failure. Leukotrienes are produced by cells of the lung, connective tissue, smooth muscle, macrophages, and mast cells that mediate the reactions characteristic of anaphylaxis. Leukotrienes are 1,000 times more potent than histamines in promoting capillary leakage. They also are effective promoters of leukocyte adherence, neutrophil activation, bronchoconstriction, and vasoconstriction. The role of lipoxins is not well understood but they are believed to induce neutrophil activation and production of superoxides and degranulation.

The metabolic effects of eicosanoids are well recognized. In the regulation of glucose, products of the cyclooxygenase pathway inhibit pancreatic beta cell release of insulin while products of the lipoxygenase pathway promote beta cell activity. Hepatocytes also express specific receptors for PGE₂ that, when activated, inhibit gluconeogenesis. PGE₂ inhibits hormone-stimulated lipolysis.

Eicosanoids modulate the immune response in multiple ways. Small amounts of PGE₂ suppress proliferation of human T lymphocytes by mitogens, an effect mediated by down-regulation of IL-2 production. Enhanced lymphocyte activation by mitogens can be achieved with the administration of indomethacin, a PGE₂ inhibitor. During phagocytosis, PMNs release eicosanoids such as LTB₄ to serve as chemoattractants for other leukocytes. PGE₂ and LTD₄ are commonly present in local areas of injury and are believed to have a direct influence on the inflammatory response.

Kallikrein-Kinin System

Bradykinins are potent vasodilators produced through kininogen degradation by the serine protease kallikrein. Kallikrein exists in blood and tissues as inactive prekallikrein and is activated by various chemical and physical factors. Among these are Hageman factor, trypsin, plasmin, factor XI, glass surfaces, kaolin, and collagen. Kinins are rapidly metabolized by kinase I and II. Kinase I degrades the anaphylatoxins C3a, C4a, and C5a. Kinase II is identical to angiotensin-converting enzyme. The use of angiotensin-converting enzyme inhibitors (ACE inhibitors) in controlling hypertension may serve partially to block kinin degradation in some patients and enhance the kinin-induced injurious effects on the bronchial tree.

Kinins increase capillary permeability and tissue edema, evoke pain, and increase bronchoconstriction. They also increase renal vasodilation and consequently reduce renal blood flow. The resulting increase in renin formation activates sodium and water retention via the renin-angiotensin system.

Bradykinin release is stimulated by hypoxic and ischemic injury. Increased kallikrein activity and bradykinin levels have been detected after hemorrhage, sepsis, endotoxemia, and tissue injury. These observations are positively correlated with the magnitude of injury and mortality. Clinical trials using bradykinin antagonists in attempts to reduce the deleterious sequelae of septic shock have demonstrated only modest reversal in gram- negative sepsis and no overall improvement in survival. Metabolically, kinins increase glucose clearance by inhibiting gluconeogenesis. Bradykinin infusion also may increase nitrogen retention.

Serotonin

The neurotransmitter serotonin (5-hydroxytryptamine, 5-HT) is a tryptophan derivative that is found in enterochromaffin cells of the intestine and in platelets. Patients with midgut carcinoid tumors often secrete excessive 5-HT. This neurotransmitter stimulates vasoconstriction, bronchoconstriction, and platelet aggregation. It also is capable of acting as a myocardial chronotrope and inotrope. Although it is released at sites of injury, its role in the injury response is unclear.

Histamine

Histamine is derived from histidine and stored in neurons, skin, gastric mucosa, mast cells, basophils, and platelets. Its release is activated by increased calcium levels. There are two receptor types for histamine binding. H₁ binding mediates increased histamine precursor uptake, I- histidine, and stimulates bronchoconstriction, intestinal motility, and myocardial contractility. H₂ binding inhibits histamine release. H₁ and H₂ receptor activation induces vasodilation and increases vascular permeability. Histamine administration causes hypotension, peripheral pooling of blood, increased capillary permeability, decreased venous return, and myocardial failure. Histamine is released in hemorrhagic shock, trauma, thermal injury, endotoxemia, and sepsis. Histamine levels are correlated with mortality from septic shock.

METABOLIC RESPONSE TO INJURY

The description of human biochemical responses to injury and the classification of such responses into an ebb and flow phase by Cuthbertson and others provides a useful model by which the metabolic response to injury may be characterized (Fig. 1-20). The ebb phase corresponds to the earliest moments to hours after injury, often in association with hemodynamic instability or reductions in effective circulating blood volume. The metabolic consequences of this phase are less well studied but are generally associated with reductions in total body energy expenditure and losses of urinary nitrogen. The ebb phase is characterized by an early enhancement of neuroendocrine hormone appearance, including catecholamines and cortisol. Less is known about the microendocrine mediator response. It is difficult to analyze the immune cell mediator response engendered during the ebb phase separately from that occurring in response to fluid or volume resuscitation, and the resultant tissue reperfusion and reoxygenation that initiates the onset of the flow phase.

Except with the most minor injury, the flow phase is ushered in by compensatory mechanisms resulting from volume repletion and cessation of initial injury conditions. The metabolic response associated with the flow phase serves to direct energy and protein substrates so as to preserve critical organ function and repair damaged tissues. This includes an increase in whole-body oxygen consumption and metabolic rate, enhancement of critical enzyme pathways for readily oxidizable substrates such as glucose, and stimulation of immune system functions required for repair of tissue destruction and protection from additional breaks in epithelial barriers. A reprioritization of substrate processing occurs to support the production of acute-phase reactants, immunoreactive proteins, and coagulation factors. The biologic priority of wound healing also is established during the early flow phase.

Metabolic Response to Fasting

A comparison between the metabolic physiology of injury to that of unstressed fasting is useful for assessing the relative magnitudes of altered physiology under these widely varying conditions. Factors such as antecedent health status, age, and lean body mass also influence the absolute rates of substrate turnover after fasting and injury.

Substrate Metabolism

A healthy adult of 70 kg body weight expends 1700 to 1800 kcal/day of energy obtained from lipid, carbohydrate, and protein sources (Fig. 1-21). Obligate glycolytic cells, such as neurons, leukocytes, and erythrocytes, require 180 g glucose per 24 h for basal energy needs. During acute starvation, glucose is derived from existing storage pools, including approximately 75 g glucose stored as hepatic glycogen. Skeletal muscle cannot directly release free glucose, because it lacks the glucose-6- phosphatase necessary for free glucose release. The reduction of circulating glucose during prolonged fasting serves as a primary stimulus to hormonal release that modulates gluconeogenesis and substrate substitution for those tissues that require glucose for energy. Glucose concentration falls within hours after the onset of fasting in association with decreases in insulin release and sustained increases of circulating glucagon and more transient elevations of GH, catecholamines, AVP, and angiotensin II. Glucagon and epinephrine enhance cAMP to promote glycogenolysis, and cortisol and glucagon promote gluconeogenesis. The actions of norepinephrine, AVP, and angiotensin II are mediated by the intracellular signals of phosphatidylinositol and calcium to promote glycogenolysis. Cortisol and epinephrine limit pyruvate use. The effect of these actions is an increase in glucose production. Sustained glucose production depends on presentation of amino acids, glycerol, and fatty acids to the liver (Figs. 1-22 and 1-23). The primary gluconeogenic precursors used by the liver and to a lesser extent by the kidney for gluconeogenesis are lactate, glycerol, and amino acids such as alanine and glutamine. Skeletal muscle releases lactate by breakdown of endogenous glycogen stores and by glycolysis of transported glucose. Lactate also is released by erythrocytes and white blood cells after aerobic glycolysis and release of newly formed lactate into the circulation. This lactate is reconverted to glucose in the liver by the Cori cycle (Fig. 1-24).

The quantity of glucose made from lactate produced by skeletal muscle is not sufficient to maintain glucose homeostasis. Consequently, approximately 75 g of protein must be degraded daily during fasting and starvation to provide gluconeogenic amino acids to the liver. Proteolysis, which results primarily from decreased insulin and increased cortisol, is associated with an increase in urinary nitrogen excretion from the normal 6 to 8 g/day to approximately 8 to 11 g within the initial 5 days of fasting. Protein mobilized in starvation is derived primarily from skeletal muscle, but the loss of protein from other organs also occurs.

The amino nitrogen load resulting from deamination of amino acids for gluconeogenesis increases urinary ammonia excretion. The renal excretion of ammonium ion becomes the primary route of elimination of alpha-amino nitrogen during starvation because the normally active hepatic enzymes are diminished. Renal gluconeogenesis increases through metabolism of glutamine and glutamate. The kidney may account for up to 45 percent of glucose production during late starvation.

After approximately 5 days, the rate of whole-body proteolysis diminishes to a level of 15 to 20 g/day and urinary nitrogen excretion stabilizes at 2 to 5 g/day for several weeks. This reduction in proteolysis occurs as the nervous system and other previous glucose-utilizing tissues adapt to ketone oxidation as the predominant energy source. During starvation, transport systems in the blood-brain barrier increase the rate of ketone body transport, and metabolism of ketone bodies by the brain increases. Consequently, the amount of protein required for gluconeogenesis is significantly reduced (see Fig. 1-23). A reduction in anabolic growth factors such as IGF-1 (formerly somatomedin C) also is observed during the first several days of fasting (Fig. 1-25). The reduction in this factor and its associated binding proteins reduces an important signal for transcellular amino acid transport. Tissue protein synthesis also falls correspondingly with reduced proteolysis.

Energy requirements for gluconeogenesis and basal enzymatic and muscular function, such as neural transmission and cardiac contraction, can be met by the mobilization of approximately 160 g of triglycerides from adipose tissue in the form of free fatty acids and glycerol in a resting, fasting 70-kg subject (Fig. 1-26). Free fatty acid release is stimulated by a reduction in the serum insulin concentration. Increased glucagon may participate in this alteration, as do catecholamines. The free fatty acids and ketone bodies generated by the liver are used as a source of energy by tissues such as the heart, kidney, muscle, and liver. Lipid stores provide up to 40 percent of the caloric expenditure during starvation. Lipid oxidation during starvation diminishes the absolute glucose requirement to sustain tissue and body energy expenditure. Fatty acid use occurs at a rate that is proportional to serum fatty acid concentration. Ketone bodies spare glucose by inhibition of pyruvate dehydrogenase in most tissues. The use of fat as a main fuel source decreases the amount of mandatory glycolysis, which diminishes the requirements for gluconeogenesis and protein degradation.

Once the initial obligatory neuroendocrine stress hormone response recedes, whole-body energy expenditure also decreases during prolonged fasting. This reduction in resting energy expenditure is a consequence of decreased sympathetic nervous system activity and reduced skeletal muscle activity. Reduced secretory enzyme production and intestinal metabolic demands contribute to this adjustment.

Metabolism After Injury

The metabolic consequences of injury differ in many fundamental ways from those of simple starvation. Well-defined changes in hormone levels and associated substrates accompany injury. These changes can reflect the degree of underlying injury (Fig. 1-27). A useful construct in the attendant changes in interorgan substrate flux has been proposed by Cahill (Fig. 1-28). It is the sustained activities of macroendocrine hormones in conjunction with immune cell activation that provides the signals that differentiate injury metabolism from unstressed starvation.

Energy Balance

Injury of any magnitude beyond the most trivial is associated with an increase in energy expenditure and increases in oxygen consumption that vary directly with the severity of injury (Fig. 1-29) or burn surface area (Fig. 1-30). A linear relationship between body cell mass and resting energy expenditure often is observed in injured patients. The increase in energy expenditure observed after injury results initially from the increased activity of the sympathetic nervous system and increased circulating concentrations of catecholamines. Increases in energy expenditure can be replicated by the administration of catecholamines to healthy subjects. Conversely, glucocorticoid excess does not significantly enhance energy expenditure. The mechanism for this catecholamine effect may be related to influences on cell membrane sodium permeability and the energy required for ion pump action to maintain normal transmembrane concentrations. This influence is observed during the endotoxin/cytokine interactivity, in which reductions in muscle resting transmembrane potential are readily observed. Under these conditions, increases in circulating catecholamines occur with restoration of membrane potential and before changes in energy expenditure. Young estimated that such cellular ion pumping and transport activities may account for over 40 percent of total body energy expenditure.

Lipid Metabolism

Free fatty acids are a principal source of energy after injury. Lipolysis is enhanced by the immediate elevations in ACTH, cortisol, catecholamines, glucagon, and growth hormone levels, reduction in insulin level, and increased sympathetic nervous system activity. Catecholamines are the chief stimulus to hormone-sensitive lipase. The sympathetic nervous system and circulating catecholamines are important in the lipolytic response to stress. Investigations into the mechanism of catecholamine- induced posttraumatic lipolysis suggest that lipolysis may be increased by changes in adrenergic postreceptor (protein kinase hormone-sensitive lipase) response after elective operation.

Lipolysis observed during the ebb phase results in elevated levels of plasma free fatty acids and glycerol. Increased reesterification of fatty acids, such as that seen in the presence of high concentrations of lactate, may decrease net free fatty acid release. This explanation is supported by the observed rise in plasma glycerol level noted after injury, which suggests that lipolysis is occurring, and by increased concentrations of lactate in studies in which there is no change in free fatty acid concentration. Acidosis, hyperglycemia, and anesthetic agents also alter lipid mobilization after injury. For example, lipolysis is directly inhibited by pentobarbital anesthesia, and hemorrhage in the presence of pentobarbital usually results in a fall in the plasma free fatty acid and ketone body concentrations. Experimental hemorrhage using other anesthetic agents or in awake animals increases free fatty acid and ketone body concentrations.

During the flow phase, net lipolysis continues, as reflected by increased concentrations and clearance of plasma free fatty acids. In the presence of oxygen, the released fatty acids can be oxidized by cardiac and skeletal muscle to produce energy.

The precise role of fatty acids in the inhibition of glycolysis after injury is controversial. Evidence suggests that fatty acid-induced inhibition of glycolysis may be a major mechanism for reduced glycolysis during the flow phase after minor to moderate injury. This mechanism may not operate in severe injury, hemorrhage, or sepsis, conditions in which persistent glycolysis and net proteolysis are observed. Lipoprotein lipase, the endothelial cell membrane enzyme responsible for clearing plasma triglycerides, is suppressed in adipose tissue after trauma, but not in muscle. In sepsis this enzyme activity is suppressed in both muscle and adipose tissue. The roles of cytokines, such as TNF (which inhibits lipogenesis and decreases lipoprotein lipase activity), IL-1, and PGE, in fat metabolism are not fully understood.

The high concentrations of intracellular fatty acids and the elevated concentration of glucagon during the ebb and flow phases inhibit fatty acid synthesis. In hepatocytes, this also stimulates the transport of acyl coenzyme A (acyl CoA) into the mitochondria for oxidation and ketogenesis. Ketogenesis is variable and is inversely correlated with the severity of injury. Ketogenesis is decreased after major injury, severe shock, and sepsis and is suppressed by increases in levels of insulin and other energy substrates, by increased uptake and oxidation of free fatty acids, and by an associated counterregulatory hormone response. After minor injury or mild infection, ketogenesis increases but to a lesser extent than that seen during nonstressed starvation. Injuries that are associated with minor ketone body formation also appear to be associated with a small or absent increase in plasma free fatty acid concentrations.

Carbohydrate Metabolism

Systemic glucose intolerance is well documented in injured patients. By contrast, basal insulin levels are elevated by several times during the early flow phase, indicating a state of relative insulin resistance. Regional tissue catheterization and isotope dilution studies provide a more precise description of this insulin resistance. A 50 to 60 percent increase in net splanchnic glucose output is observed in septic patients, and a 50 to 100 percent increase is noted in thermally injured patients. The associated macroendocrine hormone milieu contributes to this net gluconeogenic response and is believed to be largely under the active control of glucagon with permissive requirement for cortisol. The precise contributions of other macroendocrine hormones are unclear, although there is evidence to suggest that proinflammatory mediators such as IL-6 also may exert an influence on hepatic glucose production. Definable acute changes in substrate turnover are associated with the proinflammatory mediator activity induced by endotoxin administration or TNF infusion (Table 1-8).

Increases in plasma glucose levels are proportional to the severity of injury and to some extent are correlated with survival. With the presence of hyperglycemia, resulting largely from increased hepatic production, a ready source of substrate is provided to tissues such as those of the nervous system, wound, and red blood cells, which do not require insulin for glucose transport. Elevated concentrations of glucose and of some amino acids may be necessary for leukocyte energy requirements in inflamed tissues and in defense of epithelial barriers or other sites of microbial invasion. Insulin resistance is of teleologic benefit to the host in that the accompanying neuroendocrine hormone response precludes the adaptation to ketone body production. To a large extent, the deprivation of glucose to nonessential organs such as skeletal muscle and adipose tissues is mediated by catecholamines. This relationship is suggested by a close correlation between plasma glucose and prevailing catecholamine levels (Fig. 1-31). Peripheral insulin resistance has been noted in isolated catecholamine excess in healthy subjects, as has been increased hepatic glucose production. Conversely, glucocorticoids do not alter these parameters (Fig. 1-32). Although the mechanisms for reduced glucose oxidation are not fully understood, a mediator-induced reduction of skeletal muscle pyruvate dehydrogenase activity diminishes the conversion of glucose to acetyl CoA and subsequent entry into the tricarboxylic acid cycle. The consequent accumulation and shunting of three carbon skeletons to the liver provides substrate for gluconeogenesis.

Glucose must be provided to inflammatory and healing cells in the wound environment. Glucose uptake and lactate production in wounded tissue are significantly increased. Wound inflammatory cells require glucose as an energy substrate and accelerated glucose uptake in wounded and burned tissue is correlated with the inflammatory cellular infiltrate.

The increase in glucose uptake in wounded tissue is associated with an increase in the activity of phosphofructokinase, a major rate-limiting enzyme in glycolysis. Despite the increase in glucose uptake and phosphofructokinase activity, wounded and burned tissues demonstrate decreased insulin sensitivity and fail to normally increase glucose uptake or glycogenesis in response to insulin.

Protein and Amino Acid Metabolism

The intake of protein for a healthy young adult is approximately 80 to 120 g, or 13 to 20 g of nitrogen per day. Daily fecal and urinary excretion of nitrogen is 2 to 3 g and 13 to 20 g, respectively. After injury, daily nitrogen excretion in the urine increases to 30 to 50 g as urea nitrogen and represents net proteolysis. The increased excretion of urea after injury also is associated with the urinary loss of sulfur, phosphorus, potassium, magnesium, and creatinine, which indicates breakdown of intracellular compounds. Isotope dilution studies suggest that decreases in cell mass are responsible for the increased loss of these metabolites. Sophisticated methods of body mass assessment, such as neutron activation analysis, confirm a loss of lean tissues after significant injury. A predominant loss of skeletal muscle protein is suggested by increased urinary 3- methylhistidine/creatinine ratios. Radiolabeled amino acid incorporation studies and protein analyses confirm that skeletal muscle is depleted while visceral tissues, such as liver and kidney, are relatively preserved. The mechanisms for this visceral protein preservation are unclear, but animal studies suggest that proinflammatory cytokine activity may contribute to this process.

Data on total body protein turnover suggest that after injury the net changes in catabolism and synthesis depend on the severity of the injury. Elective operations and minor injuries result in decreased protein synthesis and normal rates of protein breakdown. Severe trauma, burns, and sepsis are associated with increased whole-body protein turnover and increased net protein catabolism (Table 1-9). Accelerated proteolysis and gluconeogenesis persist after major injury and during sepsis. The rise in urinary nitrogen and negative nitrogen balance begins shortly after injury, reaches a peak about the first week and may continue for 3 to 7 weeks. The magnitude of nitrogen loss also is related to the age, sex, and physical condition of the patient. Young healthy males lose more protein in response to an injury than do women or the elderly, presumably because they have a higher lean body mass than the latter two patient subsets.

The amino acid composition of normal human beings varies according to tissue origin. After trauma, substrate cycling occurs between skeletal muscle, liver, and the wound (see Fig. 1-28). Quantitatively, the major source of amino acids is skeletal muscle, in which the proportions of specific amino acids as protein and free intracellular components varies dramatically from that in normal plasma (Fig. 1-33). Increases by several times in the splanchnic uptake of alanine and glutamine in conjunction with similar trends for peripheral tissue efflux are observed after injury. Although the precise mechanisms for the net increase in skeletal muscle protein breakdown remain unclear, the combined extracellular hormonal milieu of relative insulin resistance, cortisol excess, and proinflammatory cytokine activity exert a synergistic influence. Within the cell, enhanced oxidative species and diminished antioxidant activities, such as glutathione, all enhance the potential for protein instability. Many of these same mediators serve to increase ubiquitin-dependent proteolytic pathways (Fig. 1-34), one of several potential pathways for degradation of cellular proteins. The ubiquitin-proteasome pathway may not be operative in all catabolic conditions. For example, this pathway is of major importance for muscle protein degradation during sepsis, but it is only partly responsible for the protein catabolism seen after burn injury.

The intracellular muscle concentrations of several essential amino acids decrease at the same time that net efflux is occurring from skeletal muscle. The release of glutamine and alanine are greater than can be predicted from its relative abundance in muscle tissue protein, indicating their net synthesis in muscle before release. Glutamine is a major energy source for lymphocytes, fibroblasts, and the gastrointestinal tract, especially during conditions of increased stress. Glutamine may act as a conditional essential amino acid during periods of catabolism, since depletion of this substrate has pronounced negative effects on enterocytes and mucosal integrity and since administration of glutamine reverses these effects. Although it is hypothesized that provision of this amino acid might preserve or enhance immune cell and enterocyte function during stress, clinical studies documenting such beneficial effects during injury in human subjects are lacking.

NUTRITION IN THE SURGICAL PATIENT

Most patients undergoing elective surgical operations withstand the brief period of catabolism and starvation without noticeable difficulty. Maintaining an adequate nutritional regimen may be of critical importance in managing seriously ill surgical patients with preexisting weight loss and depleted energy reserves. Between these two extremes are patients for whom nutritional support is not essential for life but may serve to shorten the postoperative recovery phase and minimize the number of complications. Frequently, a patient may become ill or even die from complications secondary to starvation rather than the underlying disorder. It is essential that the surgeon have a sound grasp of the fundamental metabolic changes associated with surgery, trauma, and sepsis and an awareness of the methods available to reverse or ameliorate these events.

Surgery, Trauma, Sepsis

In contrast to the whole-body and tissue-specific energy and protein conservation response exhibited during unstressed starvation, the injured patient manifests variable, but obligatory, increases in energy expenditure and nitrogen excretion (Fig. 1-35). While the extent and duration of this response to injury are modified by a variety of factors, including the adequacy of resuscitation, infection, and medication, the inability to down- regulate body energy expenditure and nitrogen losses may rapidly deplete labile and functional energy stores. The postinjury metabolic environment precludes the efficient oxidation of fat and ketone production, thereby promoting the continued erosion of protein pools. This enhanced net protein catabolic process, if unchecked by effective disease-specific therapy and allowed to progress for an extended period without nutritional intervention, eventuates in critical organ failure.

The sequence of flow phase metabolic and endocrine events occasioned by injury may be divided into several phases. The magnitude of the changes and the duration of each phase vary considerably and are directly related to the severity of the injury. The benefits of exogenous nutritional support in each of these recovery phases is controversial. Nevertheless, the integrative biology of endocrine and immune system interactions would predict a greater likelihood of achieving lean tissue restoration and substrate- dependent immune competence during periods of attenuated mediator activity. Such observations do not preclude a rationale for earlier efforts at nutritional intervention. Rather, they provide a biologic basis for reasonable expectations of therapy and for the design of future therapeutic adjuncts.

Catabolic Phase

Once patients have received initial resuscitation and stabilization of wounds, the earliest definable metabolic response is one of catabolism. This phase has been termed the *adrenergic-corticoid phase*, because it corresponds to the period during which changes induced by adrenergic and adrenal corticoid hormones are most striking. It is also likely that components of the micromediator systems exert significant influences during this phase. The detection of proinflammatory mediators or their surrogate markers (such as soluble receptors) usually is noted to peak during this period, and the maximum changes in substrate turnover are observed during this period (see Table 1-9). To variable degrees, rates of gluconeogenesis, acute phase protein production, and immune cell activity are all altered during the catabolic phase. The administration of moderate amounts of glucose to these individuals produces little or no change in the rate of protein catabolism, although evidence using isotopic determinations suggests that provision of sufficient nonprotein calories in combination with amino acids may reduce the rate of body protein breakdown (Table 1-10).

In the catabolic phase, glucose turnover is increased, while Cori cycle activity is stimulated and three-carbon intermediates are converted back to glucose in the liver by pyruvate carboxylase and phosphoenolpyruvate carboxylase. Increased synthesis of these two enzymes occurs in the presence of elevated levels of glucagon, glucocorticoids, and catecholamines and low concentration of insulin—the hormonal environment present during the catabolic phase of injury. Lipolysis also is stimulated by this hormonal milieu, and an obligatory oxidation of fatty acids is evident.

Efforts directed at interruption of afferent neurogenic stimuli by extradural anesthesia have met with partial success in attenuating some of these abnormalities of energy substrate turnover. The impact of such therapy on nitrogen loss has been far less dramatic, suggesting that circulating or tissue paracrine factors other than classical neuroendocrine hormones are of major importance in early postinjury metabolic responses. It has been widely speculated that one or more proinflammatory cytokine mediators are major determinants of increased rates of energy expenditure and substrate turnover after injury, but no clear evidence of this influence in human beings has been presented. Blockade of TNF and IL-1 activities during conditions of human endotoxemia does not prevent the characteristic increase in metabolic rate and glucose or protein turnover.

Early Anabolic Phase

Depending on the severity of injury, the body turns from a catabolic to an anabolic phase. This may occur within 3 to 8 days after uncomplicated elective surgery or after weeks in patients with extensive cross-sectional tissue injury, sepsis, or ungrafted thermal injury. This turning point, also known as the *corticoid-withdrawal phase*, is characterized by a sharp decline in nitrogen excretion and restoration of appropriate potassium- nitrogen balance. This phase is also biochemically characterized by a reprioritization of acute phase reactants, as early inflammatory response proteins are supplanted by tissue repair and anabolic factors, such as IGF- 1. Clinical manifestations of this transition period are brief and coincide with initial diuresis of retained water and renewed interest in oral nutrition.

The early anabolic phase may last from a few weeks to a few months depending on the capacity to ingest adequate nutrition and the extent to which erosion of protein stores has occurred. Nitrogen balance is positive, indicating synthesis of proteins, and there is a rapid and progressive gain in weight and muscular strength. Positive nitrogen balance reaches

a maximum of approximately 4 g/day, which represents the synthesis of approximately 25 g of protein and the gain of over 100 g of lean body mass/day. The total amount of nitrogen gain ultimately equals the amount lost during the catabolic phase, although the rate of gain will be much slower than the rate of initial loss.

Late Anabolic Phase

The final period of convalescence or the late anabolic phase may last from several weeks to several months after a severe injury. This phase is associated with the gradual restoration of adipose stores as the previously positive nitrogen balance declines toward normal. Weight gain is much slower during this phase because of the higher caloric content of fat and can be realized only if intake is in excess of caloric expenditure. In most individuals, the phase ends with a gradual return to the previously normal body weight. The patient who is partially immobilized during this period of time, however, may exhibit a marked gain in weight as a result of decreased energy expenditure.

Assessment and Requirements

Nutritional homeostasis assumes that proper timing and administration of nutrients has a favorable impact on the outcome of therapy. Nutritional assessment is undertaken to determine the severity of nutrient deficiencies or excesses and to aid in predicting nutritional requirements (Fig. 1-36). Important information is obtained by determining the presence of weight loss and of chronic illnesses or dietary habits influencing the quantity and quality of food intake. Social habits predisposing to malnutrition and the use of medications that may influence food intake or urination should be investigated. Physical examination seeks to assess loss of muscle and adipose tissues, organ dysfunction, and subtle change in skin, hair, or neuromuscular function reflecting an impending nutritional deficiency. Anthropometric data (weight change, skin fold thickness, and arm circumference muscle area) and biochemical determinations (levels of creatinine excretion, albumin, and transferrin) can be used to substantiate the patient's history and physical findings. It is imprecise to rely on any single or fixed combination of these findings to assess nutritional status or morbidity. Appreciation for the stresses and natural history of the disease process, in combination with nutritional assessment, is the basis for identifying patients in acute or anticipated need of nutritional support.

The caloric and nitrogen requirements necessary to maintain an individual in balance after severe injury depend on the extent of injury, the source and route of administered nutrients, and, to some extent, the degree of antecedent malnutrition. A fundamental goal of nutritional support is to meet the energy requirements for metabolic processes, core temperature maintenance, and tissue repair. Failure to provide adequate nonprotein energy sources leads to dissolution of lean tissue stores. The requirements for energy may be measured by indirect calorimetry or estimated from urinary nitrogen excretion, which is proportional to resting energy expenditure. Basal energy expenditure (BEE) also can be estimated by the equations of Harris and Benedict:

BEE (men)

$$= 66.47 + 13.75(W) + 5.0(H) - 6.76(A) \text{ kcal/day}$$

BEE (women)

$$= 655.1 + 9.56(W) + 1.85(H) - 4.68(A) \text{ kcal/day}$$

where W = weight, kg

H = height, cm

A = age, years

These equations are suitable for estimating energy requirements in at least 80 percent of hospitalized patients. Nonprotein calories are supplied in excess of energy expenditure because the use of exogenous nutrients is decreased and energy substrate demands are increased after traumatic or septic insult. Appropriate nonprotein caloric needs are 1.2 to 1.5 times resting energy expenditure (REE) during enteral nutrition and 1.5 to 2.0 times REE during intravenous nutrition. It is seldom, if ever, appropriate to exceed this level of nonprotein energy intake during the height of the catabolic phase. Some authorities have suggested that targeted underfeeding might be a more appropriate strategy in such patients.

The second objective of nutritional support is to meet the substrate requirements for protein synthesis. Maintenance of protein synthesis depends on many factors, including the nature and degree of the insult, the source and amount of exogenous protein, and previous nutritional status. Consequently, no single nutritional formulation is appropriate for all patients. An appropriate calorie-nitrogen ratio (150 to 200:1) should be maintained, but evidence suggests that increased protein intake (and a lower calorie-nitrogen ratio) may be efficient in selected hypermetabolic patients. In the absence of severe renal or hepatic dysfunction precluding the use of standard nutritional regimens, approximately 0.25 to 0.35 g of nitrogen/kg of body weight should be provided daily. Specialized nutritional formulations designed to improve nitrogen use in organ dysfunction such as acute renal and hepatic failure are targeted either to supplement deficiencies associated with the disease process or to correct characteristic amino acid abnormalities.

The requirements for vitamins and essential trace minerals can be easily met in the typical patient with an uncomplicated postoperative course. Vitamins usually are not given in the absence of preoperative deficiencies. Patients maintained on elemental diets or parenteral hyperalimentation require complete vitamin and mineral supplementation. The commercial defined-formula enteral diets contain varying amounts of essential minerals and vitamins (Table 1-11). It is necessary to ensure that adequate replacement is available in the diet or by supplementation. Numerous commercial vitamin preparations are available for intravenous or intramuscular use, although most do not contain vitamin K and some do not contain vitamin B₁₂ or folic acid. Supplemental trace minerals may be given intravenously. Essential fatty acid supplementation also may be necessary, especially in patients with depletion of adipose stores. Patients receiving intravenous feeding require all the above micronutrients to prevent the development of deficiencies.

Indications and Methods for Nutritional Support

The selection of patients who require partial or complete nutritional support has become increasingly important as constraints on hospitalization and resource management escalate. The ability to provide nutritional support to stressed patients and to attenuate nitrogen losses in catabolic states is an important adjunct to surgical care. The need for nutritional support should be assessed during the preoperative and postoperative courses of all but the most routine cases. Most surgical patients, however, do not require special nutritional regimens. The reasonably well-nourished and otherwise healthy individual who undergoes an uncomplicated major operative procedure has sufficient body fuel reserves to withstand the catabolic insult and partial starvation for at least 1 week. Adequate quantities of parenteral fluids with appropriate electrolyte composition and a minimum of 100 g glucose daily to minimize protein catabolism will be all that is necessary in most patients. Assuming that the patient has a relatively uncomplicated postoperative course and resumes normal oral intake at the end of this period, defined-formula diets or parenteral alimentation are unnecessary and inadvisable because of the associated risks. During the early anabolic phase, the patient needs an adequate caloric intake of proper composition to meet the energy needs of the body and to allow protein synthesis. A high calorie-to-nitrogen ratio (optimally approximately 150 kcal/g nitrogen) and an adequate supply of vitamins and minerals are necessary for maximum anabolism during this period. In contrast to this group, there are populations of surgical patients for whom an adequate nutritional regimen can be of critical importance for a successful outcome. These include some patients who are chronically debilitated preoperatively from their diseases or from malnutrition and patients who have suffered trauma, sepsis, or surgical complications and cannot maintain an adequate caloric intake. In many cases the need for nutritional therapy during the early catabolic phase is apparent. This most certainly includes patients for whom there is a high expectation of prolonged hospitalization and diminished capacity for voluntary nutrient intake, such as patients with extensive burns or those with other severe injuries and incipient or overt organ failure. Despite the intuitively obvious decision to initiate nutritional support in such populations, documentation of nutrition-specific benefits or improvements in outcome are generally lacking. Nevertheless, such highly stressed and at-risk patients should receive consideration of nutritional support early. The dilemma more commonly presented to the clinician is the identification of other patients in whom a reasonable expectation of benefit from nutritional intervention can be met. Prospective, randomized trials have significantly narrowed the populations in whom this expectation might be met. In general, the indications for preoperative nutritional support, at least in hospitalized patients, appear largely confined to patients with evidence of more severe erosion of lean body mass and adipose tissue stores. This does not preclude the possibility that nutritional support in the ambulatory setting or those with evidence of organ failure or immunosuppression might achieve benefit from specialized nutritional support before elective surgery.

Specialized nutritional support can be given enterally, or enterally with supplements via peripheral vein, or by central venous routes. The enteral route should always be used when possible because it is considered to be more economical and well tolerated in many patients, including those having undergone recent abdominal surgery. Nasopharyngeal, gastrostomy, and jejunostomy tube feedings may be considered for alimentation in patients who have a relatively normal gastrointestinal tract but cannot or will not eat. Elemental diets may be administered by similar routes when bulk and fat-free nutrients requiring minimal digestion are indicated. Parenteral alimentation may be used for supplementation in the patient with limited oral intake or, more commonly, for complete nutritional management in the absence of oral intake. Clinical studies demonstrate that parenteral feeding potentially enhances the magnitude of macroendocrine (stress hormones) and microendocrine (cytokine) mediator responses to an antigenic challenge (Fig. 1-37). While the mechanisms for amplification of counterregulatory hormone and proinflammatory mediator levels in parenterally fed subjects remain to be fully elucidated, a loss of intestinal barrier function permitting acute or chronic host exposure to luminal toxins has been proposed. In human beings, it has not been clearly determined whether parenteral nutrition significantly alters intestinal barrier function instead of intracellular and intercellular anatomy. The incidence of systemic immune compromise resulting from parenteral nutrition has proved difficult to document. While several studies suggest a higher incidence of infectious complications in parenterally fed subjects compared with an enterally fed cohort, this observation is largely confined to traumatically injured populations (Table 1-12).

Despite the failure to document clinical differences between the enteral and parenteral feeding routes for exogenous nutrients, the gastrointestinal tract serves a number of synthetic and immunologic functions that bear consideration in the design of nutritional support regimens. A number of approaches for preserving gastrointestinal mucosal integrity and gut mass, including luminal stimulation by digestible or nondigestible substrates, and infusion of critical intestinal fuel sources such as glutamine or short-chain fatty acids, are undergoing clinical trials. To date, these products have not been clearly documented to improve outcome in the majority of populations studied.

The patient's ability to tolerate and absorb enteral feedings is determined by the rate of infusion, the osmolality, and the chemical nature of the product. Enteral feedings often are begun at a rate of 30 to 50 mL/h and are increased by 10 to 25 mL/h per day until the optimal volume is delivered. After full volume is attained, the concentration of the solution is increased slowly to the desired strength. If esophageal or gastric feedings are given, residual gastric volume should be monitored to reduce the risk of a major aspiration episode. If abdominal cramping or diarrhea occurs, the rate of administration or the concentration of the solution should be decreased. All feeding tubes should be thoroughly irrigated clear of solutions if feedings are interrupted or medications are given by this route.

Enteral Feeding Nasoenteric Tube Feeding

The use of feeding tubes that traverse the gastroesophageal junction should normally be used only in alert patients. While some exceptions might transiently exist, prolonged use of such feeding tubes should be discouraged. The foremost contraindication for nasoesophageal or gastric tube feeding is unconsciousness or lack of protective laryngeal reflexes, which may result in life-threatening pulmonary complications from aspiration. Even with a tracheostomy, it is inadvisable to feed mentally obtunded patients via such route, since feedings often can be recovered from tracheostomy suction, indicating continued aspiration of gastric contents. Esophageal feedings are seldom permissible.

The nasojejunal tube may allow feeding beyond dysfunctional gastric stomas and high gastrointestinal fistulas. In such cases it may be possible to maintain nutrition without a jejunostomy tube until stomal dysfunction relents or the fistula heals. Such tubes may be positioned in the upper small intestine by positioning the patient in a manner that promotes passage of the mercury-weighted tube into the desired intestinal segment. If this technique is unsuccessful, placement may be accomplished with fluoroscopic guidance or by an experienced endoscopist. Proper position of the tube must be confirmed radiographically.

Whenever dietary preparations are administered into the gastrointestinal tract via tubes, it is advisable to use bedside infusion pumps to ensure a constant rate of delivery over each 24-h period. The use of such pumps decreases the incidence of gastrointestinal side effects induced by overly rapid delivery of hyperosmolar solutions, and it allows safer administration of larger daily volumes of nutrients, because gastric distention is minimized. Investigation is required for all abdominal complaints in such patients in view of reports of intussusception around feeding tubes placed more distally in the small intestine.

Gastrostomy Tube Feeding

The administration of blended food through a gastrostomy tube is a good method for feeding patients with a variety of chronic gastrointestinal lesions arising at or above the cardioesophageal junction. Gastrostomy tube feeding is contraindicated for mentally obtunded patients with inadequate laryngeal reflexes. This feeding method should be used only in alert patients or in patients with total obstruction of the distal esophagus.

Surgically constructed gastrostomies of the Stamm (serosa-lined, temporary) or the modified Glassman (mucosa-lined, permanent) type are an acceptable means of providing enteral feedings, although percutaneous endoscopic gastrostomies (PEGs) have proved to be safe and equally effective. The feeding mixture may be ordinarily prepared food converted by a blender into a semiliquid consistency. Hyperosmolality of the feeding formula is not generally a problem as long as the pylorus is intact.

Jejunostomy Tube Feeding

Jejunostomy tube feedings usually are required for patients in whom nasoesophageal or gastrostomy tube feedings are contraindicated. This includes comatose patients, patients with high gastrointestinal fistulas or obstructions, and patients in whom a nasojeunal feeding tube cannot be placed. The jejunostomy may be of the Roux-en-Y (permanent) or the Witzel (temporary) type. The latter is constructed by inserting a number 18 French rubber catheter into the proximal jejunum approximately 30 cm distal to the ligament of Treitz.

The wall of the jejunum is inverted over the tube for about 3 cm as it emerges from the bowel to create a serosa-lined tunnel that allows rapid sealing of the jejunal opening when the tube is removed. An alternative procedure is the placement of a smaller-bore polyethylene or Silastic catheter. The tube is brought out through a stab incision in the left upper quadrant of the abdomen. The jejunum is sutured to the anterior abdominal wall at the point of tube entry to seal it from the peritoneal cavity. It is acceptable to place jejunostomy tubes concomitantly with major or trauma-related abdominal operations if prolonged nutritional support is anticipated in such patients.

Under some circumstances, when the need for postoperative enteral feedings is evident or likely, the placement of a needle-catheter feeding tube may obviate the need for more extensive procedures. The ability to secure jejunal feeding tubes endoscopically makes this a reasonable alternative to surgically constructed jejunostomies. In a large series, Maurer and colleagues reported the feasibility, safety, and cost-effectiveness of fluoroscopically guided percutaneous insertion of jejunostomy tubes in a diverse population of patients who would otherwise have required intraoperative or endoscopic feeding-tube placement.

If the jejunostomy tube is inadvertently removed, blind attempts at reinsertion should not be attempted. If discovered within a few hours, the tube may be reinserted under fluoroscopic guidance to be certain that it is in the bowel before feedings are resumed. The patient is observed for signs of peritonitis for 12 to 18 h after feedings are restarted. If there is any doubt about the position of the tube, it should be replaced surgically.

Feedings are safely begun 12 to 18 h after jejunostomy construction, even though peristalsis is not audible. Jejunostomy tube feedings usually are initiated with one of the many commercially available defined-formula diets (see Table [1-11](#)). When provided by continuous infusion, such formulas usually are well tolerated, but all patients must be serially evaluated to assure efficacy.

With proper care, about 85 percent of jejunostomy patients tolerate their feedings. When diarrhea occurs it usually can be controlled by temporarily reducing the concentration and volume of formula. Failing this, feeding is halted for a day, then resumed from the beginning of the feeding regimen, progressing more slowly than before. In many cases symptoms are relieved if the rate and volume of infusion are reduced and cold formula avoided. Failing control of diarrhea by these means, or as an alternative method to opiates, the periodic administration of bulk-forming agents might be helpful.

If the patient with a jejunostomy has a proximal bowel or biliary fistula draining more than 300 mL daily for a prolonged period, the fistular drainage may be collected by sump suction, cooled in an ice basin at bedside, and promptly re-fed in small increments throughout the day. To avoid jejunal overloading, the fistular fluid is re-fed between formula feedings. Aspirated gastric juice should not be re-fed, because it may cause jejunal irritation and profuse diarrhea. If the fistular drainage is profuse, it usually is not possible to re-feed more than 2 L/day, and fluid and electrolyte losses must be replaced with appropriate intravenous supplements. Additional water may be given with the feedings or administered between the feedings as indicated. Occasionally an elemental diet is indicated when other jejunostomy formulas are not tolerated.

Defined-Formula Diets

Commercial production of nutritionally complete liquid diets, derived in purified form from natural foods or from foods prepared synthetically are widely used in acute and chronic nutritional support efforts. These diets may be used for complete nutritional support or as dietary supplements for patients who are unable to eat or digest enough food to meet their energy requirements. They may be preferable to high-calorie parenteral feedings for patients who have part of the small bowel available for the absorption of simple sugars and amino acids. Elemental diets have been found useful for patients with depleted protein reserves secondary to gastrointestinal tract disease, such as ulcerative or granulomatous colitis and malabsorption syndrome, and for patients with only partial function of the gastrointestinal tract, such as the short bowel syndrome or gastric or small-bowel fistulas with feeding distal to the fistula. These diets also have been used during preoperative bowel preparation.

The commercially prepared diets also contain baseline electrolytes, water, fat-soluble vitamins (except vitamin K in some cases), and trace minerals. They contain no bulk and therefore produce a minimum of residue. They do not contain lactose and are more readily tolerated in such lactase-deficiency states as gastroenteritis, intestinal resection, radiation, or genetic predisposition. There are several products whose protein content is partially hydrolyzed or completely hydrolyzed to amino acids or dipeptides. When digestion and absorption are normal, there appears to be little therapeutic advantage to the use of crystalline amino acid formulas. A listing of the basic constituents for several commercial preparations as well as the volume necessary to achieve minimal daily requirements is given in Table [1-11](#). The practitioner should consult product information to ascertain further details regarding the precise composition of a prescribed formula.

Special products designed for use in the presence of organ dysfunction also are available (see Table [1-11](#)). Fat may contribute less than 1 percent or as much as 47 percent of the calories in these commercial formulas. Most contain long-chain fats as corn oil, soy oil, or safflower oil. Some include medium-chain triglycerides. Because the high caloric density of fat does not increase the osmolality of the formula when significant maldigestion or malabsorption is present, a diet low in fat or one supplemented with medium-chain triglycerides may be useful. To prevent the development of essential fatty acid deficiency, the clinician must be mindful of the possible need for additional lipids when providing such specialized diets to patients.

Specific products are limited in their overall clinical usefulness by virtue of the fixed content of nutrients. There has been a trend toward preparing enteral diets in modular form whereby certain critical items, such as sodium, potassium, and fat, can be modified in concentration as needed.

The amount of elemental diet required to maintain weight and nitrogen balance varies with the individual patient. In severe catabolic states, the standard diet often fails to achieve positive nitrogen balance. Careful attention to water and electrolyte balance is mandatory, particularly when large quantities of fluid are being lost through fistulas or other routes. Additional sodium and potassium may be added to the mixture (not to exceed a total of 100 mEq), although they should be given in intravenous fluids when larger quantities are needed. Water may be added to the mixture in the face of excessive pure water losses.

Complications include nausea, vomiting, and diarrhea that develop because of the high osmolality of the diets. This generally can be controlled by decreasing the rate and/or concentration of the mixture. Hypertonic nonketotic coma may occur in the presence of excessive water losses or if the diets are administered at concentrations above those recommended. Hyperglycemia and glycosuria can occur in any severely ill patient, particularly latent diabetics, and insulin may be indicated.

A number of prospective trials have suggested that one or more enteral nutritional regimens may reduce complications and improve outcome. These formulations are promoted as enhancing various aspects of immune or solid organ function. There is no evidence to suggest that this is uniformly the case given that other prospective randomized trials failed to document any benefits.

Parenteral Alimentation

Parenteral alimentation involves the continuous infusion of a hyperosmolar solution containing carbohydrates, proteins, fat, and other necessary nutrients through an indwelling catheter inserted into the superior vena cava. In order to obtain the maximum benefit, the ratio of calories to nitrogen must be adequate (at least 100 to 150 kcal/g nitrogen) and the two materials must be infused simultaneously. When the sources of calories and nitrogen are given at different times, there is a significant decrease in nitrogen use. These nutrients can be given in quantities considerably greater than the basic caloric and nitrogen requirements, and this method has proved highly successful in achieving growth and development, positive nitrogen balance, and weight gain in a variety of clinical situations.

Indications for the Use of Intravenous Hyperalimentation

It is difficult to demonstrate that parenteral feeding significantly alters the clinical course or outcome in most nonsurgical patient populations. Clinical trials and metaanalysis of parenteral feeding in the perioperative period have, however, suggested that preoperative nutritional support may benefit some surgical patients, particularly those with extensive malnutrition. By contrast, definitive evidence of benefit accruing from use of nutritional support in the postoperative setting is lacking. The routine use of parenteral alimentation in the critical-care environment has yet to be adequately assessed, and so it is currently used intuitively. The evidence underlying the application of parenteral nutrition in situations of surgical relevance was reviewed before the formulation of clinical practice guidelines published by a recent Georgetown University panel. The principal indications for parenteral alimentation are found in seriously ill patients suffering from malnutrition, sepsis, or surgical or accidental trauma when use of the gastrointestinal tract for feedings is not possible. It has been used in many instances in which it is not needed or in which the use of the gastrointestinal tract is more appropriate. In some instances intravenous nutrition may be used to supplement inadequate oral intake. The safe and successful use of this regimen requires proper selection of patients with specific nutritional needs, experience with the technique, and an awareness of the associated complications. The fundamental goals are to provide sufficient calories and nitrogen substrate to promote tissue repair and to maintain the integrity or growth of lean tissue mass. Listed below are situations in which parenteral nutrition has been used in an effort to achieve these goals. Indications 1 and 2 below usually are exclusively used for intravenous nutrition. Indications 3 to 13 might be appropriate for enteral or parenteral nutrition.

(1) Newborn infants with catastrophic gastrointestinal anomalies, such as tracheoesophageal fistula, gastroschisis, omphalocele, or massive intestinal atresia.

(2) Infants who fail to thrive nonspecifically or secondarily to gastrointestinal insufficiency associated with the short bowel syndrome, malabsorption, enzyme deficiency, meconium ileus, or idiopathic diarrhea.

(3) Adult patients with short bowel syndrome secondary to massive small-bowel resection or enteroenteric, enterocolic, enterovesical, or enterocutaneous fistulas.

(4) Patients with high alimentary tract obstructions without vascular compromise, secondary to achalasia, stricture, or neoplasia of the esophagus, gastric carcinoma, or pyloric obstruction.

(5) Surgical patients with prolonged paralytic ileus after major operations, multiple injuries, or blunt or open abdominal trauma, or patients with reflex ileus complicating various medical diseases.

(6) Patients with normal bowel length but with malabsorption secondary to sprue, hypoproteinemia, enzyme or pancreatic insufficiency, regional enteritis, or ulcerative colitis.

(7) Adult patients with functional gastrointestinal disorders such as esophageal dyskinesia after cerebrovascular accident, idiopathic diarrhea, psychogenic vomiting, or anorexia nervosa.

(8) Patients who cannot ingest food or who regurgitate and aspirate oral or tube feedings because of depressed or obtunded sensorium after severe metabolic derangements,

neurologic disorders, intracranial surgery, or central nervous system trauma.

- (9) Patients with excessive metabolic requirements secondary to severe trauma, such as extensive full-thickness burns, major fractures, or soft- tissue injuries.
- (10) Patients with granulomatous colitis, ulcerative colitis, and tuberculous enteritis, in which major portions of the absorptive mucosa are diseased.
- (11) Paraplegics, quadriplegics, or debilitated patients with indolent decubitus ulcers in the pelvic areas, particularly when soilage and fecal contamination are a problem.
- (12) Patients with malignancy, with or without cachexia, in whom malnutrition might jeopardize successful delivery of a therapeutic option.
- (13) Patients with potentially reversible acute renal failure, in whom marked catabolism results in the liberation of intracellular anions and cations, inducing hyperkalemia, hypermagnesemia, and hyperphosphatemia.

Contraindications to hyperalimentation include the following:

- (1) Lack of a specific goal for patient management, or when instead of extending a meaningful life, inevitable dying is prolonged.
- (2) Periods of cardiovascular instability or severe metabolic derangement requiring control or correction before attempting hypertonic intravenous feeding.
- (3) Feasible gastrointestinal tract feeding; in the vast majority of instances, this is the best route by which to provide nutrition.
- (4) Patients in good nutritional status, in whom only short-term parenteral nutrition support is required or anticipated.
- (5) Infants with less than 8 cm of small bowel, since virtually all have been unable to adapt sufficiently despite prolonged periods of parenteral nutrition.
- (6) Patients who are irreversibly decerebrate or otherwise dehumanized.

Insertion of Central Venous Infusion Catheter

The successful use of intravenous hyperalimentation generally depends on the proper placement and management of the central venous feeding catheter. A 16-gauge, 8- or 12-inch radiopaque catheter is introduced percutaneously through the subclavian or internal jugular vein and threaded into the superior vena cava. Although the technique for subclavian vein puncture (Fig. 1-38) has been the preferable and more widely used technique, the internal jugular approach also may be useful (Fig. 1-39). For insertion of the intravenous catheter through the subclavian vein, the patient is placed supine in a 15-degree head-down position with a small pad placed between the shoulder blades to allow the shoulders to drop posteriorly. This allows expansion of the subclavian vein and easier penetration. The skin may be scrubbed with acetone to defat the surface and then with an iodophor compound. Drapes are carefully placed, and *scrupulous* aseptic precautions are observed. Local anesthetic is infiltrated into the skin, forming a wheal, as well as into the subcutaneous tissue and periosteum at the inferior border of the midpoint of the clavicle. Most commercially prepared kits for central venous catheter insertion are equipped with a 2- to 2½-inch long, 16-gauge needle. This needle attaches to a 5 to 10 mL syringe and is inserted, beveled down through the wheal, and advanced toward the tip of the operator's finger, which is pressed well into the patient's suprasternal notch. The needle should hug the inferior clavicular surface and go over the first rib into the subclavian vein. With slight negative pressure applied to the syringe, entrance into the vein will be noted by the appearance of blood. The needle is advanced a few millimeters further to be sure that it is entirely within the lumen of the vein. The thumb is held over the needle hub as the syringe is removed to avoid air embolism. A flexible guide-wire is then introduced into the vein through the bore of the needle, leaving at least half the length of the guide-wire visible above the skin to prevent losing the wire in the vein or inducing arrhythmias. With control of the guide-wire, the needle is then withdrawn and removed over the guide-wire. The puncture site can be slightly extended with a number 11 blade, and a dilator forms a tract in the soft tissue by being passed in and out over the guide-wire. A 16-gauge, 8- or 12-inch radiopaque catheter is then introduced over the guide-wire until the catheter tip reaches the superior vena cava. The guide-wire is then withdrawn and venous blood return is ascertained by syringe aspiration. The catheter hub is then connected to a sterile intravenous administration tubing, and a slow infusion is begun while the catheter is sewn to the skin with a small suture. Antibiotic ointment is routinely applied around the entrance of the catheter into the skin, and an occlusive dressing is applied over it, including the junction of the intravenous tubing with the catheter. A chest film is immediately obtained to confirm the position of the radiopaque catheter in the vena cava and to check for a possible pneumothorax. Every 2 or 3 days, the intravenous tubing is changed at the catheter entry site over a guide-wire. The catheter site is scrubbed as for an operative procedure, and antibiotic ointment and a new occlusive dressing are applied. Withdrawal or administration of blood through the catheter or the use of the catheter for central venous pressure measurements should be avoided, since the risk of contamination and catheter occlusion are significantly increased. The use of the internal jugular approach is satisfactory but more prone to developing local or systemic infection. It is unwise, unless absolutely necessary, to place catheters into the inferior vena cava from the lower extremities because of the greater likelihood of sepsis and thromboembolic phenomena. Long catheters inserted through the antecubital cephalic or basilic veins may be advanced into the superior vena cava. Initial experience with polyvinyl chloride catheters was disappointing because of a high incidence of thrombophlebitis. The use of polyurethane catheters has decreased the incidence of such complications. Other vascular access methods may be achieved by surgical cutdowns and passing the catheters through the subcutaneous tissue of the anterior chest wall or the neck (Fig. 1-40).

Preparation and Administration of Solutions

The basic solution contains a final concentration of 20 to 25% dextrose and 3 to 5% crystalline amino acids. The solutions are usually prepared sterily in the pharmacy from commercially available kits containing the component solutions and transfer apparatus. Preparation in the pharmacy under laminar flow reduces the incidence of bacterial contamination of the solution. Proper preparation with suitable quality control is essential to avoid septic complications. Because of the considerable variability in amino acid and electrolyte concentrations among commercially available alimentation formulations, physicians must become familiar with the solutions used in their institutions. Only in this manner may additives, in the form of additional electrolytes, be rationally planned to meet the specific metabolic needs of the patient. Electrolyte requirements may vary considerably from patient to patient, depending on routes of fluid and electrolyte loss, renal function, metabolic rate, cardiac function, and underlying disease state. Intravenous vitamin preparations also should be added to parenteral formulations. Vitamin deficiencies are rare occurrences if such preparations are used. In addition, phytonadione (vitamin K₁) 10 mg and folic acid 5 mg should be administered intramuscularly once a week, since these are unstable in the hyperalimentation solution. Cyanocobalamin (vitamin B₁₂) 1 mg is given by intramuscular injection once a month. Intramuscular administration of iron may be required for patients with iron deficiency anemia, although adequate mobilization of iron stores may occur once the patient is anabolic. During prolonged fat-free parenteral nutrition essential fatty acid deficiency may become apparent, manifested by a dry, scaly dermatitis and loss of hair. The syndrome may be prevented by periodic infusion of a fat emulsion at a rate equivalent to 10 to 15 percent of total calories (Table 1-13). Essential trace minerals may be required after prolonged total parenteral nutrition and can be supplied by direct addition of commercial preparations of dextrose amino acids solutions. The most frequent presentation of trace mineral deficiencies is the eczematoid rash that develops diffusely and at intertriginous areas in zinc-deficient patients. Other rare trace mineral deficiencies include a microcytic anemia associated with copper deficiency and glucose intolerance presumably related to chromium deficiency. These complications are seldom seen except in patients receiving parenteral nutrition for extended periods. The daily administration of commercially available trace mineral supplements obviates most such problems. Depending on fluid and nitrogen tolerance, parenteral nutrition solutions usually can be increased over 2 to 3 days to achieve the desired infusion rate. Insulin may be supplemented as necessary to ensure glucose tolerance. Wolf and Elwyn demonstrated that maximum efficiency of glucose use occurs at an infusion rate of 7 mg/kg body weight per min. Dextrose infusions above this level result in increased fat synthesis and provide no additional suppression of amino acid gluconeogenesis. Rarely, additional intravenous fluids and electrolytes might be necessary with continued abnormal large losses of fluids. The patient should be carefully monitored for development of electrolyte, volume, acid-base, and septic complications. Vital signs and urinary output are regularly observed, and the patient should be weighed daily. Frequent adjustments of the volume and composition of the solutions are necessary during the course of therapy. Electrolyte determinations are made daily until stable and every 2 or 3 days thereafter, and the hemogram, liver function chemistries, and levels of blood urea nitrogen, phosphate, and magnesium are determined weekly. The urine or capillary blood glucose level is checked every 6 h and serum glucose concentration checked at least once daily during the first few days of the infusion and at frequent intervals thereafter. Relative glucose intolerance may occur after initiation of parenteral alimentation. Insulin may be supplemented as necessary to improve carbohydrate tolerance. The response of blood glucose to exogenous insulin is evaluated by frequent capillary blood determinations, rather than reliance on glycosuria. If the blood glucose levels remain elevated or glycosuria persists, the dextrose concentration may be decreased, the infusion rate slowed, or regular insulin added to each bottle. The rise in blood glucose concentration observed after initiating an intravenous alimentation program may be temporary, as the normal pancreas increases its output of insulin in response to the continuous carbohydrate infusion. In patients with diabetes mellitus, additional crystalline or human insulin may be required. The administration of adequate amounts of potassium is essential to achieve positive nitrogen balance and to replace depleted intracellular stores. In addition, a significant shift of potassium ion from the extracellular to the intracellular space can take place because of the large glucose infusion, with resultant hypokalemia, metabolic alkalosis, and poor glucose utilization. In some cases as much as 240 mEq of potassium ion daily may be required. Hypokalemia may cause glycosuria, which would be treated with potassium, not insulin. Before giving insulin, the serum potassium level must be checked to avoid compounding the hypokalemia. Patients with insulin-dependent diabetes mellitus may exhibit wide fluctuations in blood glucose levels during parenteral nutrition. Partial replacement of lipid emulsions for dextrose calories may alleviate these problems in selected patients.

Fat Emulsions

Lipid emulsions derived from soybean or safflower oils are widely used as an adjunctive nutrient to prevent the development of essential fatty acid deficiency. They have been used as a major energy source in parenteral alimentation, but there is no evidence of enhanced metabolic efficacy if greater than 10 to 15 percent of calories is provided as lipid emulsions. Fat emulsion, dextrose, and amino acid combinations are as effective as carbohydrate and amino acid solutions in the repletion of nonstressed patients. The efficiency of fat as a caloric source in the traumatized, hypermetabolic patient is not well documented. There is a theoretic advantage to the use of lipid emulsions in some septic and trauma patients when nonsuppressible fat oxidation and increased norepinephrine excretion accompany glucose infusion. Patients with abnormal fat transport or metabolism, lipid nephrosis, coagulopathy, or serious pulmonary disease should not receive fat emulsions. Most investigators advise limitation of administered fat emulsions to between 2.0 and 2.5 g/kg of body weight per day.

Special Formulations

Numerous studies have documented the safety of parenteral alimentation in patients with renal failure. In these patients special formulations of essential amino acids may be indicated. Selection of the appropriate calorie and nitrogen concentration must be judged by fluid tolerance, associated illnesses, and the frequency of dialysis. Appropriate use of

dialysis is additive to nutritional support in improving survival of these patients. Solutions for patients with acute, oliguric renal failure contain a final dextrose concentration of 40 to 45 percent and only essential l-amino acids. In patients with nonoliguric renal failure, it may be possible to use essential and nonessential amino acids to promote protein synthesis. Solutions designed for patients with hepatic failure contain increased levels of branched-chain amino acids and decreased concentrations of aromatic amino acids. Such solutions improve encephalopathy but may not improve survival, which is dictated by the underlying hepatic pathology. Patients with moderate hepatic reserve and alcoholic hepatitis also may be treated with standard parenteral formulas to control encephalopathy and ascites. Cachexia related to severe cardiac disease may be judiciously treated with highly concentrated dextrose and amino acid formulas that are low in sodium content. Given the rather modest benefits demonstrable with nutritional support during the catabolic phase, there has been much recent investigation into possible adjunctive therapies. Efforts to enhance anabolism via growth factor (growth hormone, IGF-1) or anabolic steroid administration generally do not appear to be of additional benefit under full-calorie feeding conditions. Some benefit might be achieved in patients who exhibit more chronic disease conditions in which weaning from nutritional support might be indicated.

Complications

Problems may arise in the placement and maintenance of venous access or in the formulation and delivery of parenteral solutions. One of the more common and serious complications associated with long-term parenteral feeding is sepsis secondary to contamination of the central venous catheter. Contamination of solutions should be considered but is rare when proper pharmacy protocols have been followed. This problem occurs more frequently in patients with systemic sepsis and in many cases is a consequence of hematogenous seeding of the catheter with bacteria. Usually it is a result of failure to observe strict aseptic precautions during preparation and administration of the solutions. One of the earliest signs of systemic sepsis may be the sudden development of glucose intolerance (with or without temperature increase) in a patient who previously has been maintained on parenteral alimentation without difficulty. When this occurs or if fever develops without obvious cause, a diligent search for a potential septic focus is indicated. Other causes of fever also should be investigated. If fever persists, the infusion catheter should be removed and cultured. Some centers are now replacing catheters considered at low risk for infection over a J-wire. Should evidence of infection persist over 24 to 48 h without a definable source, the catheter should be replaced in the opposite subclavian vein or into one of the internal jugular veins and the infusion restarted. It may be advisable to wait a short period before reinserting the catheter, especially if bacteremia or hemodynamic instability are present.

Other complications related to catheter placement include the development of pneumothorax, hemothorax, or hydrothorax; subclavian artery injury; cardiac arrhythmia if the catheter is placed into the atrium or the ventricle; air embolism or catheter embolism; and, rarely, cardiac perforation with tamponade. Clinically evident thrombophlebitis or thrombosis of the superior vena cava has been rare, but radiographically proved thrombophlebitis has been noted in up to 25 percent of selected patients.

Although the use of multiple-lumen catheters for infusion therapy and monitoring critically ill patients is occasionally indicated, the risks (particularly of sepsis and of venous thrombosis) attending the prolonged use of such catheters may be increased. Efforts should be directed toward replacing these catheters with standard single-lumen intravenous feeding catheters as early as possible. The acute nutritional management of surgical patients seldom requires the use of permanently implanted catheters (see Fig. 1-38). Use of these catheters should be restricted to nonseptic or high-risk patients requiring prolonged periods of nutritional or fluid therapy or for selected patients requiring frequent blood sampling.

Hyperosmolar nonketotic hyperglycemia may develop with normal rates of infusion in patients with impaired glucose tolerance or in any patient if the hypertonic solutions are administered too rapidly. This is a particularly common complication in latent diabetics and in patients who have had severe surgical stress or trauma. Treatment of the condition consists of volume replacement with correction of electrolyte abnormalities and the administration of insulin. This serious complication can be avoided with careful attention to daily fluid balance and frequent determinations of urine and blood glucose levels and serum electrolyte levels.

A number of volume, concentration, and compositional abnormalities also may develop, but these are largely avoided by careful attention to the details of patient management. This is particularly important for elderly patients and for patients with significant cardiovascular, renal, or hepatic disorders. It is important not to “overfeed” the parenterally nourished patient. This is particularly true of the depleted patient in whom excess calorie infusion may result in carbon dioxide retention and respiratory insufficiency. Excess feeding also has been related to the development of hepatic steatosis or marked glycogen deposition in certain patients. Mild abnormalities of serum transaminase, alkaline phosphatase, and bilirubin levels may occur in many parenterally nourished patients. Failure of the tests to plateau or return toward normal over 7 to 14 days should suggest another cause.

Home Parenteral Nutrition

Patients who do not require a hospital environment for management of their primary disease, yet cannot tolerate adequate enteral or oral feeding, may be candidates for home parenteral nutrition. As opposed to the temporary methods of vascular access, Silastic catheters have proved to be durable portals for long-term parenteral nutrition (see Fig. 1-40). Alternatives to this technique include the placement of subcutaneous infusion ports. An absolute catheter-related infection rate of 0.3 per year per patient is still encountered. While home parenteral nutrition is more cost-effective than similar inpatient methods, criteria for selection of patients must be more stringent than those listed above for hospitalized patients. Patients with terminal illnesses, lack of self-care ability, or lack of a supportive home environment are candidates for this method. Inflammatory bowel disease, motility disorders, and ischemic bowel infarction are also indications.

A period of inpatient training is necessary to acquaint the patient and family with appropriate methods of solution preparation and delivery. This is best done in a multidisciplinary setting in which professionals are thoroughly familiar with the acute and chronic complications of home parenteral nutrition. There is interest in new technologies that might promote the transition of parenteral nutrition–dependent patients to partial or complete enteral nutrition. Such methods have demonstrated some promise in initial trials but will most likely prove inadequate for most parenteral nutrition–dependent patients.

Acknowledgments

Supported in part by National Institutes of Health grant GM 34695. Dr. Lin is also supported by the Surgical Society of the New York Hospital–Queens.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 2 - Fluid and Electrolyte Management of the Surgical Patient

G. Tom Shires III
Annabel Barber
G. Tom Shires

ANATOMY OF BODY FLUIDS

One of the most critical aspects of patient care is management of the composition of body fluids and electrolytes. Most diseases, many injuries, and even operative trauma have a great impact on the physiology of fluids and electrolytes in the body. These changes often exceed those brought about by acute lack of alimentation. A thorough understanding of the metabolism of salt, water, and electrolytes and of certain metabolic responses is essential to the care of surgical patients.

In the sections that follow, the anatomy of body fluids and the physiologic principles that maintain normal fluid and electrolytes are defined, and a classification of derangements is outlined to allow an organized therapeutic approach.

A prerequisite to the understanding of fluid and electrolyte management is knowledge of the extent and composition of the various body fluid compartments. Early attempts to define these compartments were relatively accurate, but a more precise definition has been obtained by many investigators through the use of isotope tracer techniques. The wide range of normal values is a function of body size, weight, and sex, but these compartments are relatively constant in the individual patient in the normal steady state. The figures used in this section are approximate and presented as a percentage of body weight.

Total Body Water

Water constitutes 50 to 70 percent of total body weight. Using deuterium oxide or tritiated water for measurement of total body water, the average normal value is 60 percent of body weight for young adult males and 50 percent for young adult females. A normal variation of ± 15 percent applies to both groups. The actual figure for each healthy individual is remarkably constant and is a function of several variables, including age and lean body mass. Because fat contains little water, the lean individual has a greater proportion of water to total body weight than the obese person. The lower percentage of total body water in females correlates well with a relatively large amount of subcutaneous adipose tissue and small muscle mass. Moore and associates demonstrated that total body water as a percentage of total body weight decreases steadily and significantly with age to a low of 52 and 47 percent in males and females, respectively. Conversely, the highest proportion of total body water to body weight is found in newborns, with a maximum of 75 to 80 percent. During the first several months after birth, there is a gradual "physiologic" loss of body water as infants adjust to their environment. At 1 year of age, the total body water averages approximately 65 percent of the body weight and remains relatively constant throughout the remainder of infancy and childhood.

The water of the body is divided into three functional compartments (Fig. 2-1). The fluid within the body's diverse cell population represents 30 to 40 percent of the body weight. The extracellular water represents 20 percent of the body weight and is divided between the intravascular fluid, or plasma (5 percent of body weight), and the interstitial, or extravascular, extracellular fluid (15 percent of body weight).

Intracellular Fluid

Measurement of intracellular fluid is determined indirectly by subtraction of the measured extracellular fluid from the measured total body water. The intracellular water is 30 to 40 percent of the body weight, with the largest proportion in the skeletal muscle mass. Because of the smaller muscle mass in females, the percentage of intracellular water is lower than in males.

The chemical composition of the intracellular fluid is shown in Fig. 2-2; potassium and magnesium are the principal cations, and phosphates and proteins the principal anions. This is an approximation, because few data concerning the intracellular fluid are available.

Extracellular Fluid

The total extracellular fluid volume represents approximately 20 percent of the body weight. The extracellular fluid compartment has two major subdivisions. The plasma volume comprises approximately 5 percent of the body weight in the normal adult. The interstitial, or extravascular, extracellular fluid volume comprises approximately 15 percent of the body weight.

The interstitial fluid is further complicated by having a rapidly equilibrating functional component as well as several more slowly equilibrating nonfunctioning components. The nonfunctioning components include connective tissue water and water that has been termed transcellular, which includes cerebrospinal and joint fluids. This nonfunctional component normally represents only 10 percent of the interstitial fluid volume (1 to 2 percent of body weight) and should not be confused with the relatively nonfunctional extracellular fluid, often called a "third space," found in burns and soft-tissue injuries.

The normal constituents of the extracellular fluid are shown in Fig. 2-2; sodium is the principal cation, and chloride and bicarbonate the principal anions. Plasma and interstitial fluid differ slightly in ionic composition. Because plasma has a higher protein content (organic anions), its total concentration of cations is higher and its concentration of inorganic anions somewhat lower than those of interstitial fluid, as explained by the Gibbs-Donnan equilibrium equation (i.e., the product of the concentrations of any pair of diffusible cations and anions on one side of a semipermeable membrane is equal to the product of the same pair of ions on the other side). For practical considerations, however, they may be considered equal. The total concentration of intracellular ions exceeds that of the extracellular compartment and seems to violate the concept of osmolar equilibrium between the two compartments. This apparent discrepancy is due to the fact that the concentration of ions is expressed in milliequivalents (mEq), without regard to osmotic activity. In addition, some of the intracellular cations probably exist in undissociated form.

Osmotic Pressure

The physiologic and chemical activity of electrolytes depend on three factors: (1) the number of particles present per unit volume (moles or millimoles [mmol] per liter); (2) the number of electric charges per unit volume (equivalents or milliequivalents per liter); and (3) the number of osmotically active particles or ions per unit volume (osmoles or milliosmoles [mOsm] per liter). Measurements in grams or milligrams per 100 milliliters express the weight of the electrolytes per unit volume but do not allow a physiologic comparison of the solutes in a solution.

A mole of a substance is the molecular weight of that substance in grams, and a millimole is that figure expressed in milligrams. For example, a mole of sodium chloride is 58 g (Na-23, Cl-35), and a millimole is 58 mg. The expression, however, gives no direct information about the number of osmotically active ions in solution or the electric charges that they carry.

The electrolytes of the body fluids can be expressed in terms of chemical combining activity, or "equivalents." An equivalent of an ion is its atomic weight expressed in grams divided by the valence, and 1 mEq of an ion is that figure expressed in milligrams. In the case of univalent ions, 1 mEq is the same as a millimole. In the case of divalent ions, such as calcium or magnesium, 1 mmol equals 2 mEq. The importance of this expression is that 1 mEq of any substance will combine chemically with 1 mEq of any other substance; in any given solution, the number of milliequivalents of cations present is balanced by precisely the same number of milliequivalents of anions.

When the osmotic pressure of a solution is considered, it is more descriptive to use the units osmole and milliosmole. These units refer to the actual number of osmotically active particles present in solution but are not dependent on the chemical combining capacities of the substances. Thus a millimole of sodium chloride, which dissociates nearly completely into sodium and chloride, contributes 2 mOsm, and 1 mmol of sodium sulfate (Na_2SO_4), which dissociates into three particles, contributes 3 mOsm. One millimole of non-ionized substance, such as glucose, is equal to 1 mOsm of the substance.

The differences in ionic composition between intracellular and extracellular fluid are maintained by the semipermeable cell membrane. The total number of osmotically active particles is 290 to 310 mOsm in each compartment. Although the total osmotic pressure of a fluid is the sum of the partial pressures contributed by each of the solutes in that fluid, the effective osmotic pressure is dependent on those substances that fail to pass through the pores of the semipermeable membrane. The dissolved proteins in the plasma, therefore, are primarily responsible for effective osmotic pressure between the plasma and the interstitial fluid compartments. This is frequently referred to as the colloid oncotic pressure. The effective osmotic pressure between the extracellular and intracellular fluid compartments would be contributed to by any substance that does not traverse the cell membranes freely. While sodium, as the principal cation of the extracellular fluid, contributes a major portion of the osmotic pressure, other substances that fail to penetrate the cell membrane freely, such as glucose, also increase the effective osmotic pressure.

Because the cell membranes are completely permeable to water, the effective osmotic pressures in the two compartments are considered to be equal. Any condition that alters the effective osmotic pressure in either compartment results in redistribution of water between the compartments. Thus an increase in effective osmotic pressure in the extracellular fluid, which would occur typically as a result of increased sodium concentration, would cause a net transfer of water from the intracellular to the extracellular fluid compartment. This transfer of water would continue until the effective osmotic pressures in the two compartments were equal. Conversely, a decrease in the sodium concentration in the extracellular fluid will cause a transfer of water from the extracellular to the intracellular fluid compartment. Depletion of the extracellular fluid volume without a change in the concentration of ions will not result in transfer of free water from the intracellular space.

The intracellular fluid shares in losses that involve a change in concentration or composition of the extracellular fluid, but shares only slowly in changes involving loss of isotonic volume alone. For practical considerations, most losses and gains of body fluid are directly from the extracellular compartment.

NORMAL EXCHANGE OF FLUID AND ELECTROLYTES

Knowledge of the basic principles governing the internal and external exchanges of water and salt is mandatory for care of the patient undergoing major operative surgery. The stable internal fluid environment, which is maintained by the kidneys, brain, lungs, skin, and gastrointestinal tract, may be compromised by surgical stress or by direct damage to any of these organs.

Water Exchange

The normal individual consumes an average of 2000 to 2500 mL water per day; approximately 1500 mL water is taken by mouth, and the rest is extracted from solid food, either from the contents of the food or as the product of oxidation (Table 2-1). The daily water losses include 250 mL in stools, 800 to 1500 mL in urine, and approximately 600 mL as insensible loss. A patient deprived of all external access to water must still excrete a minimum of 500 to 800 mL urine per day in order to excrete the products of catabolism, in

addition to the mandatory insensible loss through the skin and lungs.

Insensible loss of water occurs through the skin (75 percent) and the lungs (25 percent) and is increased by hypermetabolism, hyperventilation, and fever. The insensible water loss through the skin is not from evaporation of water from sweat glands but from water vapor formed within the body and lost through the skin. With excessive heat production (or excessive environmental heat), the capacity for insensible loss through the skin is exceeded and sweating occurs. These losses may, but seldom do, exceed 250 mL/day per degree of fever. An unhumidified tracheostomy with hyperventilation increases the loss through the lungs and results in a total insensible loss up to 1.5 L/day.

A frequently overlooked source of gain is the water of solution, which is the water that holds carbohydrates and proteins in solution in the cell. Normally, gain of water from this source is zero, but after 4 to 5 days without food intake, the postoperative patient may begin to gain significant quantities of water (up to 500 mL/day) from excessive cellular catabolism.

Salt Gain and Losses

In the normal individual, daily salt intake varies from 50 to 90 mEq (3 to 5 g) as sodium chloride (Table 2-2). Balance is maintained primarily by the kidneys, which excrete the excess salt. Under conditions of reduced intake or extrarenal losses, the normal kidney can reduce sodium excretion to less than 1 mEq/day within 24 h after restriction. In the patient with salt- wasting kidneys, however, the loss may exceed 200 mEq/L of urine. Sweat represents a hypotonic loss of fluids with an average sodium concentration of 15 mEq/L in the acclimatized person. In the unacclimatized person, the sodium concentration in sweat may be 60 mEq/L or more. Insensible fluid lost from the skin and lungs, by definition, is pure water. For practical considerations, normal losses may be relatively free of salt in the healthy individual with normal renal function.

The volume and composition of various types of gastrointestinal secretions are shown in Table 2-3. Gastrointestinal losses usually are isotonic or slightly hypotonic, although there is considerable variation in the composition. These should be replaced by an essentially isotonic salt solution. It is also important to reiterate that distributional or sequestration losses of extracellular fluid at any point in the operative or postoperative course also represent isotonic losses of salt and water.

CLASSIFICATION OF BODY FLUID CHANGES

The disorders in fluid balance may be classified in three general categories: disturbances of (1) volume, (2) concentration, and (3) composition. Of primary importance is the concept that although these disturbances are interrelated, each is a separate entity.

If an isotonic salt solution is added to or lost from the body fluids, only the volume of the extracellular fluid is changed. The acute loss of an isotonic extracellular solution, such as intestinal juice, is followed by a significant decrease in the extracellular fluid volume and little, if any, change in the intracellular fluid volume. Fluid will not be transferred from the intracellular space to refill the depleted extracellular space as long as the osmolarity remains the same in the two compartments.

If water alone is added to or lost from the extracellular fluid, the concentration of osmotically active particles changes. Sodium ions account for 90 percent of the osmotically active particles in the extracellular fluid and generally reflect the tonicity of body fluid compartments. If the extracellular fluid is depleted of sodium, water will pass into the intracellular space until osmolarity is again equal in the two compartments.

The concentration of most other ions within the extracellular fluid compartment can be altered without significant change in the total number of osmotically active particles, thus producing only a compositional change. For instance, a rise of the serum potassium concentration from 4 to 8 mEq/L would have a significant effect on the myocardium, but it would not significantly change the effective osmotic pressure of the extracellular fluid compartment. Normally functioning kidneys minimize these changes considerably, particularly if the addition or loss of solute or water is gradual. An internal loss of extracellular fluid into a nonfunctional space, such as the sequestration of isotonic fluid in a burn, peritonitis, ascites, or muscle trauma, is termed a distributional change. This transfer or functional loss of extracellular fluid internally may be extracellular (e.g., peritonitis), or intracellular (e.g., hemorrhagic shock), or both (e.g., major burns). In any event, all distributional shifts or losses result in a contraction of the functional extracellular fluid space.

Volume Changes

Volume deficit or excess usually is diagnosed by clinical examination of the patient. There are no readily available, useful laboratory tests in the acute phase except measurement of the plasma volume. Direct measurement of the extracellular fluid volume using radioisotopic tracers is feasible only in a research setting. There are several laboratory tests, however, that indirectly reflect changes in extracellular fluid volume. The blood urea nitrogen (BUN) level rises with an extracellular fluid deficit of sufficient magnitude to reduce glomerular filtration. The serum creatinine level may not increase proportionally in young people with healthy kidneys, and this discrepancy often is used as one test to differentiate prerenal and renal azotemia. The concentration of formed elements in the blood, such as the hematocrit, increases with an extracellular fluid deficit and decreases with an extracellular fluid excess. The concentration of serum sodium is not related to the volume status of extracellular fluid; a severe volume deficit may exist with a normal, low, or high serum level.

Volume Deficit

Extracellular fluid volume deficit is the most common fluid disorder in the surgical patient. The lost fluid is not water alone, but water and electrolytes in approximately the same proportion as they exist in normal extracellular fluid. The most common causes of extracellular fluid volume deficit are losses of gastrointestinal fluids from vomiting, nasogastric suction, diarrhea, and fistular drainage. Other common causes include sequestration of fluid in soft-tissue injuries and infections, intraabdominal and retroperitoneal inflammatory processes, peritonitis, intestinal obstruction, and burns. The signs and symptoms of volume deficit are easily recognized and are listed in Table 2-4. The central nervous system and cardiovascular signs occur early with acute rapid losses, but tissue signs may be absent until the deficit has existed for at least 24 h. The central nervous system signs are similar to barbiturate intoxication and may be missed if the volume deficit is mild. The cardiovascular signs are secondary to a decrease in plasma volume and may be associated with varying degrees of hypotension in the patient with a severe extracellular fluid volume deficit. Skin turgor may be difficult to assess in the elderly patient or in the patient with recent weight loss and is not diagnostic in the absence of other confirmatory signs. The body temperature tends to vary with the environmental temperature. In a cool room, the patient may be slightly hypothermic and the febrile response to illness may be suppressed. This occurs frequently and can be very misleading during clinical evaluation of the septic patient. After partial correction of the volume deficit, the temperature generally will rise to the appropriate level. Severe volume depletion depresses all body systems and interferes with the clinical evaluation of a patient. For example, a volume-depleted patient with severe sepsis from peritonitis may have a normal temperature and white blood cell count, complain of little pain, and have unremarkable findings on abdominal examination. The clinical picture may change dramatically, however, when the extracellular fluid volume is restored.

Volume Excess

Extracellular fluid volume excess may be generally iatrogenic or secondary to renal insufficiency, cirrhosis, or congestive heart failure. Plasma and interstitial fluid volumes are increased. In the healthy young adult the signs are generally those of circulatory overload, manifested primarily in the pulmonary circulation, and of excessive fluid in other tissue (see Table 2-4). In the elderly patient, congestive heart failure with pulmonary edema may develop quickly with a moderate volume excess.

Concentration Changes

Sodium is primarily responsible for the osmolarity of the extracellular fluid space: determination of the serum concentration of sodium generally indicates the tonicity of body fluids. Hyponatremia and hypernatremia can be diagnosed on clinical grounds (Table 2-5), but signs and symptoms generally are not present until the changes are severe. Clinical signs of hyponatremia or hypernatremia occur early and with greater severity when the rate of change in extracellular sodium concentration is very rapid. Changes in concentration should be noted early by laboratory tests and corrected promptly.

Hyponatremia

Acute symptomatic hyponatremia (sodium less than 130 mEq/L) clinically is characterized by central nervous system signs of increased intracranial pressure and tissue signs of excessive intracellular water. The hypertension probably is induced by the rise in intracranial pressure, and the blood pressure usually returns to normal with the administration of hypertonic solutions of sodium salts. Of importance with severe hyponatremia is the relatively rapid development of oliguric renal failure, which may not be reversible if therapy is delayed.

Many chronic hyponatremic states are asymptomatic until the serum sodium level falls below 120 mEq/L. One important exception is the patient with increased intracranial pressure after closed head injury, in whom mild hyponatremia may be fatal, because of the progressive increase in intracellular water as the extracellular fluid osmolarity falls.

Hypernatremia

Central nervous system and tissue signs characterize acute symptomatic hypernatremia. This is the only state in which dry, sticky mucous membranes are characteristic. This sign does not occur with pure extracellular fluid volume deficit alone and may be misleading in the patient who breathes through the mouth. Body temperature generally is elevated and may approach a lethal level, as in the patient with heatstroke.

While volume changes occur frequently without any change in serum sodium concentration, the reverse is not true. The disease states that cause a significant acute alteration in the serum sodium level frequently produce a concomitant change in the extracellular fluid volume.

Mixed Volume and Concentration Abnormalities

Mixed volume and concentration abnormalities may develop as a consequence of the disease state or occasionally as a result of inappropriate parenteral fluid therapy. Moyer noted that the clinical picture associated with a combination of fluid abnormalities will be an algebraic composite of the signs and symptoms of each state. Like signs produced by both abnormalities will be additive, and opposing signs will nullify one another. For example, the tendency for the body temperature to fall with an extracellular volume deficit may be counteracted by the tendency for it to rise with severe hypernatremia.

One of the more common mixed abnormalities is an extracellular fluid deficit and hyponatremia. This state is readily produced in the patient who continues to drink water while losing large volumes of gastrointestinal fluids. It may also occur in the postoperative period when gastrointestinal losses are replaced with inadequate volumes of only 5% dextrose in water or a hypotonic sodium solution. An extracellular volume deficit accompanied by hypernatremia may be produced by the loss of a large amount of hypotonic salt solution, such as sweat, in the absence of fluid intake.

The prolonged administration of excessive quantities of sodium salts with restricted water intake may result in an extracellular volume excess and hypernatremia. This may also occur when pure water losses (such as insensible loss of water from the skin and lungs) are replaced with sodium- containing solutions only. Similarly, the excessive administration of water or hypotonic salt solutions to the patient with oliguric renal failure may rapidly produce an extracellular volume excess and hyponatremia.

Normally functioning kidneys may minimize these changes to some extent and compensate for many of the imprecise replacements associated with parenteral fluid administration. In contrast, the patient in anuric or oliguric renal failure is particularly prone to develop these mixed volume and osmolar concentration abnormalities. Fluid and electrolyte management in these patients, therefore, must be precise. Unfortunately, the fact that a patient with normal kidneys who develops a significant volume deficit may be in a state of "functional" renal failure often is not appreciated. As the volume deficit progresses, the glomerular filtration rate falls precipitously, and the kidneys' unique functions for maintaining fluid homeostasis are lost. These changes may occur with only a mild volume deficit in the elderly patient with borderline renal function. In these elderly patients, the blood urea nitrogen level may rise higher than 100 mg/dL in response to the fluid deficit with a concomitant rise in the serum creatinine level. Fortunately, these changes usually are reversible with early and adequate correction of the extracellular fluid volume deficit.

Composition Changes

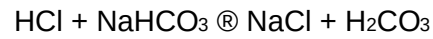
Compositional abnormalities of importance include changes in acid-base balance and changes in the concentration of potassium, calcium, and magnesium.

Acid-Base Balance

The pH of the body fluids is normally maintained within narrow limits in spite of the large load of acid produced endogenously as a by-product of body metabolism. The acids are neutralized efficiently by several buffer systems and subsequently excreted by the lungs and kidneys.

The important buffers include proteins and phosphates, which play a primary role in maintaining intracellular pH, and the bicarbonate–carbonic acid system, which operates principally in the extracellular fluid space. The proteins and hemoglobin have only minor influence in the extracellular fluid space, but the latter is of prime significance as an intracellular buffer in the red blood cell.

A buffer system consists of a weak acid or base and the salt of that acid or base. The buffering effect is the result of the formation of an amount of weak acid or base equivalent to the amount of strong acid or base added to the system. The resultant change in pH is considerably less than if the substance were added to water alone. Thus inorganic acids (e.g., hydrochloric, sulfuric, phosphoric) and organic acids (e.g., lactic, pyruvic, keto acids) combine with base bicarbonate, producing the sodium salt of the acid and carbonic acid:



The carbonic acid formed is then excreted via the lungs as CO_2 . The inorganic acid anions are excreted by the kidneys with hydrogen or as ammonium salts. The organic acid anions generally are metabolized as the underlying disorder is corrected, although some renal excretion may occur with high levels.

The functions of the buffer systems are expressed in the Henderson- Hasselbalch equation, which defines the pH in terms of the ratio of the salt and acid. The pH of the extracellular fluid is defined primarily by the ratio of the amount of base bicarbonate (the majority as sodium bicarbonate) to the amount of carbonic acid (related to the CO_2 content of alveolar air) present in the blood:

$$\text{pH} = \text{pK} + \log \frac{\text{HCO}_3^-}{\text{H}_2\text{CO}_3} = 7.38 + \log \frac{27 \text{ mEq/L}}{1.33 \text{ mEq/L}} = 7.38 + 1 = 8.38$$

The term pK represents the dissociation constant of carbonic acid in the presence of base bicarbonate, which by measurement is 6.1. At a body pH of 7.4, the ratio must be 20:1, as depicted. From a chemical standpoint, this is an inefficient buffer system, but the unusual property of CO_2 of behaving as an acid or changing to a neutral gas subsequently excreted by the lungs makes it quite efficient biologically.

As long as the 20:1 ratio is maintained, regardless of the absolute values, the pH will remain at 7.4. When an acid is added to the system, the concentration of bicarbonate (the numerator in the Henderson-Hasselbalch equation) decreases. Ventilation immediately increases to eliminate larger quantities of CO_2 , with a subsequent decrease in the carbonic acid (the denominator in the Henderson-Hasselbalch equation) until the 20:1 ratio is reestablished. Slower, more complete compensation is effected by the kidneys with increased excretion of acid salts and retention of bicarbonate. The reverse occurs if an alkali is added to the system. Respiratory acidosis and alkalosis are produced by disturbances of ventilation, with an increase or decrease in the denominator and hence a change from the 20:1 ratio. Compensation is primarily renal, with retention of bicarbonate and increased excretion of acid salts in respiratory acidosis and the reverse process in respiratory alkalosis.

The four types of acid-base disturbances are listed in Table 2-6. Use of the CO_2 combining power (approximates the plasma bicarbonate) or CO_2 content (includes bicarbonate, carbonic acid, and dissolved CO_2) and knowledge of the patient's disease may allow an accurate diagnosis in the uncomplicated case. Use of the serum CO_2 content or CO_2 combining power alone is generally inadequate as an index of acid-base balance. This test principally reflects the level of plasma bicarbonate, dissolved CO_2 and carbonic acid contribute no more than a few millimoles under most circumstances. In the acute phase, therefore, respiratory acidosis or alkalosis may exist without any change in the serum CO_2 content; determinations of the pH and P_{CO_2} from a freshly drawn arterial blood sample are necessary for diagnosis.

More complex acid-base disturbances are encountered frequently. Combinations of respiratory and metabolic changes may represent compensation for the initial acid-base disturbance or may indicate two or more coexisting primary disorders (Table 2-7).

A knowledge of the pH, bicarbonate concentration, and Pa_{CO_2} allows an accurate diagnosis of most acid-base disturbances. However, the clinical interpretation of these measurements is associated with some inherent problems. Although the Pa_{CO_2} is considered an accurate index of primary respiratory disturbances, changes in the level may represent compensation for a primary metabolic alteration. Thus a depressed Pa_{CO_2} (below 40 mmHg) is characteristic of respiratory alkalosis but also represents the normal compensatory response to a metabolic acidosis. Similarly, the level of plasma bicarbonate cannot be regarded exclusively as an index of metabolic disturbances. An elevated plasma bicarbonate level may indicate a primary metabolic alkalosis or a compensatory response to chronic respiratory acidosis. Astrup and colleagues proposed the use of the standard bicarbonate and base excess values. Base excess (or deficit) directly expresses, in mEq/L, the amount of fixed base (or acid) added to each liter of blood. This defines the metabolic component of acid-base disorders.

One useful approach to defining pure, combined, or compensated disturbances relates measured changes in Pa_{CO_2} and pH to calculated changes that would be expected from pure etiologies. Within reasonable physiologic ranges, a 10-mmHg change in Pa_{CO_2} yields a 0.08 change in pH from the normal values of Pa_{CO_2} (40 mmHg) and pH (7.4).

Respiratory Acidosis

This condition is associated with retention of CO_2 secondary to decreased alveolar ventilation. The more common causes are listed in Table 2-6. Initially the Pa_{CO_2} is elevated (usually above 50 mmHg), and the serum bicarbonate concentration (measured as CO_2 content) is normal. In the chronic form, the Pa_{CO_2} remains elevated, and the bicarbonate concentration rises as renal compensation occurs.

This problem may be particularly serious in the patient with chronic pulmonary disease, in whom preexisting respiratory acidosis may be accentuated in the postoperative period. A number of conditions resulting in inadequate ventilation (e.g., airway obstruction, atelectasis, pneumonia, pleural effusion, pain from an upper abdominal incision, or abdominal distention limiting diaphragmatic excursion) may exist singly or in combination to produce respiratory acidosis. Although restlessness, hypertension, and tachycardia in the immediate postoperative period may be caused by pain, similar signs indicate inadequate ventilation with hypercapnia. The use of narcotics in this situation compounds the problem by depressing respiration.

Management involves prompt correction of the pulmonary defect, when feasible, and measures to ensure adequate ventilation. Endotracheal intubation and mechanical ventilation occasionally are necessary. Strict attention to tracheobronchial hygiene during the postoperative period is an important preventive measure in all patients, and particularly in those with chronic pulmonary disease. Encouraging deep breathing and coughing, using humidified air to prevent inspissation of secretions, and avoiding oversedation are all indicated.

Respiratory Alkalosis

Respiratory alkalosis is a more common problem in the surgical patient than previously recognized. Hyperventilation because of apprehension, pain, hypoxia, central nervous system injury, and assisted ventilation are all common causes. Any of these conditions may cause a rapid decrease in the Pa_{CO_2} and increase in serum pH. The serum bicarbonate concentration is normal in the acute phase but falls with compensation if the condition persists.

The majority of patients who require ventilatory support in the postoperative period develop varying degrees of respiratory alkalosis. This may be the inadvertent result of improper use of the mechanical respirator, or it may occur during attempts to raise the P_{O_2} in a hypoxic patient. Proper management of the patient on a mechanical ventilator requires frequent measurements of blood gases and appropriate corrections of the ventilatory pattern when indicated. The Pa_{CO_2} should not be allowed to fall below 30 mmHg, as serious complications may occur, particularly in the presence of a complicating hypokalemia or metabolic alkalosis. Generally the Pa_{CO_2} can be maintained at an acceptable level by proper adjustments of the ventilatory rate and volume.

The dangers of a severe respiratory alkalosis are those related to potassium depletion and include the development of ventricular arrhythmias and fibrillation, particularly in patients who are digitalized or have preexisting hypokalemia. Other complications include a shift of the oxyhemoglobin dissociation curve to the left, which limits the ability of hemoglobin to unload oxygen at the tissue level except at low tissue oxygen tensions, and the development of tetany and convulsions if the level of ionized calcium is significantly depressed. The development of hypokalemia may be quite sudden and is related to entry of potassium ions into the cells in exchange for hydrogen and an excessive urinary potassium loss in exchange for sodium. Severe and persistent respiratory alkalosis often is difficult to correct and may be associated with a poor prognosis because of the underlying cause of hyperventilation. The treatment of alkalosis is directed primarily toward preventing the condition by the proper use of mechanical ventilation and correcting preexisting potassium deficits.

Metabolic Acidosis

Metabolic acidosis results from the retention or gain of fixed acids (diabetic ketoacidosis, lactic acidosis, azotemia) or the loss of base bicarbonate (diarrhea, small-bowel fistula, renal insufficiency with inability to resorb bicarbonate). The excess of hydrogen ions results in lower pH and serum bicarbonate concentration. The initial compensation is pulmonary, with an increase in the rate and depth of breathing and depression of the Pa_{CO_2} .

Renal damage may interfere with the important role of the kidneys in the regulation of acid-base balance. The kidneys serve a vital function in this regard through the excretion of nitrogenous waste products and acid metabolites and the resorption of bicarbonates. If renal damage occurs and these functions are lost, metabolic acidosis develops rapidly and may be difficult to control.

With normal kidneys, metabolic acidosis may develop when the capacity of the kidneys for handling a large chloride load is exceeded. This is particularly common in patients who have excessive losses of alkaline gastrointestinal fluids (biliary, pancreatic, small-bowel secretions) and are maintained on parenteral fluids for an extended period. Continued replacement of these losses with fluids that have an inappropriate chloride- bicarbonate ratio, such as isotonic sodium chloride solution, will not correct the pH change; the use of a balanced salt solution, such as lactated Ringer's solution, is indicated. Anion gap has become a useful tool in the management of acid-base disorders. This value can be determined routinely when evaluating serum electrolytes. The gap is calculated from the sum of serum chloride and bicarbonate levels subtracted from the serum sodium concentration. The normal value is 10 to 15 mEq/L. The anion gap is a laboratory anomaly because routine clinical laboratory tests measure the cations sodium and potassium and the anions chloride and bicarbonate. The unmeasured anions that account for the "gap" are sulfate and phosphate plus lactate and other organic anions. If the acidosis is a result of loss of bicarbonate (e.g., diarrhea) or gain of a chloride acid (e.g., administration of ammonium chloride), the anion gap will be normal. Conversely, if the acidosis is a result of increased production of an organic acid (e.g., lactic acid in circulatory shock) or the retention of sulfuric or phosphoric acid (e.g., renal failure), the concentration of unmeasured anions (anion gap) will be increased.

Conditions associated with an elevated anion gap are listed in Table 2-8. The most common cause of an elevated anion gap is shock or inadequate tissue perfusion from any number of causes, resulting in accumulation of large quantities of lactic acid. Diabetic ketoacidosis, starvation, and ethanol intoxication cause elevation of the anion gap by the formation of keto acids; renal failure and uremia cause the elevation by the retention of sulfuric and phosphoric acids. Poisoning by methanol, ethylene glycol, and aspirin produce increased anion gaps by elevation of their organic acid counterparts (formic, oxalic, and salicylic acids). These causes should be considered, singly or in combination, in a patient with an elevated anion gap.

One of the most common causes of severe metabolic acidosis in surgical patients is acute circulatory failure with accumulation of lactic acid. This is a reflection of tissue hypoxia from inadequate perfusion, although it is only one of the manifestations of cellular dysfunction. Acute hemorrhagic shock may result in a rapid and profound drop in the pH, and attempts to raise the blood pressure with vasopressors will compound the problem by further compromising tissue perfusion. Similarly, attempts to correct the acidosis by the infusion of large quantities of sodium bicarbonate without restoration of flow are futile. After restoration of adequate tissue perfusion by proper volume replacement, the lactic acid is quickly metabolized and the pH returned to normal. The use of lactated Ringer's solution to replace the extracellular fluid deficit incurred with hemorrhagic shock concomitant with administration of whole blood does not accentuate the lactic acidosis. Instead there is a rapid decrease in the serum lactate and return of pH toward normal, which is not the case when whole blood alone is used (see Table 2-8).

The indiscriminate use of sodium bicarbonate during the resuscitation of patients in hypovolemic shock is discouraged for several reasons. A mild metabolic alkalosis is a common finding after resuscitation, in part because of the alkalizing effects of blood transfusions and the administration of lactated Ringer's solution. After infusion (and partial restoration of hepatic blood flow), the citrate contained in the transfused blood and the lactate in lactated Ringer's solution are metabolized and bicarbonate is formed. The organic acidosis (lactic acid) that developed during the shock episode is rapidly cleared once adequate tissue perfusion is restored. Lactic acid production ceases, the hydrogen ion load is buffered and excreted via the lungs as CO₂, and the organic anion, lactate, is metabolized to bicarbonate by the liver. If excessive quantities of sodium bicarbonate are administered simultaneously, severe metabolic alkalosis can result. An alkaline pH may be highly undesirable in this situation, particularly in patients with hypoxia or a low fixed cardiac output, because it shifts the oxyhemoglobin dissociation curve to the left. Other factors that shift the oxygen dissociation curve to the left in this situation include the depressed level of erythrocyte 2,3- diphosphoglycerate in the transfused blood and the development of hypothermia. If the curve shifts far enough to the left, significant interference with oxygen unloading at the tissue level may occur.

The treatment of metabolic acidosis should be directed toward correction of the underlying disorder when possible. Bicarbonate therapy properly may be reserved for the treatment of severe metabolic acidosis, particularly after cardiac arrest, when partial correction of the pH may be essential to restore myocardial function. Studies indicate that the acidosis accompanying cardiac arrest is well compensated for a significant period of time if the patient is well ventilated and not previously acidotic. In addition, the administration of bicarbonate in the usual recommended doses may induce an acute and severe hyponatremia and hyperosmolarity. Bicarbonate should be used judiciously during cardiac arrest; the initial dose should not exceed 50 mL of 7.5% solution (45 mEq NaHCO₃ containing 90 mOsm), and the decision for additional doses should be based on measurements of pH and PaCO₂ when possible.

Similarly, pH correction of more protracted states of metabolic acidosis may be indicated, but it should be accomplished slowly. Since no satisfactory formula for estimating the amount of alkali needed has been devised, frequent measurements of serum electrolyte levels and blood pH are the best guide to therapy.

Metabolic Alkalosis

Metabolic alkalosis results from the loss of fixed acids or the gain of bicarbonate and is aggravated by any preexisting potassium depletion. The pH and plasma bicarbonate concentration are elevated. Compensation for metabolic alkalosis is primarily by renal mechanisms; respiratory compensation generally is small and cannot be detected in most patients. Rarely, hypercapnia may represent a compensatory response to metabolic alkalosis in patients without chronic pulmonary disease. When this is suspected, rapid reduction in PaCO₂ by mechanical ventilation should be avoided. Rather, the PaCO₂ will fall as the metabolic alkalosis is corrected.

The majority of patients with metabolic alkalosis have some degree of hypokalemia, a result in part of the influx of potassium ions into the cells as hydrogen ions efflux into the serum. The dangers of metabolic alkalosis are the same as those discussed for respiratory alkalosis.

A problem commonly encountered in the surgical patient is hypochloremic, hypokalemic metabolic alkalosis resulting from persistent vomiting or gastric suction in the patient with pyloric obstruction. Unlike vomiting with an open pylorus, which involves a combined loss of gastric, pancreatic, biliary, and intestinal secretions, vomiting with an obstructed pylorus results in loss of fluid with high chloride and hydrogen ion concentrations in relation to sodium. Initially the urinary excretion of bicarbonate increases to compensate for the alkalosis. This increase in urinary bicarbonate excretion results from net hydrogen ion resorption by the renal tubular cells, with accompanying potassium ion excretion. As the volume deficit progresses, aldosterone-mediated sodium resorption is accompanied by potassium excretion. The resulting hypokalemia leads to excretion of hydrogen ions in place of potassium ions by this mechanism, producing paradoxical aciduria. The net result is a self-perpetuating alkalosis with hypokalemia. Proper management includes replacement of the extracellular fluid volume deficit with isotonic sodium chloride solution in addition to replacement of potassium. Volume repletion should be started and a good urine output obtained before potassium is administered.

Rarely, severe hypokalemic metabolic alkalosis in a patient with pyloric outlet obstruction may be refractory to standard therapy. This occurs most often in patients who also have severe hypochloremia and several liters of nasogastric drainage daily. In the past, the infusion of ammonium chloride or arginine hydrochloride was the usual method for increasing the level of nonvolatile acids. However, infusion of ammonium chloride may produce ammonia toxicity, and arginine hydrochloride is no longer available commercially. The use of 0.1 N to 0.2 N hydrochloric acid is a safe and effective therapy for correction of severe, resistant metabolic alkalosis. The infusion should be administered over a period of 6 to 24 h, with measurements of pH, PaCO₂, and serum electrolytes every 4 h. Generally, 1 or 2 L of solution over a period of 24 h is sufficient, but additional hydrochloric acid should be infused when indicated by appropriate clinical and laboratory evidence. Temporary control of the alkalosis with this method usually is successful, but the underlying cause should be controlled as soon as possible.

Potassium Abnormalities

The normal dietary intake of potassium is approximately 50 to 100 mEq daily, and in the absence of hypokalemia, the majority of this is excreted in the urine. Ninety-eight percent of the potassium in the body is located within the intracellular compartment at a concentration of approximately 150 mEq/L, and it is the major cation of intracellular water. Although the total extracellular potassium in a 70-kg male would amount approximately to only 63 mEq (4.5 mEq/L × 14 L), this small amount is critical to cardiac and neuromuscular function. In addition, the turnover rate in the extracellular fluid compartment may be extremely rapid.

The intracellular and extracellular distribution of potassium is influenced by many factors. Significant quantities of intracellular potassium are released into the extracellular space in response to severe injury or surgical stress, acidosis, and the catabolic state. A significant rise in serum potassium concentration may occur in these states in the presence of oliguric or anuric renal failure, but dangerous hyperkalemia (greater than 6mEq/L) rarely is encountered when renal function is normal. After severe trauma normal or excessive urinary volumes may not reflect the kidney's ability to clear solutes or to excrete potassium.

Hyperkalemia

The signs of a significant hyperkalemia are limited to the cardiovascular and gastrointestinal systems. The gastrointestinal symptoms include nausea, vomiting, intermittent intestinal colic, and diarrhea. The cardiovascular signs are apparent on the electrocardiogram (ECG) initially, with high, peaked T waves, widened QRS complex, and depressed ST segments. Disappearance of T waves, heart block, and diastolic cardiac arrest may develop with increasing levels of potassium.

Treatment of hyperkalemia consists of immediate measures to reduce the serum potassium level, withholding of exogenous potassium, and correction of the underlying cause if possible. Temporary suppression of the myocardial effects of a sudden rapid rise of potassium level can be accomplished by the intravenous administration of 1 g of 10% calcium gluconate under ECG monitoring. Serum potassium levels may be transiently decreased by administration of bicarbonate and glucose with insulin (45 mEq NaHCO₃ in 1000 mL/D₁₀W with 20 units regular insulin), which promote cellular uptake of potassium. However, the definitive treatment of hyperkalemia requires the enteral administration of cation exchange resins (Kayexalate) or dialysis.

Hypokalemia

The more common problem in the surgical patient is hypokalemia, which may occur as a result of: (1) excessive renal excretion; (2) movement of potassium into cells; (3) prolonged administration of potassium-free parenteral fluids with continued obligatory renal loss of potassium (20 mEq/day or more); (4) total parenteral hyperalimentation with inadequate potassium replacement; and (5) loss in gastrointestinal secretions.

Potassium has an important role in the regulation of acid-base balance. Increased renal excretion occurs with respiratory and metabolic alkalosis. Potassium is in competition with hydrogen ions for renal tubular excretion in exchange for sodium ions. Thus, in alkalosis, the increased potassium ion excretion in exchange for sodium ions permits hydrogen ion conservation. Hypokalemia may produce a metabolic alkalosis, because an increase in excretion of hydrogen ions occurs when the concentration of potassium in the tubular cell is low. In addition, movement of hydrogen ions into the cells as a consequence of potassium loss is partly responsible for the alkalosis. In metabolic acidosis the reverse process occurs, and the excess hydrogen ion is exchanged for sodium, with retention of greater amounts of potassium.

Renal tubular excretion of potassium ions is increased when large quantities of sodium are available for excretion. The more sodium ions available for resorption, the more potassium is exchanged for sodium in the lumen. Potassium requirements for prolonged or massive isotonic fluid volume replacement are increased, probably on this basis.

The renal excretion of potassium may be small in comparison with the amount of potassium that may be lost in gastrointestinal secretions. The amount per liter in various types of

gastrointestinal fluids is shown in Table [2-3](#). Although the average potassium concentration of some of these fluids is relatively low, significant hypokalemia will result if potassium-free fluids are used for replacement.

Hypokalemia also may be a serious problem in the patient who is maintained on intravenous nutrition. Large quantities of supplemental potassium generally are necessary to restore depleted intracellular stores and to meet the requirements for tissue synthesis during the anabolic phase.

Most of the factors that influence potassium metabolism result in excess excretion, and a tendency toward hypokalemia occurs frequently in the surgical patient, except when shock or acidosis interferes with the normal renal handling of potassium.

The signs of potassium deficit are related to failure of normal contractility of skeletal, smooth, and cardiac muscle and include weakness that may progress to flaccid paralysis, diminished to absent tendon reflexes, and paralytic ileus. Sensitivity to digitalis, with cardiac arrhythmias and ECG signs of low voltage, flattening of T waves, and depression of ST segments, are characteristic. Signs of potassium deficit may be masked by those of a severe extracellular fluid volume deficit. Repletion of the volume deficit may further aggravate the situation by lowering the serum potassium level through dilution.

Hypokalemia is best dealt with through prevention. In the replacement of gastrointestinal fluids, it is safe to replace the upper limits of loss, because an excess is readily handled by the patient with normal renal function. No more than 40 mEq should be added to a liter of intravenous fluid, and the rate of administration should not exceed 40 mEq/h unless the ECG is being monitored. In the absence of specific indications, potassium should not be given to the oliguric patient or to patients during the first 24 h after severe surgical stress or trauma.

Calcium Abnormalities

The majority of the 1000 to 1200 g of body calcium in the average-sized adult is found in the bone in the form of phosphate and carbonate. Normal daily intake of calcium is 1 to 3 g. Most of this is excreted via the gastrointestinal tract, and 200 mg or less is excreted in the urine daily. The normal serum level is 8.5 to 10.5 mg/dL, about half of which is non-ionized and bound to plasma protein. An additional non-ionized fraction (5 percent) is bound to other substances in the plasma and interstitial fluid, and the remaining 45 percent is the ionized portion that is responsible for neuromuscular stability. Determination of the plasma protein level, therefore, is essential for proper analysis of the serum calcium level. The ratio of ionized to non-ionized calcium is also related to the pH; acidosis causes an increase in the ionized fraction, and alkalosis causes a decrease.

Disturbances of calcium metabolism generally are not a problem in the postoperative patient without complications, with the exception of skeletal loss during prolonged immobilization. Routine administration of calcium to the surgical patient, therefore, is not needed in the absence of specific indications.

Hypocalcemia

The symptoms of hypocalcemia, which may be seen when serum levels are below 8 mg/dL, include numbness and tingling of the circumoral region and the tips of the fingers and toes. The signs are of neuromuscular origin and include hyperactive tendon reflexes, positive Chvostek's sign, muscle and abdominal cramps, tetany with carpopedal spasm, convulsions (with severe deficit), and prolongation of the Q-T interval on the ECG.

The common causes include acute pancreatitis, massive soft-tissue infections (necrotizing fasciitis), acute and chronic renal failure, pancreatic and small-bowel fistulas, and hypoparathyroidism. Transient hypocalcemia is a frequent occurrence in the hyperparathyroid patient after removal of a parathyroid adenoma, owing to atrophy of the remaining glands and avid bone uptake. Asymptomatic hypocalcemia may occur with hypoproteinemia (normal ionized fraction), but symptoms may appear with a normal serum calcium level in a patient with severe alkalosis. In this case there is a decrease in the physiologically active or ionized fraction of total serum calcium. Calcium levels also may fall with a severe depletion of magnesium.

Treatment is directed toward correction of the underlying cause and repletion of the deficit. Acute symptoms may be relieved by the intravenous administration of calcium gluconate or calcium chloride. Calcium lactate may be given orally, with or without supplemental vitamin D, in the patient requiring prolonged replacement. The routine administration of calcium during massive transfusions of blood is controversial and reflects the paucity of studies in which calcium ion levels have been measured. In the majority of studies, calcium ion concentrations have been estimated from measured total serum calcium levels. Available data indicate that the majority of patients receiving blood transfusions do not require calcium supplementation. The binding of ionized calcium by citrate generally is compensated for by the mobilization of calcium from body stores. For patients receiving blood as rapidly as 500 mL every 5 to 10 min, calcium administration is recommended. An appropriate dose, according to Moore's data, is 0.2 g calcium chloride (2 mL 10% calcium chloride solution), administered intravenously in a separate line, for every 500 mL of blood transfused. To avoid dangerous levels of hypercalcemia, this dose of calcium is recommended only while blood is being transfused at the rate noted above. The total dose of calcium generally should not exceed 3 g unless there is objective evidence of hypocalcemia. Larger doses rarely are indicated because there is some mobilization of calcium and citrate breakdown with release of calcium ions even with shock and inadequate peripheral perfusion. During massive transfusions, some attempt should be made to monitor the calcium level. An approximation of calcium ion concentration can be obtained by monitoring the Q-T interval on the ECG, although techniques for the rapid measurement of calcium ion concentration are available.

Hypercalcemia

The symptoms of hypercalcemia are vague and of gastrointestinal, renal, musculoskeletal, and central nervous system origin. The early manifestations of hypercalcemia include easy fatigue, lassitude, weakness of varying degree, anorexia, nausea, vomiting, and weight loss. With higher serum calcium levels, lassitude gives way to somnambulism, stupor, and, finally, coma. Other symptoms include severe headaches, pains in the back and extremities, thirst, polydipsia, and polyuria. The critical level for serum calcium is 15 mg/dL or higher, and unless treatment is instituted promptly, the symptoms may rapidly progress to death. The two major causes of hypercalcemia are hyperparathyroidism and cancer with bony metastasis. The latter is seen most frequently in the patient with metastatic breast cancer who is receiving estrogen therapy.

A serum calcium concentration of 15 mg/dL or higher requires emergency treatment. Most patients have an extracellular fluid volume deficit because of the effects of hypercalcemia (vomiting, polyuria), and vigorous volume repletion with salt solutions lowers the calcium level by dilution and increased urinary calcium excretion. Rapid correction of the associated extracellular fluid volume deficit immediately lowers the serum calcium level by dilution and by increased renal clearance, which may be augmented by furosemide administration.

Oral or intravenous inorganic phosphates effectively lower serum calcium by inhibiting bone resorption and forming calcium-phosphate complexes that are deposited in soft tissues and bone. Intravenous use may cause an abrupt fall in calcium, and tetany, hypotension, and acute renal failure have been reported with this form of therapy. If intravenous phosphorus is used, it should be given slowly over a period of approximately 12 h once daily for no more than 2 or 3 days. Inorganic phosphates are contraindicated in patients with hyperphosphatemia or renal failure. Intravenous sodium sulfate also lowers serum calcium by increasing urinary excretion of calcium. It is less effective than phosphate salts, however, and probably is no more effective than normal saline.

Corticosteroids decrease resorption of calcium from bone and reduce the intestinal absorption of vitamin D. They have been useful in treating hypercalcemic patients with sarcoidosis, myelomas, lymphomas, and leukemias, although the reduction in serum calcium may not be apparent for 1 or 2 weeks. Mithramycin, a cytotoxic drug, effectively lowers serum calcium in 24 to 48 h by direct action on the bones. The drug is relatively safe in the small doses used, and the calcium level may remain normal for several days to weeks after a single dose. Calcitonin induces a moderate decrease in serum calcium, but the effect is diminished with repeated administration. The definitive treatment of acute hypercalcemic crisis in patients with hyperparathyroidism is immediate surgery.

In the patient with metastatic cancer, a preventive approach to hypercalcemia should be taken. The serum calcium level is checked frequently; if it is elevated, the patient is placed on a low-calcium diet and measures to ensure adequate hydration are instituted.

Magnesium Abnormalities

The total body content of magnesium in the average adult is approximately 2000 mEq, about half of which is incorporated in bone and only slowly exchangeable. The distribution of magnesium is similar to that of potassium, the major portion being intracellular. Serum magnesium concentration normally ranges from 1.5 to 2.5 mEq/L. The normal dietary intake of magnesium is approximately 20 mEq (240 mg) daily. The larger part is excreted in the feces and the remainder in the urine. The kidneys show a remarkable ability to conserve magnesium; on a magnesium-free diet, renal excretion of this ion may be less than 1mEq/day.

Magnesium Deficiency

Magnesium deficiency is known to occur with starvation, malabsorption syndromes, protracted losses of gastrointestinal fluid, prolonged intravenous fluid therapy with magnesium-free solutions, and during total parenteral nutrition when inadequate quantities of magnesium have been added to the solutions. Other causes include acute pancreatitis, treatment of diabetic ketoacidosis, primary aldosteronism, chronic alcoholism, amphotericin B therapy, and a protracted course after thermal injury.

The magnesium ion is essential for proper function of most enzyme systems, and depletion is characterized by neuromuscular and central nervous system hyperactivity. The signs and symptoms are similar to those of calcium deficiency, including hyperactive tendon reflexes, muscle tremors, and tetany with a positive Chvostek's sign. Progression to delirium and convulsions may occur with a severe deficit. A concomitant calcium deficiency occasionally is noted and is refractory to treatment in the absence of magnesium repletion.

The diagnosis of magnesium deficiency depends on an awareness of the syndrome and clinical recognition of the symptoms. Laboratory confirmation is available, but it is not reliable because the syndrome may exist in the presence of a normal serum magnesium level. The possibility of magnesium deficiency should always be considered in the surgical patient who exhibits disturbed neuromuscular or cerebral activity in the postoperative period. This is particularly important in patients who have had protracted dysfunction of the gastrointestinal tract with long-term maintenance on parenteral fluids and in patients on parenteral hyperalimentation. Routine magnesium administration is always indicated in the management of these patients.

Treatment of magnesium deficiency is by the parenteral administration of magnesium sulfate or magnesium chloride solution. If renal function is normal, as much as 2 mEq of magnesium per kg of body weight per day can be administered intravenously or intramuscularly daily in the face of severe depletion. The intravenous route is preferable for the initial treatment of a severe symptomatic deficit. The solution is prepared by the addition of 80 mEq of magnesium sulfate (20 mL 50% solution containing 4 mEq/mL magnesium) to 1 L of intravenous fluid and is administered over a 4-h period. If the patient is not symptomatic, the infusion should be given over a longer period. The possibility of acute magnesium toxicity should be kept in mind when giving magnesium intravenously. When large doses are given, the heart rate, blood pressure, respiration, and ECG should be monitored closely for signs of magnesium toxicity, which could lead to cardiac arrest. It is advisable to have calcium chloride or calcium gluconate available to counteract any adverse effects of a rapidly rising serum magnesium level.

Partial or complete relief of symptoms may follow this infusion as a result of increased concentration of magnesium ion in the extracellular fluid compartment, although continued

replacement over a period of 1 to 3 weeks is necessary to replenish the intracellular compartment. For this purpose and for the asymptomatic patient who may have significant magnesium depletion, 10 to 20 mEq of 50% magnesium sulfate solution per day may be given intramuscularly or in intravenous fluids; alternatively, 800 mg magnesium oxide per day may be given orally. When intramuscular magnesium sulfate is used, it should be given in divided doses or at multiple sites because the intramuscular injection of this salt is painful. After complete repletion of intracellular magnesium, and in the absence of abnormal loss, balance may be maintained by the administration of as little as 4 mEq of magnesium ion daily. The amount of magnesium supplementation required for patients on parenteral hyperalimentation varies but 12 to 24 mEq daily is effective for the average patient.

Magnesium should not be given to the oliguric patient or in the presence of a severe volume deficit unless actual magnesium depletion has been demonstrated. If given to a patient with renal insufficiency, considerably smaller doses are used and the patient is carefully observed for signs or symptoms of toxicity.

Magnesium Excess

Symptomatic hypermagnesemia, although rare, is most commonly seen with severe renal insufficiency. Retention and accumulation of magnesium may occur in any patient with impaired glomerular or renal tubular function, and the presence of acidosis may rapidly compound the problem. Serum magnesium levels tend to parallel changes in potassium concentration in these cases. In patients on ordinary dietary intakes of magnesium, increased serum concentrations do not occur until the glomerular filtration rate falls below 30 mL/min. Magnesium-containing antacids and laxatives (milk of magnesia, epsom salts, Gelusil, Maalox) are commonly administered in quantities that are sufficient to produce toxic serum levels of magnesium in patients with impaired renal function. Other conditions that may be associated with symptomatic hypermagnesemia include early thermal injury, massive trauma or surgical stress, severe extracellular volume deficit, and severe acidosis.

The early signs and symptoms of magnesium excess include lethargy and weakness with progressive loss of deep tendon reflexes. Interference with cardiac conduction occurs with increasing levels of magnesium, and changes in the ECG (increased P-R interval, widened QRS complex, and elevated T waves) resemble those seen in hyperkalemia. Somnolence leading to coma and muscular paralysis occur in the later stages, and death is usually caused by respiratory or cardiac arrest.

Treatment consists of immediate measures to lower the serum magnesium level by correcting coexisting acidosis, replenishing preexisting extracellular volume deficit, and withholding exogenous magnesium. Acute symptoms may be temporarily controlled by the slow intravenous administration of 5 to 10 mEq of calcium chloride or calcium gluconate. If elevated levels or symptoms persist, peritoneal dialysis or hemodialysis is indicated.

FLUID AND ELECTROLYTE THERAPY

Parenteral Solutions

Many different electrolyte solutions, with various compositions, are available for parenteral administration (Table 2-9). Several of the more commonly used solutions are discussed below. The choice of a particular fluid depends on the patient's volume status and the type of concentration or compositional abnormality present.

A cost-effective extracellular "mimic" of isotonic salt solution for replacing gastrointestinal losses and extracellular fluid volume deficits, in the absence of gross abnormalities of concentration and composition, is lactated Ringer's solution. This solution is physiologic and contains 130 mEq of sodium balanced by 109 mEq of chloride and 28 mEq of lactate. Lactate is used instead of bicarbonate because it is more stable in intravenous fluids during storage. The lactate is converted readily to bicarbonate by the liver after infusion. Concern about the ability of the liver to metabolize lactate is unwarranted even when infusing large quantities of lactated Ringer's solution to patients in hemorrhagic shock. This fluid has minimal effects on normal body fluid composition and pH even when infused in large quantities. There are other balanced salt solutions available, some with sodium acetate or bicarbonate instead of lactate; all are considered interchangeable.

Isotonic sodium chloride contains 154 mEq of sodium and 154 mEq of chloride per liter. The high concentration of chloride above the normal serum concentration of 103 mEq/L imposes on the kidneys an appreciable load of excess chloride that cannot be rapidly excreted. A dilutional acidosis may develop by reducing base bicarbonate relative to carbonic acid. This solution is ideal, however, for the initial correction of an extracellular fluid volume deficit in the presence of hyponatremia, hypochloremia, and metabolic alkalosis. In a similar situation with moderate metabolic acidosis, M/6 sodium lactate (167 mEq/L each of sodium and lactate) may be given.

For maintenance fluid in the postoperative period, 0.45% sodium chloride in 5% dextrose solution is used often to provide free water for insensible losses and some sodium for renal adjustment of serum concentration. With added potassium, this is a reasonable solution to use for maintenance requirements in a patient with no complications who requires only a short period of parenteral fluids.

Preoperative Fluid Therapy

Preoperative evaluation and correction of existing fluid disorders is an integral part of surgical care. An orderly approach to these problems requires an understanding of the common fluid disturbances associated with surgical illness and adherence to a few simple guidelines.

The analysis of a particular fluid disorder may be facilitated by categorizing the abnormalities as volume, concentration, and compositional changes. Although some disease states produce characteristic changes in fluid balance, much confusion may be avoided by regarding each disturbance as a separate entity. There are no shortcuts; close observation of the patient and frequent reevaluation of the clinical situation is the most rewarding approach. For example, volume changes cannot be accurately predicted from a knowledge of the level of serum sodium because an extracellular fluid volume deficit or excess may exist with a normal, low, or high sodium concentration. Similarly, any of the four primary acid-base disturbances may be associated with any combination of volume and concentration abnormalities.

Correction of Volume Changes

Changes in the volume of extracellular fluid are the most frequent and important abnormalities encountered in the surgical patient. Depletion of the extracellular fluid compartment without changes in concentration or composition is a common problem. The diagnosis of volume changes is made almost entirely on clinical grounds. The signs that will be present in an individual patient depend not only on the relative or absolute quantity of extracellular fluid that has been lost but also on the rapidity with which it is lost and the presence or absence of signs of associated disease.

Volume deficits in the surgical patient may result from external loss of fluids or from an internal redistribution of extracellular fluid into a nonfunctional compartment. It usually involves a combination of the two, but the internal redistribution frequently is overlooked.

The phenomenon of internal redistribution or translocation of extracellular fluid is peculiar to many surgical diseases; in the individual patient, the loss may be quite large. Although the concept of a "third space" is not new, it usually is considered only in relation to patients with massive ascites, burns, or crush injuries. Of more importance is the third-space loss into the peritoneum, the bowel wall, and other tissues with inflammatory lesions of the intraabdominal organs. The magnitude of these losses may not be fully appreciated without realization that the peritoneum alone has approximately 1.8 m² of surface area. A slight increase in thickness from sequestration of fluid, which would not be appreciated on casual observation, may result in a functional loss of several liters of fluid. Swelling of the bowel wall and mesentery and secretion of fluid into the lumen of the bowel cause even larger losses. Similar deficits may occur with massive infection of the subcutaneous tissues (necrotizing fasciitis) or with severe crush injury.

These "parasitic" losses remain a part of the extracellular fluid space and may be measured as a slowly equilibrating volume. The term *nonfunctional* is used because the fluid is no longer able to participate in the normal functions of the extracellular compartment and may just as well have been lost externally. Any transfer of intracellular fluid to the extracellular compartment for replenishment of the loss is insignificant in the acute phase. The patient with ascites may have an enormous total extracellular fluid volume, but the functional component is severely depleted. The same is true of extensive inflammatory or obstructive lesions of the gastrointestinal tract, although the loss is not as obvious. These losses evoke the signs and symptoms of an extracellular fluid volume deficit with or without the concomitant external loss of fluids.

Exact quantification of these deficits is impossible and probably unnecessary. The defect can be estimated on the basis of the severity of the clinical signs. A mild deficit represents a loss of approximately 4 percent of body weight; a moderate loss is 6 to 8 percent of body weight; and a severe deficit is approximately 10 percent of body weight. Cardiovascular signs predominate when there is acute rapid loss of fluid from the extracellular fluid compartment with few or no tissue signs. In addition to the estimated deficit, fluids lost during treatment must be replaced.

Immediately after diagnosis of a volume deficit, prompt fluid replacement with a balanced salt solution should be started. Continuing therapy is tailored to the response of the patient, based on frequent clinical examination. Reliance on a formula or a single clinical sign to determine the adequacy of resuscitation is perilous. Rather, reversal of the signs of the volume deficit, combined with stabilization of the blood pressure and pulse and an hourly urine volume of 30 to 50 mL are used as general guidelines. An adequate hourly urine output, although usually a reliable index of volume replacement, may be totally misleading. The excessive administration of glucose (over 50 g in a 2- to 3-h period) may result in osmotic diuresis, while an osmotic agent such as mannitol tends to produce urine at the expense of the vascular volume. Patients with chronic renal disease or incipient acute renal damage from shock and injury also may have inappropriately high urinary volumes. In addition, the rapid administration of salt solutions may transiently expand the intravascular volume, increase the glomerular filtration rate, and result in an immediate outpouring of urine, although the total extracellular fluid space remains quite depleted.

The choice of the proper fluid for replacement depends on concomitant concentration or compositional abnormalities. With pure extracellular fluid volume loss or when only minimal concentration or compositional abnormalities are present, the use of a balanced salt solution, such as lactated Ringer's, is desirable.

Rate of Fluid Administration

The rate of fluid administration varies considerably, depending on the severity and type of fluid disturbance, the presence of continuing losses, and the cardiac status. The most severe volume deficits may be safely replaced initially with isotonic solutions at rates up to 2000 mL/h, with the rate reduced as the fluid status improves. Constant observation by a physician is mandatory when the administration exceeds 1000 mL/h. At these rates a significant portion may be lost as urinary output because of a transient overexpansion of the plasma volume.

In elderly patients, associated cardiovascular disorders do not preclude correction of existing volume deficits, but they do necessitate slower, more careful correction with constant monitoring of the cardiopulmonary system. If urinary output is not promptly restored, measurements of central filling pressures and cardiac output may be required in order to prevent renal injury from insufficient volume restoration.

Correction of Concentration Changes

If severe symptomatic hyponatremia or hypernatremia complicates the volume loss, prompt correction of the concentration abnormality to the extent that symptoms are relieved is

necessary. Volume replenishment should be accomplished with slower correction of the remaining concentration abnormality. For immediate correction of severe hyponatremia, 5% sodium chloride solution or molar sodium lactate solution is used, depending on the patient's acid-base status. In any case, the sodium deficit can be estimated by multiplying the decrease in serum sodium concentration below normal (in mEq/L) by the total body water (in L). Initially, up to one-half of the calculated amount of sodium may be administered slowly, followed by clinical and chemical reevaluation of the patient before any additional infusion of sodium salts.

This estimate is based on total body water, because the effective osmotic pressure in the extracellular compartment cannot be increased without increasing this function proportionally in the intracellular compartment. Although absolute reliance on any formula is undesirable, proper use of this estimate allows a safe quantitative approximation of the sodium deficit. Only a portion of the total deficit is replaced initially to relieve acute symptoms. Further correction is facilitated when renal function is restored by correction of the volume deficit. If the total calculated deficit were given rapidly, severe hypervolemia might occur, particularly in patients with limited cardiac reserve. The infusion of small, successive increments of hypertonic saline solution with frequent evaluation of the clinical response and serum sodium concentration is recommended.

In the treatment of moderate hyponatremia with an associated volume deficit, volume replacement can be started immediately with concomitant correction of the serum sodium deficit. Isotonic sodium chloride solution (normal saline) is used initially in the presence of metabolic alkalosis, and M/6 sodium lactate (167 mEq/L each of sodium and lactate) is used to correct an associated acidosis. Correction of the serum sodium concentration may require only a few liters of these solutions; the remainder of the volume deficit may be replaced with lactated Ringer's solution.

Treatment of hyponatremia associated with volume excess is by restriction of water. In the presence of severe symptomatic hyponatremia, a small amount of hypertonic salt solution may be infused cautiously to alleviate symptoms. Because this will cause additional volume expansion, it is contraindicated in patients with limited cardiac reserve; peritoneal dialysis or hemodialysis is preferred in this situation.

For the correction of severe, symptomatic hypernatremia with an associated volume deficit, 5% dextrose in water may be infused slowly until symptoms are relieved. If the extracellular osmolality is reduced too rapidly, however, convulsions and coma may result. For this reason, correction of hypernatremia concomitant with repletion of the volume deficit by half-strength sodium chloride or half-strength lactated Ringer's solution is safer in most cases. In the absence of a significant volume deficit, water should be administered cautiously since dangerous hypervolemia may result; constant observation and frequent determinations of the serum sodium concentration are indicated. The problem is simplified when a sufficient quantity of fluid has been given to permit renal excretion of the solute load.

Composition and Miscellaneous Considerations

Correction of existing potassium deficits should be started after an adequate urine output is obtained, particularly in the patient with metabolic alkalosis because this may be secondary to or aggravated by potassium depletion. Calcium and magnesium rarely are needed during preoperative resuscitation but should be given as indicated, particularly to patients with massive subcutaneous infections, acute pancreatitis, or chronic starvation.

Fluid abnormalities also must be suspected in the patient for whom an elective procedure is planned. Chronic illnesses frequently are associated with extracellular fluid volume deficits, and concentration and compositional changes are not uncommon. Correction of anemia and recognition of the fact that a concentrated blood volume may exist in the chronically debilitated patient is of obvious importance. The hematocrit level increases approximately 3 percent after the infusion of one unit of packed red blood cells into the adult of average size. The increase may be significantly greater in the patient with a contracted intravascular volume, indicating the need for concurrent volume replacement.

The prevention of volume depletion during the preoperative period is important. Prolonged periods of fluid restriction in preparation for various diagnostic procedures, and the use of cathartics and enemas for preparation of the bowel may cause a significant acute loss of extracellular fluid. Prompt recognition and treatment of these losses is necessary to prevent complications during the operative period.

Intraoperative Fluid Management

If preoperative replacement of extracellular fluid volume has been incomplete, hypotension may develop promptly with the induction of anesthesia. This can be insidious because the ability of the awake patient to compensate for a mild volume deficit is revealed only when the compensatory mechanisms are abolished with anesthesia. This problem is prevented by maintaining baseline requirements and replacing abnormal losses of fluids and electrolytes by intravenous infusions in the preoperative period.

In addition to blood losses during operation, there appear to be extracellular fluid losses during major operative procedures. Some of these, including edema from extensive dissection, collections within the lumen and wall of the small bowel, and accumulations of fluid in the peritoneal cavity, are clinically discernible and well recognized. They are believed to represent distributional shifts, in that the functional volume of extracellular fluid is reduced, but not externally lost from the body. These functional losses often are referred to as "parasitic losses," "third space edema," or "sequestration" of extracellular fluid. Another source of extracellular fluid loss during major operative trauma is the wound, though this is a smaller loss and difficult to quantify except in extensive and major operative procedures.

At the beginning of the twentieth century surgeons became aware that many changes occurred in urinary output, blood volume, and fluid and electrolyte composition during and after surgery. Assessment of these changes awaited the development of analytic techniques and their application to patient studies. In the following 25 years, saline solutions in varying combinations were given to patients undergoing operation, often in excessive amounts. Work in the late 1930s and early 1940s by Moyer and others indicated that during and after operative procedures, saline and water solutions should be withheld entirely because most of the fluid administered is retained.

The possibility was recognized that the operative and postoperative retention of salt and water administered in relatively small amounts might be physiologic retention to replace a deficit of salt and water incurred by the operative procedure. Subsequent studies revealed that functional extracellular fluid decreases with major abdominal operations, largely as sequestered loss into the operative site. This extracellular fluid volume deficit can be replaced during the operative procedure. These data led to the conclusion that the need for an extracellular "mimic" in the form of balanced salt solution can be clinically estimated. Intraoperative correction of the volume deficit with salt solution markedly reduces postoperative oliguria but is not intended to substitute for blood replacement. It is believed to be a physiologic supplement, or an adjunct, to replace sequestered losses.

The pendulum thus swung from the indiscriminate use of salt solutions in the first quarter of the twentieth century to almost total withholding of fluid and electrolytes from surgical patients in the second quarter of the century; today indications are that proper management lies between these two extremes. Some guidelines are necessary for the intraoperative administration of saline solutions as a "mimic" for the sequestered extracellular fluid. Because this varies from an almost imperceptible minimum to a high of approximately 3 L during an uncomplicated procedure, quantification is extremely difficult with available means of measuring functional extracellular fluid. Consequently, no accurate formula for intraoperative fluid administration can be derived. Some arbitrary but clinically useful guidelines are as follows: (1) Blood should be replaced to maintain an acceptable red blood cell mass irrespective of any additional fluid and electrolyte therapy. (2) The replacement of extracellular fluid should begin during the operative procedure. (3) Balanced salt solution needed during operation is approximately 0.5 to 1 L/h, but only to a maximum of 2 to 3 L during a 4-h major abdominal procedure, unless there are other measurable losses.

Using a similar fluid regimen, Thompson and associates reported experiences in a series of 670 patients undergoing major aortoiliac reconstructive procedures. In this group of patients, the average amount of lactated Ringer's solution administered was 3555 mL, giving an average intraoperative replacement of salt solution of 677 mL/h of operative procedure. In the last 6 years of this study there were only two deaths in 298 operations, an operative mortality of 0.67 percent. Among all 670 patients, only two patients died of renal failure, an incidence of 0.3 percent. No patient died of pulmonary insufficiency. This extremely low incidence of renal failure, even in the presence of extensive operative trauma, is similar to the authors' data for major abdominal operative procedures.

Data reported by Virgilio and others indicate that in the previously healthy surgical patient, the addition of albumin to intraoperative blood and extracellular fluid replacement is not only unnecessary but also potentially harmful. Data by Shires on operative measurements of cardiac function and extravascular lung water indicate that optimal function is obtained with replacement of blood and an extracellular "mimic" without the addition of extra albumin.

The addition of crystalloid fluid resuscitation, in appropriate volume, to blood replacement in the past quarter century has markedly improved the ability to maintain intraoperative homeostasis and avoid organ injury associated with inadequate volume replacement.

Postoperative Fluid Management

Immediate Postoperative Period

Orders for postoperative fluids are not written until the patient is in the recovery room and the fluid status has been assessed. Evaluation at this point should include a review of preoperative fluid status, the amount of fluid loss and gain during operation, and clinical examination of the patient with assessment of the vital signs and urinary output. Initial fluid orders are written to correct any *existing* deficit, followed by maintenance fluids for the remainder of the day. For the patient with complications who has received or lost large amounts of fluid, it is frequently difficult to estimate what the fluid requirements will be for the first 24 h after operation; hence intravenous fluids are ordered 1 L at a time, and the patient is checked frequently until the situation is clarified. Proper replacement of fluids during this relatively short period facilitates subsequent fluid management.

Immediately after operation, extracellular fluid volume depletion may occur as a result of continued losses of fluid at the site of injury or operative trauma, e.g., into the wall or lumen of the small intestine. Several liters of extracellular fluid may be slowly deposited in such areas within a few hours or over the first postoperative day or so. Unrecognized deficits of extracellular fluid volume during the early postoperative period are manifested primarily as circulatory instability. The signs of volume deficiency in other organ systems may be delayed for several hours with this type of fluid loss. Postoperative hypotension and tachycardia require prompt investigation and appropriate therapy. The usually accepted adequate blood pressure of 90/60 mmHg and a pulse of less than 120 in postoperative patients may not be sufficient to prevent renal ischemia unless, in addition to lack of signs of shock, urine flow is adequate (30 to 50 mL/hour). Evaluation of the level of consciousness, pupillary size, airway patency, breathing patterns, pulse rate and volume, skin warmth and color, body temperature, urine output, and a critical review of the operative procedure and the operative fluid management, are recommended. Because operative trauma frequently involves loss or transfer of significant quantities of whole blood, plasma, or extracellular fluid that can be only grossly estimated, circulatory instability is most commonly caused by underestimated initial losses or insidious, concealed continued losses. Operative blood loss usually is estimated by the operating surgeon to be 15 to 40 percent less than the isotopically measured blood loss from that patient. For a patient with circulatory instability, further volume replacement of an additional 1000 mL isotonic salt solution, while determining whether continuing losses or other causes are present, often resolves the problem.

It is unnecessary and probably unwise to administer potassium during the first 24 h after operation unless a definite potassium deficit exists. This is particularly important for the patient subjected to prolonged operative trauma involving one or more episodes of hypotension and for the posttraumatic patient with hemorrhagic hypotension. Oliguric renal failure or the more insidious high-output renal failure may develop, and the administration of even a small quantity of potassium may be detrimental.

Later Postoperative Period

The problem of volume management during the postoperative convalescent phase is one of accurate measurement and replacement of all losses. In the otherwise healthy individual, this involves the replacement of measured sensible losses, which usually are of gastrointestinal origin, and the estimation and replacement of insensible losses.

The insensible loss usually is relatively constant and averages 600 mL/day. This may be increased by hypermetabolism, hyperventilation, and fever to a maximum of approximately 1500 mL/day. The estimated insensible loss is replaced with 5% dextrose in water. This loss may be partially offset by an insensible gain of water from excessive tissue catabolism in the postoperative patient with complications, particularly if associated with oliguric renal failure.

Approximately 1 L of fluid should be given to replace that volume of urine required to excrete the catabolic end products of metabolism (800 to 1000 mL/day). In the individual with normal renal function, this may be given as 5% dextrose in water because the kidneys are able to conserve sodium with excretion of less than 1 mEq daily. It is probably unnecessary to stress the kidneys to this degree, however, and a small amount of salt solution may be given in addition to water to cover urinary loss. In elderly patients with salt-losing kidneys or in patients with head injuries, an insidious hyponatremia may develop if urinary losses are replaced with water. Urinary sodium in these circumstances may exceed 100 mEq/L and result in a daily loss of significant amounts of sodium. Measurement of urinary sodium facilitates accurate replacement.

Urine volume is not replaced on a milliliter-for-milliliter basis. A urinary output of 2000 to 3000 mL on a given day may represent diuresis of fluids given during surgery or may represent excessive fluid administration. If these large losses are completely replaced, the urine output progressively increases, and this may proceed to a situation resembling diabetes insipidus, with urinary outputs in excess of 10 L/day.

Sensible losses, by definition, can be measured or, as in the case of sweating, estimated. Gastrointestinal losses usually are isotonic or slightly hypotonic, and they are replaced with an essentially isotonic salt solution. When the estimated loss is slightly above or below isotonicity, appropriate corrections can be made in the daily water administration, while isotonic salt solutions are used to replace these losses volume for volume. Sweating usually is not a problem except with the febrile patient, in whom losses may, but seldom do, exceed 250 mL/day per degree of fever. Excessive sweating may represent a considerable loss of sodium in the unacclimatized individual.

Determination of serum electrolyte levels usually is unnecessary in the patient with an uncomplicated postoperative course maintained on parenteral fluids for 2 to 3 days. A more prolonged period of parenteral replacement or one complicated by excessive fluid losses requires frequent determinations of the serum sodium, potassium, and chloride levels and of carbon dioxide combining power. Adjustments then can be made with intravenous fluids of appropriate composition.

Daily maintenance fluid should be administered at a steady rate while the losses are incurred. If given over a shorter period, renal excretion of the excess salt and water may occur while the normal losses continue over the full 24-h period. For the same reason, fluids of different composition are alternated, and additives to intravenous fluids (e.g., potassium chloride and antibiotics) are evenly distributed in the total volume of fluid given.

Daily fluid orders should begin with an assessment of the patient's volume status and a check for possible concentration of compositional disorders as reflected by proper laboratory determinations. All measured and insensible losses are replaced with fluids of appropriate composition, with allowances made for any preexisting deficit or excess. The amount of potassium replacement is 40 mEq daily for renal excretion of potassium in addition to approximately 20 mEq/L for replacement of gastrointestinal losses. Inadequate replacement may prolong the usual postoperative ileus and contribute to the insidious development of a resistant metabolic alkalosis. Calcium and magnesium are replaced when needed.

Special Considerations in the Postoperative Patient Volume Excesses

The administration of isotonic salt solutions in excess of volume losses (external or internal) may result in overexpansion of the extracellular fluid space. The otherwise normal person in a postoperative state tolerates an acute overexpansion extremely well. Excesses administered over several days, however, will soon exceed the kidneys' ability to excrete sodium. It is important to determine accurately, from intake and output records and serum sodium concentrations, the actual needs of the patient managed over several postoperative days. Attention to the signs and symptoms of overload usually prevents volume excess. It occurs most frequently with attempts to meet excessive volume losses that are not measurable, such as those resulting from incompletely controlled fistula drainage.

The earliest sign of volume overload is weight gain during the catabolic period, when the patient should be losing $\frac{1}{4}$ to $\frac{1}{2}$ lb/day. Heavy eyelids, hoarseness, or dyspnea on exertion may appear rapidly. Circulatory and pulmonary signs of overload appear late and represent a massive overload. Peripheral edema may be a sign, but it does not necessarily indicate volume excess. In the absence of additional evidence for volume overload, other causes for peripheral edema should be considered. Overexpansion of the *total* extracellular fluid may coexist with *depletion* of the functional extracellular fluid compartment, along with decreased effective circulating plasma volume.

Hyponatremia

Significant postoperative alterations in serum sodium concentration are infrequent when the fluid resuscitation during operation has included adequate volumes of isotonic salt solutions. The kidneys retain the ability to excrete moderate excesses of salt water administered in the early postoperative period if functional extracellular fluid has been adequately replaced during the operative or immediate postoperative period. Previous studies of sodium balance revealed that patients do excrete sodium after the functional deficit incurred by the shift of extracellular fluid has been replaced. Wright and Gann demonstrated normal capacity to excrete water postoperatively when isotonic salt solutions are administered before a challenge with a water load. The commonly described hyponatremia associated with surgical procedures and traumatic injury is prevented by the replacement of extracellular fluid deficits. The daily maintenance of normal osmolarity is simplified by the replacement of observable losses of sodium content.

Hyponatremia may occur easily when water is given to replace losses of sodium-containing fluids or when water administration consistently exceeds water losses. The latter may occur with oliguria or in association with decreased water loss through the skin and lungs, intracellular shifts of sodium, or the cellular release of excessive amounts of endogenous water. Severe or refractory hyponatremia is unlikely to occur when renal function remains normal.

In the presence of hyperglycemia, determination of the glucose concentration is necessary to evaluate the significance of a depressed serum sodium level. Because glucose does not enter cells by passive diffusion, it exerts an osmotic force in the extracellular compartment. This contribution to osmotic pressure is normally small, but with an elevated glucose concentration the increased osmotic pressure causes the transfer of cellular water into the extracellular compartment, resulting in a dilutional hyponatremia. Hence hyponatremia may be observed when the total effective osmotic pressure in the extracellular compartment is normal or even above normal. Each 100-mg/dL rise above normal in the blood glucose level results in a decrease in the serum sodium concentration of 1.6 to 3 mEq/L.

Endogenous Water Release

The patient maintained on intravenous fluids without adequate caloric intake will, between the fifth and tenth days, gain significant quantities of water (maximum 500 mL/day) from excessive cellular catabolism, thereby decreasing the quantity of exogenous water required per day.

Intracellular Shifts

Systemic bacterial sepsis often is accompanied by a precipitous drop in serum sodium concentration. This sudden change is poorly understood, but it usually accompanies loss of extracellular fluid as interstitial or intracellular sequestrations. The condition can be treated by withholding free water, restoring extracellular fluid volume, and initiating treatment of the sepsis.

Hypernatremia

Hypernatremia (serum sodium concentration above 150 mEq/L), although uncommon, is a dangerous abnormality. Unlike decreased serum sodium concentration, hypernatremia is produced easily when renal function is normal. The extracellular fluid hyperosmolarity results in a shift of intracellular water from within the cell to the extracellular fluid compartment; in this situation, a high serum sodium level may indicate a significant deficit of total body water. In surgical patients hypernatremia is most often the result of excessive or unexpected water losses, but it may result from use of salt-containing solutions to replace water losses. Classification of water losses may be helpful in preventing and treating this abnormality.

Excessive Extrarenal Water Losses

With increased metabolism from any cause, but particularly when it is associated with fever, the water loss through evaporation of sweat may reach a level of several liters daily. Patients with a tracheostomy in a dry environment can (with high minute volumes) lose as much as 1 to 1.5 L of water/day by this route. Increased water evaporation from a granulating surface is of significant magnitude in the thermally injured patient, with losses as great as 3 to 5 L/day.

Increased Renal Water Losses

Extremely large volumes of solute-poor urine may result from hypoxic damage to the distal tubules and collecting ducts or loss of antidiuretic hormone stimulation from damage to the central nervous system. In both instances, facultative water resorption is impaired. The former occurs in high-output renal failure; this is the most common type of renal failure after severe injury or operative trauma. The latter occurs with extensive head injuries accompanied by temporary diabetes insipidus.

Solute Loading

High protein intake may produce an increased osmotic load of urea that necessitates the excretion of large volumes of water. Hypernatremia, azotemia, and extracellular fluid volume deficits follow. These can be prevented by an intake of 7 mL of water per gram of dietary protein.

Excessive glucose administration results in the need for a large volume of water for excretion. Osmotic diuretics, such as mannitol and urea, also result in the obligatory excretion of a large volume of water and increased urinary sodium losses. In addition, isotonic salt solutions, if used to replace pure water losses, rapidly produce hypernatremia.

Acute Renal Failure

Acute renal insufficiency after trauma or surgical stress is a lethal complication. The diagnosis is based on persistent oliguria and chemical evidence of uremia after stabilization of the circulation. The clinical course is characterized by oliguria lasting from several days to several weeks followed by a progressive rise in daily urine volume until the excretory and concentrating functions of the kidney are gradually restored.

Acute renal failure is classified according to its cause as prerenal, renal, or postrenal (Table 2-10). The most common cause is sequestered or third-space loss in the area of the surgical procedure. Shock from blood loss and occlusion of small arteries (e.g., renal artery emboli) also may cause prerenal failure. Renal artery occlusions are less common causes than volume depletion or hypotensive shock. The common intrarenal causes of renal failure in the postoperative patient include endotoxemia, trauma, drugs (such as aminoglycosides), or the generation of pigment delivery to the kidneys (as with myoglobin), or destabilized hemoglobin (as with a cardiac bypass machine). Postrenal causes are

almost always because of obstruction of the ureter, the bladder, or the urethra.

Therapy of acute renal failure after surgery begins with removal of the cause. With prerenal azotemia, for example, correction of the extracellular fluid volume deficit with an extracellular fluid volume mimic, such as lactated Ringer's solution, will correct the oliguria and increase retention of nitrogenous products such as creatinine. With intrarenal causes, correction of sepsis or removal of nephrotoxic drugs such as aminoglycosides is mandatory. Postrenal obstruction is generally the result of stone, tumor, or surgical misadventures resulting in ureteral occlusion, or perhaps benign prostatic hypertrophy. One of the more common causes of postoperative oliguria is a blocked Foley catheter.

Therapy is directed not only at the cause of the renal failure but also at maintaining homeostasis within the biochemical abnormalities that ensue. The significant biochemical abnormalities are as follows:

(1) Metabolic acidosis results from failure of renal excretion of fixed acids and the inability to maintain respiratory compensation.

(2) Hyperkalemia is a result of large amounts of intracellular potassium released in acute renal failure. Serum potassium level is elevated by the systemic acidosis and must be corrected early.

(3) Hyponatremia is a commonly observed electrolyte disturbance in patients with acute renal failure. Production of metabolic water from metabolism of nutrients and liberation of water from intracellular breakdown contribute to the excess of free water.

(4) Hyperphosphatemia and hypocalcemia usually evolve in patients with posttraumatic acute renal failure because of inadequate excretion and excessive release from injured tissue.

(5) Hypermagnesemia occurs regularly because the kidney is the major organ for regulating magnesium balance. Magnesium levels can rise rapidly in patients with acute renal failure, particularly if magnesium-containing preparations such as antacids are administered.

Predisposing Factors

A number of specific factors may contribute to the development of postoperative renal failure, including the following:

Trauma

Trauma generally contributes to acute renal failure in the surgical patient as a result of the hypovolemic shock caused by blood loss. In addition, myoglobinuria may accompany severe trauma if there has been significant crush injury with rhabdomyolysis. Similarly, if the patient has received blood transfusions after trauma, small amounts of incompatible blood will produce intravascular hemolysis and hemoglobinuria, leading to acute renal failure. Extracellular fluid volume depletion is a common complicating factor in the trauma patient because of the development of prolonged ileus, peritonitis, or sequestration of fluid at a site of injury producing hypovolemia. The development of sepsis is a further contributing factor in these patients.

Sepsis

The onset of sepsis from any cause, including specific infections, such as urinary, biliary, or intraperitoneal sepsis and contamination from colon procedures or from severe trauma, causes acute renal failure. Endotoxin from any source is the priming agent for the release of endogenously produced cytokines, such as tumor necrosis factor (TNF), that have been shown clearly to produce acute renal failure. It must be remembered that nephrotoxic antibiotics that are used to treat sepsis may themselves cause or worsen acute renal failure.

Cardiopulmonary bypass

As many as 5 to 25 percent of patients who have had prolonged cardiopulmonary bypass have oliguric renal failure or the less frequently lethal nonoliguric renal failure. This is most likely from hypoperfusion of the kidneys.

Renal transplantation

Failure of a transplanted kidney to function in the early postoperative period should raise questions concerning technical problems with the renal artery, obstruction of urinary flow, or intravascular volume problems because of continued bleeding in the patient. Hyperacute rejection, which is uncommon, also should be considered.

Urologic surgery

Urologic procedures pose additional and specific problems for development of acute renal failure after surgery. These problems usually are caused by obstruction in one form or another. Obstruction may occur at the level of the kidney, ureter, bladder, or urethra as a result of urologic procedures. Removal of the obstruction or cause of obstruction should relieve the acute renal failure. The presence of obstruction can be determined by radiographic, tomographic, and computed tomography studies.

Vascular disease

If the blood flow to the kidney is interrupted for a prolonged period, as in operations on the abdominal aorta or the renal artery, acute renal failure may result. Acute renal failure may be delayed because of an obstruction from an organizing hematoma around a vascular anastomosis or ureteral strictures at the site of the ureteral devascularization. Similarly, immediate postoperative hemorrhage may cause hypovolemia and acute renal failure.

Preexisting renal disease

This can contribute to postoperative renal failure. When the kidney is diseased because of nephrosclerosis, diabetes, chronic glomerulonephritis, or chronic tubular interstitial nephritis, the organ is predisposed to the development of acute renal failure.

Radiographic contrast agents

When used preoperatively, contrast agents can be a predisposing factor; a transient reduction in renal function might have occurred preoperatively.

Drugs

A number of drugs can lead to postoperative renal toxicity and acute renal failure. The most prominent are the aminoglycosides, for which renal toxicity has been reported, even with careful monitoring and maintenance of recommended drug levels. Cyclosporine, amphotericin B, and nonsteroidal anti-inflammatory drugs also may contribute to postoperative acute renal failure. The commonly used chlorinated inhalation anesthetic agents have been responsible for primary acute renal failure after operation as well.

Laboratory Studies

Urinalysis

Examination of the urine is an essential diagnostic test in patients with postoperative renal failure. The presence of blood or myoglobin is a positive diagnostic test, and red cell casts may be present in the urine of patients with urinary obstruction. Sodium, creatinine, urea, and osmolality levels should be measured in the urine.

Urine Osmolality

Patients with acute renal failure are isosthenuric; that is, the urine osmolality is close to that of plasma, typically near 300 mOsm/L. Patients with prerenal azotemia have osmolalities of 500 mOsm/L or more (Table 2-11). The ratio of urine-to-plasma osmolality has been shown to be more discriminating than urinary values alone. Urine-to-plasma osmolality ratios of less than 1:10 are consistent with acute renal failure, but prerenal azotemia usually produces ratios of 1:25 or higher.

Urine Urea and Creatinine

The urine-to-plasma urea and urine-to-plasma creatinine ratios are the most useful in diagnosing acute renal failure postoperatively. A urine-to-plasma creatinine ratio below 20 is indicative of acute renal failure, and a ratio above 40 indicates prerenal azotemia. A urine-to-plasma urea ratio of less than 3 indicates tubular injury, and a ratio above 8 usually indicates prerenal azotemia.

Urine Sodium

The underperfused kidney is sodium retaining, and a low urine sodium concentration is characteristic of prerenal azotemia. However, when acute renal failure has occurred, there is diminished sodium reabsorption by the kidney. Patients with acute renal failure usually have a higher urine sodium level (above 40 mEq/L). Values of urine sodium concentration between 20 and 40 mEq/L are nondiscriminatory. Consequently, urine sodium concentration is not as sensitive in providing diagnostic accuracy as is the urine-to-plasma urea ratio.

Management of the Patient with Established Acute Renal Failure

When the diagnosis of acute renal failure is made with rising levels of blood urea nitrogen and creatinine and low urine volume, initial efforts should be directed toward correcting reversible causes. Attention then is turned to fluid and electrolyte balance problems occurring because of acute renal failure and adjusting the dosage of any administered drug to compensate for impaired elimination.

Fluid and Electrolyte Management

Hyperkalemia

Of all the electrolyte abnormalities that are encountered in acute renal failure, hyperkalemia is the most serious, and it must be treated early. Untreated hyperkalemia leads to cardiac arrest. The severity of the hyperkalemia can be estimated by the ECG changes, including peaked T waves, prolonged PR intervals, loss of P waves, and widening of the QRS complex.

When significant ECG changes are apparent, calcium infusion with 1 to 2 g 10% calcium gluconate should be administered over 10 to 15 min to stabilize the cardiac membranes and neutralize the toxic effects of hyperkalemia. This therapy frequently is lifesaving, but the serum potassium level must be lowered quickly. This may include the administration of insulin, concentrated glucose, and intravenous sodium bicarbonate. Enteric cation-exchange resins or dialysis therapy must be used fairly soon to remove the potassium from the body.

Fluid Volume

In a patient with renal failure and little or no urine output, excessive intake of salt and water will be retained, eventually causing pulmonary edema and congestive heart failure. Fluid intake should be restricted to replacing measured fluid losses plus 500 to 600 mL per day of insensible loss. Careful balance studies of intake and output are mandatory. Once any extracellular fluid volume deficits have been corrected, the quantity of maintenance fluids amount only to measurable losses and insensible loss.

Hyponatremia

Hyponatremia evolves early, usually because of excessive free water availability from breakdown of protein, carbohydrate, and fat as well as administered free water. If serum sodium concentration falls below 120 mEq/L, dialysis is the only therapeutic endeavor that corrects hyponatremia.

Metabolic Acidosis

Metabolic acidosis is almost inevitable when the kidney fails to remove acid by-products from the body. Treatment involves the use of sodium bicarbonate, but dialysis is often required as well. Excessive restriction of protein in the catabolic injured patient in order to delay the need for dialysis is not advisable.

Other Electrolyte Abnormalities

Hypocalcemia and hyperphosphatemia may occur in the patient who has experienced a crush injury or a burn covering a significant proportion of body surface area. Severe hyperphosphatemia may require dialysis. Hypocalcemia can be managed with careful replacement of calcium.

Use of Dialysis in Acute Renal Failure

The indications for dialysis are listed in Table [2-12](#). Dialysis is best initiated before the occurrence of the life-threatening complications of acute renal failure such as hyperkalemia, severe acidosis, uremic encephalopathy, or uremic pericarditis. There are four forms of dialysis for acute renal failure: hemodialysis, peritoneal dialysis, continuous arterial-venous dialysis, and continuous venovenous ultrafiltration.

Hemodialysis is the most effective and is the treatment of choice in the very hypercatabolic patient. Hemodialysis may require 4 to 5 h of daily dialysis to counteract the effects of the hypercatabolism. In others, dialysis may be done three to four times per week, depending on the rate of hypercatabolism. A major advantage of hemodialysis is that removal of fluid by ultrafiltration is easily controlled.

Peritoneal dialysis frequently is used for the patient with severe heart disease, including coronary artery disease and myocardial infarction. For these patients peritoneal dialysis may be safer than hemodialysis, with more gradual changes in volume, particularly after surgery.

Continuous therapy, such as arterial-venous and venovenous dialysis, have virtually replaced the use of hemodialysis in most postoperative patients with acute renal failure. The advantage of venovenous continuous therapy is that an arterial line is not required. A blood pump is used to maintain extracorporeal blood flow. An additional advantage of this system is that the flow is not blood pressure dependent.

High-Output Renal Failure

Uremia occurring without a period of oliguria and accompanied by a daily urine volume greater than 1000 to 1500 mL/day is a more frequent but less well recognized disorder than acute renal insufficiency. Clinical experience and laboratory experiments suggest that high-output renal failure represents the renal response to a less severe or a modified episode of renal injury than that required to produce classic oliguric renal failure; it is thus a milder form of renal insufficiency. Through serial measurement of blood urea nitrogen and serum electrolytes, informed chemical and fluid volume management of high-output renal failure can be carried out with a much greater latitude because of the daily urine volume excretion. Normal extracellular fluid volume and normal serum sodium concentration are easily maintained when accurate daily outputs of each are obtained and replaced accordingly. The sodium-containing fluids may be administered with lactate to control the mild metabolic acidosis that occurs. Severe acidosis may develop if isotonic losses from the gastrointestinal tract or renal excretion of sodium are replaced with sodium chloride.

The primary danger of high-output renal failure is the delay in recognition because of normal urine output. The inappropriate administration of intravenous potassium in this setting can result in hyperkalemia. Good urinary output and gastrointestinal involvement requiring suction usually indicate the need for daily potassium replacement. With this type of renal failure, potassium intoxication may be produced. As little as 20 mEq potassium chloride given intravenously may rapidly produce myocardial potassium intoxication requiring exchange resin or hemodialysis treatment.

The typical course of high-output renal failure begins without a period of oliguria. The daily urine volumes are normal or higher than normal, often reaching levels of 3 to 5 L/day while blood urea nitrogen is increasing. An attempt to decrease urine output by water restriction rapidly results in hypernatremia without a change in urine volume. On the average, urea nitrogen continues to increase for 8 to 12 days before a downward trend occurs. The blood-to-urine urea ratio is about 1:10 until a decrease occurs in the blood urea concentration.

Functionally, the disorder is characterized by a glomerular filtration rate of less than 20 percent of normal and complete resistance to vasopressin for 1 to 3 weeks after the blood urea nitrogen has declined. During the next 6 to 8 weeks, the glomerular filtration rate gradually rises, and the response to vasopressin becomes normal.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 3 - Hemostasis, Surgical Bleeding, and Transfusion

Seymour I. Schwartz

BIOLOGY OF HEMOSTASIS

Hemostasis is a complex process that prevents or terminates blood loss from a disrupted intravascular space, provides a fibrin network for tissue repair, and, ultimately, removes the fibrin when it is no longer needed (Fig. 3-1). Endothelial cells functionally act to prevent clotting. They interfere with platelet recruitment by inactivating *adenosine diphosphate (ADP)*. They provide an environment in which thrombin is also inactivated by complexing with antithrombin III. Endothelial cells release thrombomodulin, which down- modulates the coagulation process. Four major physiologic events participate, both in sequence and interdependently, in the hemostatic process. Vascular constriction, platelet plug formation, fibrin formation, and fibrinolysis occur in that general order, but the products of each of these four processes are interrelated in such a way that there is a continuum and multiple reinforcements (Fig. 3-2).

Vascular Constriction

Vasoconstriction is the initial vascular response to injury, even at the capillary level. It is dependent upon local contraction of smooth muscle that has a reflex response to various stimuli. The initial vascular constriction occurs before any platelet adherence at the site of injury. Adherence of endothelial cells to adjacent endothelial cells may be sufficient to cause cessation of blood loss from the intravascular space. Vasoconstriction is subsequently linked to platelet plug and fibrin formation. *Thromboxane A₂ (TXA₂)*, which results from the release of arachidonic acid from platelet membranes during aggregation, is a powerful vasoconstrictor. By contrast, prostacyclin, which is also secreted during the platelet release reaction, is a potent vasodilator. Serotonin, 5-hydroxytryptamine (5-HT), released during platelet aggregation, is another vasoconstrictor, but it has been shown that when platelets have been depleted of serotonin in vivo, constriction is not inhibited. Bradykinin and fibrinopeptides in the coagulation schema are also capable of contracting smooth muscle. Some patients with mild bleeding disorders and a prolonged bleeding time have, as their only abnormality, capillary loops that fail to constrict in response to injury.

A lateral incision in a small artery may remain open because of physical forces, whereas complete transection of a similarly sized vessel contracts to the extent that bleeding may cease spontaneously. The vascular response factor should also include the contribution of pressure provided by surrounding tissues. Bleeding from a small venule ruptured by trauma, in the thigh of an athlete, may be negligible because of the compressive effect of surrounding muscle. In the same individual, bleeding from a similar vessel in the nasal mucosa may be significant. When there is low perivascular pressure, as seen in patients with muscle atrophy accompanying aging, in patients on prolonged steroid therapy, and in patients with the Ehlers-Danlos syndrome, bleeding tends to be more persistent. Vascular abnormalities, such as hereditary hemorrhagic telangiectasia, may predispose the patient to bleeding from the involved region.

Platelet Function

Platelets are 2-mm diameter fragments of megakaryocytes and number 200,000 to 400,000/mm³ in circulating blood with a life span of 7 to 9 days. They play an integral role in hemostasis along two pathways. Platelets, which normally do not adhere to each other or to the normal vessel wall, form a plug that stops bleeding when vascular disruption occurs. Injury to the intima exposes subendothelial collagen to which platelets adhere within 15 s of the traumatic event. This requires von Willebrand factor (vWF), a protein that is lacking in patients with von Willebrand's disease. The platelets then expand and develop pseudopodal processes and also initiate a release reaction that recruits other platelets from the circulating blood. As a consequence, a loose platelet aggregate forms, sealing the disrupted blood vessel. The aggregation up to this point is reversible and is not associated with secretion. This process is known as *primary hemostasis*. The administration of heparin does not interfere with this reaction, which is why hemostasis can occur in the heparinized patient. *ADP* and serotonin are principal mediators in this process of adhesion and aggregation. Various prostaglandins have opposing activities. Arachidonic acid, released from platelet membranes, is converted by cyclooxygenase to prostaglandin G₂ (PGG₂) and PGH₂, which in turn are converted to *TXA₂*, a potent platelet aggregator and vasoconstrictor. By contrast, PGI₂ (prostacyclin) and PGE₂ inhibit aggregation and act as vasodilators.

ADP, released from damaged tissues and platelets, plus platelet factor 4 and trace thrombin on the platelet surface in the face of Ca²⁺ and Mg²⁺, stimulate a platelet release reaction by which the content of the platelet and its granules is discharged.

Fibrinogen is required for this process. Thrombin plays a central role in this process by stimulating platelet degranulation and activating the generation of thromboxane A₂. During this process, platelet factor 4, b- thromboglobulin, platelet-derived growth factor, *ADP*, serotonin, and calcium are introduced into the plasma. The release reaction results in compaction of the platelets and the formation of an “amorphous” plug, which is no longer reversible. This process is inhibited by cyclic adenosine monophosphate (cAMP). As a consequence of the release reaction, platelet factor 3 is made available and contributes phospholipid to several stages of the coagulation cascade.

The lipoprotein surface provided by platelets catalyzes reactions that are involved in the conversion of prothrombin (factor II) into thrombin (Fig. 3-3). Platelet factor 3 is involved in the reaction by which activated factor IX (IXa), factor VIII, and calcium activate factor X. It is also involved in the reaction by which factor Xa, factor V, and Ca²⁺ activate factor II. Platelets may also play a role in the initial activation of factors XI and XII. Platelet factor 4 and b-thromboglobulin are also made available during the release reaction, and they may inhibit the activity of heparin and modify fibrin formation. The platelets also play a role in the fibrinolytic process by releasing an inhibitor of plasminogen activation.

Coagulation

Coagulation is the process by which prothrombin is converted into the proteolytic enzyme thrombin, which in turn cleaves the fibrinogen molecule to form insoluble fibrin in order to stabilize and add to the platelet plug. Coagulation consists of a series of zymogen activation stages in which circulating proenzymes are converted in sequence to activated proteases (Fig. 3-4). The traditional concept of the clotting system evolved from test tube analysis and follows two pathways: the *intrinsic* pathway involves components normally present in blood, and the *extrinsic* pathway is initiated by the tissue lipoprotein (Fig. 3-5). In the intrinsic pathway factor XII is activated by binding to subendothelial collagen. Prekallikrein and high- molecular-weight kininogen amplify this contact phase. Activated factor XII (XIIa) proteolytically cleaves factor XI and also prekallikrein to form factor XIa and kallikrein. In the presence of Ca²⁺, factor XIa activates factor IX (IXa). This in turn complexes with factor VIII, which can be activated to a more potent form by thrombin, and, in the presence of Ca²⁺ and the phospholipid platelet factor 3, activates factor X. In the extrinsic pathway, the tissue phospholipid, thromboplastin, reacts with factor VII and Ca²⁺ to activate factor X.

Activated factor X (Xa), produced by the two pathways, proteolyses prothrombin (factor II) to form thrombin. The effects of thrombin are limited to the area of endothelial disruption by several processes. Thrombin activates the fibrin stabilizing factor (XIII) and cleaves fibrinopeptides A and B from fibrinogen (factor I) to form fibrin, a monomer that is cross-linked with factor XIIIa, to form a stable clot (Fig. 3-6). The escape of thrombin into the circulation is prevented from complexing with antithrombin locally by the binding of the thrombin to thrombomodulin on the endothelium, creating a complex that cannot cleave fibrinogen and also activating protein C, which in turn inactivates factors V and VIII. Additionally, circulating thrombin is inactivated by plasma protease inhibitors. This process is accelerated by factor V, tissue lipoproteins, platelet surface phospholipids, and Ca²⁺.

All the coagulation factors except thromboplastin, Ca²⁺, and most of factor VIII are synthesized in the liver. Factors II, VII, IX, and X require vitamin K for their production (Table 3-1).

Fibrinolysis

Fibrinolysis is a natural process directed at maintaining the patency of blood vessels by lysis of fibrin deposits. Also involved in the maintenance of vascular patency is circulating antithrombin III (ATIII) which neutralizes the action of thrombin and other proteases in the coagulation cascade.

Fibrinolysis is initiated at the same time as the clotting mechanism under the influence of circulating kinases, tissue activators, and kallikrein that are present in many organs, including venous endothelium. Fibrinolysis is dependent on the enzyme plasmin, which is derived from a precursor plasma protein (plasminogen) (Fig. 3-7). Plasminogen levels are known to rise as a consequence of exercise, venous occlusion, and anoxia. Plasminogen activation is also initiated by the activation of factor XII. The plasminogen is preferentially absorbed on fibrin deposits. The enzyme plasmin lyses fibrin and acts on other coagulant proteins as well, including fibrinogen, factor V, and factor VIII. The smaller fragments of polypeptide products of fibrin that are produced interfere with normal platelet aggregation; the larger fragments are incorporated into the clot in lieu of normal fibrin monomers and result in an unstable clot. Human blood also contains an antiplasmin that inhibits plasminogen activation, and platelets are believed to possess antifibrinolytic activity.

CONGENITAL HEMOSTATIC DEFECTS

Inheritance

The modes of inheritance of hemostatic disorders, with only rare exceptions, are of three types: (1) autosomal dominant, (2) autosomal recessive, and (3) sex-linked recessive. The most common hemostatic disorder transmitted by the autosomal dominant mode is von Willebrand's disease. Hereditary hemorrhagic telangiectasia and factor XI deficiency also appear to be transmitted in this fashion.

Occasionally, in a pedigree with an autosomal dominant gene, an *apparently* normal person may transmit disease to his or her child. The parent clearly carried the gene, which clinically expressed no defect.

In inherited hemostatic disorders, the difference in clinical expression between dominant and recessive genes is a graded one rather than an “all- or-none” phenomenon. The heterozygous individual with an autosomal recessive trait may have a measurable deficiency of the factor governed by that gene, but no clinical disease. In order to demonstrate clinical expression of disease, the individual must be homozygous. This appears to be the case, for example, in factor X deficiency. Other hemostatic disorders probably inherited in this mode are deficiencies in factor V, factor VII, and factor I.

Sex-linked recessive inheritance governs true hemophilia (factor VIII deficiency) and factor IX deficiency (Christmas disease). The genes for these diseases are recessive in expression and are carried on the female (X) chromosome. When paired with the normal X chromosome (the female carrier state), clinical disease is not present. When the affected X chromosome is paired with the normal male (Y) chromosome, clinical disease is expressed.

Platelet Deficiencies

Hereditary quantitative disorders include disorders of platelet production (hereditary thrombocytopenia) and platelet destruction (Wiskott-Aldrich syndrome). The most common congenital platelet deficiency is the abnormality seen in *von Willebrand's disease*, in which the von Willebrand factor (vWF) is missing; vWF has been shown to be required for platelet adhesion to subendothelial collagen (see section below). Also, platelets from patients with von Willebrand's disease, unlike those of normal patients, fail to aggregate in vitro with the addition of ristocetin. Another inherited disorder affecting platelets is the rare *Bernard-Soulier syndrome*. Patients with Bernard-Soulier syndrome have normal levels of vWF, and the addition of vWF does not affect aggregation of platelets in the presence of ristocetin. In Bernard-Soulier syndrome, the platelet membrane receptor for vWF, a portion

of the glycoprotein I complex, is missing.

Glanzmann's thrombasthenia is a rare congenital disorder in which platelets fail to aggregate in the presence of *ADP* and mediation of factors involved in clot retention is impaired. Patients with *congenital afibrinogenemia* also have impairment of platelet aggregation because fibrinogen is required for this process to occur. Patients with congenital afibrinogenemia have disturbed platelet function, manifested by a prolonged bleeding time correctable by fibrinogen administration.

Congenital disorders of platelet secretion include *storage pool disease*, in which the platelets lack the storage capability of *ADP* required for aggregation. The Hermansky-Pudlak syndrome (oculocutaneous albinism, ceroidlike deposits in macrophages, and bleeding diathesis) is classified in this category. Congenital *primary release defects* have also been described and are responsible for prolonged bleeding time.

Congenital Defects of Coagulation Factors

Factor VIII Deficiency (Classical Hemophilia)

Classical hemophilia (hemophilia A) is a disease of males. The failure to synthesize factor VIII in normal proportions is inherited as a sex-linked recessive trait. Spontaneous mutations account for almost 20 percent of cases. The incidence of the disease is approximately 1:10,000 to 1:15,000 population, and the clinical manifestations can be extremely variable.

Clinical Manifestations

Characteristically, the severity of clinical manifestations is related to the degree of deficiency of factor VIII. Spontaneous bleeding and severe complications are the rule when virtually no factor VIII can be detected in the plasma. When plasma factor VIII concentrations are in the range of 5 percent of normal, the patient may have no spontaneous bleeding yet may bleed severely with trauma or surgical treatment. Patients with levels greater than 5 percent of normal (greater than 0.05 units/mL) are considered mild hemophiliacs. Patients whose factor VIII levels fall between 1 and 5 percent of normal are considered moderately severe hemophiliacs. Typically, members of the same pedigree with true hemophilia have approximately the same degree of clinical manifestations.

While the severely affected patient may bleed during early infancy, significant bleeding typically is noted first when the child is a toddler. At that time, in addition to the classic bleeding into joints, epistaxis and hematuria may be noted. Bleeding that is life-threatening may follow injury to the tongue or lingual frenulum. Tracheal compression and retropharyngeal bleeding may follow tonsillar infection. Intracranial bleeding, associated with trauma in half the cases, accounts for 25 percent of deaths. Vascular and neural compromise may occur in relation to pressure secondary to bleeding into a soft-tissue closed space. Talipes equinus contracture deformity may be seen in severely hemophilic patients secondary to bleeding into the calf. Volkmann's contracture of the forearm and flexion contractures of the knees and elbows are also disabling sequelae of deep soft-tissue bleeding.

Hemarthrosis is the most characteristic orthopaedic problem. Bleeding into the joint may cause few symptoms until distention of the joint capsule occurs. A large hemarthrosis generally is manifested by a tender, swollen, warm, and painful joint. Muscle spasm and pain around the joint arise from involvement of periarticular structures. These signs may mimic infection. The same orthopaedic problems are noted in association with severe factor IX deficiency (Christmas disease).

Retroperitoneal bleeding may follow lifting of a heavy object or strenuous exercise. Signs of posterior peritoneal irritation and spasm of the iliopsoas suggest the diagnosis. Hypovolemic shock may occur, since the amount of blood loss that can take place in this setting is enormous. The clinical manifestations of intramural intestinal hematoma are nausea and vomiting, crampy abdominal pain, and signs of peritoneal irritation that mimic those of appendicitis. Fever and leukocytosis may be noted. Radiographs of the abdomen may fail to reveal any abnormality or may display a modest amount of ileus. Upper gastrointestinal examination may demonstrate a uniform thickening of mucosal folds that has been described as a "picket fence" or "stack of coins" appearance (Fig. 3-8). Intramural hematomas of the intestine occur in other hemostatic disorders and therefore should be considered when any patient with a hemostatic problem presents with findings suggesting an acute intraabdominal process.

Treatment

Replacement Therapy

The plasma concentration of factor VIII necessary for maintenance of hemostatic integrity is normally quite small. Patients with as little as 2 to 3 percent of factor VIII activity usually do not bleed spontaneously. Once serious bleeding begins, however, a much higher level of factor VIII activity, probably approaching 30 percent, is necessary to achieve hemostasis.

The half-life of factor VIII is 8 to 12 h. After administration of a given dose of factor VIII, approximately one-half of the initial posttransfusion activity disappears from the plasma in 4 h. This early disappearance is thought to be due in large part to diffusion from the intravascular space. The period of equilibration may extend for as long as 8 h, at which time only about one-quarter of the initial level remains in the circulating blood. From that time on, the slope of disappearance is less steep. Twenty-four hours after a given dose, no more than 7 to 8 percent of administered factor VIII activity remains within the circulation.

One unit of factor VIII activity is considered that amount present in 1 mL of normal plasma. Actually, fresh frozen plasma contains 0.60 unit/mL. Theoretically, in a patient with 0 percent activity, to achieve an initial posttransfusion level of 60 percent of normal, using fresh plasma, a volume of plasma equal to 60 percent of the patient's estimated plasma volume would have to be administered. Table 3-2 shows approximate levels of factor VIII required for hemostasis in different disorders. The minimum hemostatic level of factor VIII for mild hemorrhages is 30 percent; for joint and muscle bleeding and major hemorrhages, it is 50 percent. For major surgery and life-threatening bleeding, levels of 80 to 100 percent should be reached preoperatively and maintained above 30 percent for 2 weeks. Remembering the loss from the circulation, one-half the initial dose would need to be supplied every 12 h. The use of fresh plasma in such circumstances would require a volume that is excessive. Factor VIII concentrates are available that circumvent this problem. Cryoprecipitate concentrates of factor VIII can be regarded as containing 9.6 units/mL. The amount of material to be given can be computed from the formula

Patient's weight (kg) × desired rise of factor VIII (% average normal) / Total units of factor VIII in dose = *R*

where *R* is a factor that is fairly constant for any given type of material and represents the rise of factor VIII obtained in the patient's plasma for every unit of transfused factor VIII per kilogram of the patient's body weight. Half that amount is subsequently administered every 4 to 6 h to maintain a safe level.

A variety of factor VIII concentrates are available. Regardless of the preparation employed, continued laboratory assessment of circulating factor VIII level is an important element in the control of these patients. Wet-frozen cryoprecipitate is preferred for replacement in patients with mild hemophilia since the risk of hepatitis is less than it is with factor VIII concentrates. The latter are preferred for major replacement problems. In mild hemophilia A and in mild von Willebrand's disease, DDAVP (1-desamino- 8-d-arginine vasopressin), a synthetic derivative of vasopressin, has been used to effect a dose-dependent increase of all factor VIII activities and to effect release of plasminogen activator. DDAVP reduces blood loss associated with major surgical procedures by 40 percent. Patients undergoing orthopaedic or neurosurgical procedures should also receive a fibrinolytic inhibitor.

After major surgical treatment of the hemophiliac, transfusion replacement of factor VIII should be continued for at least 10 days. Wounds should be well healed and all drains removed before termination of therapy. If sutures remain, transfusion should be reinstituted before they are removed. Many recent reports document the safety of major surgical procedures in hemophilic patients receiving replacement therapy. But in one large series the incidence of postoperative hemorrhage did not improve over a 16-year period despite a threefold increase in dosage of factor VIII, suggesting that circulating factor VIII levels are not the sole determinant of bleeding in these patients.

The viruses of homologous serum hepatitis and HIV have been transmitted by various plasma concentrates. Other complications of replacement therapy include the appearance of inhibitors of factor VIII, which may arise in the hemophiliacs who have had transfusion. These inhibitors have been characterized as antibodies of the gG variety. They tend to diminish in several weeks if further transfusion is not employed. Laboratory search for these factors should be carried out in every hemophilic patient who is considered a candidate for elective surgical treatment, as their presence complicates transfusion management. Paradoxical bleeding may occur in patients transfused to an appropriate factor VIII level as a result of the development of abnormal platelet function.

Adjunctive Management

Treatment of soft tissue bleeding is directed at the prevention of airway obstruction and vascular and neural damage. These are accomplished best by the administration of sufficient factor VIII. Bed rest and cold packs can be of some assistance. In general, results of fasciotomy to relieve pressure have varied from disappointing to disastrous. The occasional development of large cysts has resulted in sufficient deformity and disability to require amputation.

The primary treatment of hemophilic hemarthrosis is directed at maintaining full range of motion and minimal destruction of the cartilage. Aspiration of blood from the hemophilic joint is not uniformly endorsed, and when regarded as necessary it should be considered a major surgical event. Elevation of factor VIII level by transfusion is necessary. The procedure should be carried out in the operating room under strict sterile precautions. In most instances aspiration is not required, and the combination of factor VIII replacement and local cold packing proves sufficient. Physiotherapy plays a critical role and should consist of *active* exercises, since the patient is unlikely to move the extremity to a point where bleeding will recur. Passive exercises often result in recurrence of bleeding. The reader is referred to the review by Curtiss for details of orthopaedic management.

The management of intramural intestinal hematoma and retroperitoneal bleeding is based on appropriate transfusion therapy and avoidance of surgical treatment. Even when a relatively minor procedure, such as tracheostomy, is performed, the plasma level of factor VIII should be raised above 25 to 30 percent. Since dental hygiene usually is poor in hemophilic patients, dental and oral surgical treatment frequently are necessary.

Factor IX Deficiency (Christmas Disease)

Factor IX deficiency clinically is indistinguishable from factor VIII deficiency and also has an X-linked recessive mode of inheritance. These two entities were considered a single disease until 1952, when their unique deficiencies were documented. Factor IX deficiency accounts for 20 percent of hemophiliacs. Like classical hemophilia (hemophilia A), Christmas disease (hemophilia B), can occur in severe, moderate, or mild forms according to the level of factor IX activity in the plasma. One-half of afflicted patients have the severe form, which has a factor IX level of less than 1 percent. Patients have a prolonged *partial thromboplastin time (PTT)*.

Treatment

Most patients with severe factor IX deficiency require substitution therapy on a regular basis. All patients require substitution therapy whenever minor or major surgery is performed. Therapy is generally based on the administration of fresh frozen plasma or, rarely, factor IX concentrates. Initially the rate of disappearance of factor IX from the circulation is more rapid than that of factor VIII; subsequently factor IX, with a half-life of 18 to 40 h, has a slower disappearance rate. During severe hemorrhage, treatment should be directed at achieving plasma factor IX levels of 20 to 50 percent of normal for the first 3 to 5 days, and then maintaining the plasma level at 10 to 20 percent of normal for approximately 10

days. The usual daily dose is 30 to 50 units/kg of body weight, followed by 20 units/kg of body weight/24 h. When an operation is required, the plasma level of approximately 50 to 70 percent of normal should be achieved. During the operation and the first postoperative day, 60 units/kg of body weight/24 h is recommended, followed by 30 to 40 units/kg every 24 h for the next 2 to 3 days, and 20 units/kg every 24 h for 3 to 4 days. In all instances, the levels should be monitored by laboratory determinations. The development of antibodies against factor IX represents a serious complication that is difficult to deal with. This occurs in about 10 percent of patients with Christmas disease. These patients are managed by withholding all infusion therapy with blood or plasma. High doses of factor IX concentrates combined with cyclophosphamide have been effective.

von Willebrand's Disease

von Willebrand's disease occurs as commonly as true hemophilia. The increasing recognition of this disease is related to more reliable factor VIII assays. This hereditary disorder of hemostasis is usually transmitted as an autosomal dominant trait, but recessive inheritance may occur. The disease is characterized by a diminution of the level of factor VIII:C (procoagulant) activity that corrects the clotting abnormality in hemophilia A plasma. The reduction of factor VIII:C activity usually is not as great as that seen in classical hemophilia. Also unlike classical hemophilia, in which factor VIII:C activity remains constant, in the patient with von Willebrand's disease, variation in the level of circulating factor VIII:C activity may be noted. Characteristically, these patients also have a prolonged bleeding time, but this is less constant than the factor VIII:C reduction. A given patient may have an abnormal bleeding time on one occasion and a normal bleeding time on another. The level of factor VIII-related antigen (factor VIII:Ag) is disproportionately lower than that of factor VIII:C, and ristocetin fails to cause platelet aggregation in about 70 percent of patients with the disease. The vast majority of patients have a prolonged activated partial thromboplastin time (aPTT).

Clinical Manifestations

The manifestations of bleeding usually are mild and often are overlooked until trauma or the stress of surgical treatment makes them apparent. A careful clinical history is therefore of great importance in these patients. Spontaneous manifestations often are limited to bleeding into the skin or mild mucous membrane bleeding. Epistaxis and menorrhagia have been relatively common in the author's experience. Serious bleeding after dental extractions and tonsillectomy also are not uncommon. Fatal bleeding from the gastrointestinal tract has been described.

Treatment

Treatment is directed at correcting the bleeding time and factor VIII R:vWF (the von Willebrand factor). Only cryoprecipitate is reliably effective. Lyophilized concentrates of factor VIII:C lack the required factor VIII R:vWF. Newer factor VIII concentrates have a full complement of vWF. Bleeding time is corrected with cryoprecipitate 10 to 40 units/kg of body weight/12 h. Replacement therapy should be begun 1 day before a surgical procedure. Aspirin *must* be avoided for 10 days before an elective operation. Duration of treatment should be the same as that described for the patient with classical hemophilia. Intravenous high-dose g-globulin has been used successfully to treat an acquired type of von Willebrand's disease associated with myeloproliferative disorders, leukemias, and lymphomas.

An analogue of the antidiuretic hormone vasopressin, DDAVP (desmopressin acetate), administered intravenously temporarily (for 1 to 2 h) increases vWF:Ag, ristocetin cofactor, and VIII:C and shortens the bleeding time. Desmopressin may induce thrombocytopenia and release tissue plasminogen activator.

Rare Deficiencies of Coagulation Factors

Factor XI (plasma thromboplastin antecedent [PTA]) deficiency (Rosenthal's syndrome) is a mild disorder occurring mainly in patients of Jewish ancestry. Patients may undergo major procedures without significant bleeding. Fresh frozen plasma is therapeutic. *Factor V (proaccelerin) deficiency (parahemophilia)* is rare; significant bleeding usually occurs in homozygotes. Excessive bleeding is characteristically associated with factor levels lower than 1 percent of normal. Administration of fresh frozen plasma to raise the level to 25 percent of normal is sufficient. *Factor VII (proconvertin) deficiency* also is associated with excessive operative bleeding in homozygous patients. Administration of banked plasma to raise the factor level above 10 percent of normal provides adequate hemostasis. *Factor X (Stuart-Prower) deficiency* is associated with amyloidosis and with familial carotid body tumors. Frozen or normal plasma is therapeutic. *Factor II (prothrombin) deficiency* is rare and can be corrected with plasma infusion to achieve levels greater than 30 percent of normal.

Although the *prothrombin time (PT)* will reveal factor VII deficiency, it cannot be used as a guide to therapy because it remains abnormal when adequate levels of factor VII are achieved. Patients with factor VII deficiency have normal *PTT* and *thrombin time (TT)*. Specific assays can define any of these deficiencies. Vitamin K therapy is ineffective in raising levels of factors II, V, VII, X, and XI.

When the transfusion programs outlined earlier for deficiency of the prothrombin group of factors (factors II, V, VII, and X) are employed, the one-stage prothrombin time does not return to normal. Rather, a one-stage *PT* slightly less than twice the control value is achieved. This is sufficient to result in normal hemostasis. Of the four "prothrombin" factors, only factor V must be provided as fresh or freshly frozen plasma. Stored plasma is equally effective as therapy for factors II, VII, and X.

Inherited Fibrinogen Abnormalities

Included in this category are patients with congenital afibrinogenemia, hypofibrinogenemia, and dysfibrinogenemia. Fewer than 200 cases of afibrinogenemia have been reported. This disorder is ascribed to an autosomal recessive mode of inheritance. The affected individuals are presumably homozygous for the trait. Bleeding time may be markedly prolonged in some patients because fibrinogen is required for platelet aggregation. Conventional methods for measuring fibrinogen in the plasma give a zero value, but immunologic techniques may detect trace amounts of a fibrinogen-like protein. Patients have an indefinitely prolonged whole-blood coagulation time, which can be corrected by the addition of fibrinogen. The deficiency usually is less a clinical problem than is classical hemophilia, however. Bleeding usually begins early in life, and bleeding from the umbilical cord is a characteristic symptom. Bleeding may follow operations, dental extraction, and trauma, but the most feared complication is intracranial bleeding after minor injury to the head.

Less profound inherited deficiencies of fibrinogen have been observed and categorized as congenital hypofibrinogenemia. Two groups of hypofibrinogenemic patients have been differentiated: those with fibrinogen values below 50 mg/dL and those with higher levels. The clinical manifestations depend on the fibrinogen concentration. Another congenital disorder is dysfibrinogenemia, in which there are structural defects in the fibrinogen molecule. Both hypofibrinogenemia and dysfibrinogenemia have a dominant mode of inheritance. Dysfibrinogenemic patients are frequently asymptomatic but may have moderate or severe bleeding associated with an operation. They have a propensity for thromboembolic disorders and have a higher incidence of wound dehiscence after operative intervention. The thrombin clotting time is diagnostic for this general category of abnormalities, but definition of the precise abnormality requires a series of complex laboratory studies.

Treatment

Although the hemostatically optimal level of fibrinogen is not known, a level greater than 100 mg/dL is generally required during an operation. The patient's fibrinogen level should be raised above this before the procedure. Substitution therapy may be effected by the infusion of fresh frozen plasma or cryoprecipitate. In order to achieve a fibrinogen level near 100 mg/dL for 24 h, an initial dose of 20 to 25 mg fibrinogen/kg of body weight should be administered, followed by one-third the initial amount given on a daily basis throughout the postoperative period. But appropriate corrections must be based on actual fibrinogen measurements. Normal fibrinogen concentration should be maintained until wound healing is shown to be adequate.

Congenital Factor XIII Deficiency

This rare autosomal recessive disorder is manifest by umbilical bleeding in the newborn and slow wound healing after an operation. In general, most of the bleeding manifestations are mild, but intracranial bleeding may occur as a consequence of minor trauma. Immunologic assays have demonstrated deficiency of the protein. Therapy is accomplished with fresh frozen plasma, cryoprecipitate, or factor XIII concentrates. With major bleeding, or accompanying surgical intervention, the desired concentration of the recipient's plasma is 0.3 to 0.5 mmol/mL. With minor bleeding or as prophylaxis, a level greater than 0.05 mmol/mL is all that is required.

ACQUIRED HEMOSTATIC DEFECTS

Platelet Abnormalities

Thrombocytopenia is the most common abnormality of hemostasis that results in bleeding in the surgical patient. The patient may have a reduced platelet count as a result of a variety of disease processes, such as idiopathic thrombocytopenic purpura, thrombotic thrombocytopenic purpura, and systemic lupus erythematosus, or secondary hypersplenism and splenomegaly of sarcoid, Gaucher's disease, lymphoma, and portal hypertension. In these circumstances the marrow usually demonstrates a normal or increased number of megakaryocytes. By contrast, when thrombocytopenia occurs in patients with leukemia or uremia, and in patients on cytotoxic therapy, there is generally a reduced number of megakaryocytes in the marrow.

Thrombocytopenia may occur acutely as a result of massive blood loss followed by replacement with stored blood. Exchange of one blood volume (11 units in a 75-kg man) decreases the platelet count from approximately 250,000/mm³ to approximately 80,000/mm³. Thrombocytopenia may be induced acutely by the administration of heparin and may be associated with thrombotic and hemorrhagic complications. This situation, which is thought to have an immunologic basis, has been reported in 0.6 percent of patients receiving heparin. The lowest platelet counts occur after 4 to 15 days of treatment in patients given heparin for the first time and after 2 to 9 days in those given subsequent courses.

Thrombocytopenia is often accompanied by impaired platelet function. Impaired aggregation after the addition of *ADP* has been demonstrated in patients receiving a blood transfusion of more than 10 units. Uremia may be associated with increased bleeding time and impaired aggregation, which can be corrected by hemodialysis or peritoneal dialysis. Defective aggregation and platelet secretion can occur in patients with thrombocythemia, polycythemia vera, or myelofibrosis. A variety of drugs interfere with platelet function, including aspirin, indomethacin, ibuprofen, dipyridamole, phenothiazines, penicillins, chelating agents, lidocaine, dextran, beta-adrenergic blockers, nitroglycerin, furosemide, and antihistamines.

The presence and extent of thrombocytopenia can be defined rapidly by a platelet count. In general, 60,000 platelets/mm³ is adequate for normal hemostasis, but if there is associated platelet dysfunction, there may be a poor correlation between the platelet count and the extent of bleeding. The template bleeding time is the most reliable in vivo test of platelet function.

When thrombocytopenia is present in a patient for whom an elective operation is being considered, it is managed on the basis of how much the platelet count is reduced and the cause of the reduction. A count greater than 50,000/mm³ requires no specific therapy. If thrombocytopenia is caused by acute alcoholism, drug effect, or viral infection, the platelet level will return to near normal within 1 to 3 weeks. Occasionally, severe thrombocytopenia may be secondary to vitamin B₁₂ or folic acid deficiency, in which case it is associated with a megaloblastic bone marrow. This condition generally occurs 2 to 3 years after total gastrectomy or in association with severe intestinal malabsorption. In either case, supplying the appropriate nutrient will correct the thrombocytopenia in 2 to 3 days.

If the patient has idiopathic thrombocytopenia or lupus erythematosus, and a platelet count less than 50,000/mm³, an attempt to raise the platelet count with steroid therapy or plasmapheresis may prove successful (see Chap. 3-1). The administration of platelet transfusions in these patients with the spleen in place is generally ineffective. The administration of g-globulin may temporarily increase the platelet count. Splenectomy alone should not be performed to correct the thrombocytopenia associated with splenomegaly secondary to portal hypertension.

Prophylactic platelet administration as a routine accompaniment to massive blood transfusion is not required or indicated to prevent a hemostatic defect. Platelet packs are administered preoperatively to rapidly increase the platelet count in surgical patients with thrombocytopenia due to marrow depression or in association with massive bleeding and replacement with banked blood. Special platelet transfusion sets are used to reduce the loss of platelets due to adherence. One unit of platelet concentrate contains approximately 5.5×10^{10} platelets and would be expected to increase the circulating platelet count by about 10×10^9 /L in the average 70-kg person. Hence a transfusion of 4 to 8 pool platelet concentrates should raise the count by 40 to 80×10^9 /L and should provide adequate hemostasis, as documented by bleeding time and control of the hemorrhagic manifestations. Fever, infection, hepatosplenomegaly, and the presence of antiplatelet alloantibodies decrease the effectiveness of platelet transfusions. In patients refractory to standard platelet transfusions, the use of *human lymphocyte antigen (HLA)*-compatible platelets coupled with special processors has proved effective. Platelet aggregometry has been applied to screening for potential donors.

Acquired Hypofibrinogenemia Defibrination Syndrome

The largest proportion of patients with fibrinogen-related problems of surgical concern are in this group. The fibrinogen deficiency rarely is an isolated defect, as thrombocytopenia and factors II, V, VII, VIII, and X deficiencies of variable severity usually accompany this state.

The majority of patients with acquired hypofibrinogenemia suffer from intravascular coagulation, more properly known as *defibrination syndrome* or *consumptive coagulopathy*, and it is to this group of patients that the term *disseminated intravascular coagulation (DIC)* has been applied. Systemic bleeding, however, dominates the clinical manifestations; thrombi are rarely found at autopsy. The syndrome, now recognized with increasing frequency, is caused by the introduction of thromboplastic materials into the circulation. Because this material is found in most tissues, many disease processes may activate the coagulation system. Evidence of the thrombotic process includes patchy necrosis of the skin, hematuria and oliguria, confusion due to cerebral ischemia, gastrointestinal bleeding, and hemorrhage into the adrenal cortex causing acute onset of hypotension. The hemorrhagic disasters of the perinatal period, e.g., retained dead fetus, premature separation of the placenta, and amniotic fluid embolus, are due primarily to this pathophysiologic mechanism. The hemorrhagic state that follows hemolytic transfusion reaction is also related to this process. Defibrination has been observed as a complication of extracorporeal circulation, head trauma, mucin-producing and disseminated carcinoma, lymphomas, thrombotic thrombocytopenia, rickettsial infection, snakebite, burns, aortic surgery, and shock from any cause. Release of thromboplastic material has long been a recognized complication of gram-negative sepsis and has been attributed to the effects of circulating endotoxin on platelets. Septicemia caused by gram-positive organisms may also be associated with DIC.

The differentiation of *DIC* with secondary protective fibrinolysis from primary fibrinolytic states can be extremely difficult because the *TT* is prolonged in both cases, as is the *PT* and *PTT*. There is no laboratory test to confirm or exclude the diagnosis. The combination of a low platelet count, a positive plasma protamine test indicating the presence of fibrin monomer-fibrinogen complexes in the plasma, and reduced fibrinogen accompanied by increased fibrin degradation products viewed in the context of the patient's underlying disease is highly suggestive of the diagnosis. The fibrinogen level is generally below 100 mg/dL when there is significant diffuse bleeding.

Treatment

The most important facets of treatment are relieving the patient's causative primary medical or surgical problem and maintaining adequate capillary flow. The use of intravenous fluids to maintain volume, and sometimes vasodilators to open the arterioles, is indicated. If blood flow deficiency is related to the inability of a damaged heart to pump, the use of drugs such as digitalis or isoproterenol may be indicated. Viscosity may be affected by an increased hematocrit concentration, and therefore a plasma expander may be beneficial.

If there is active bleeding, hemostatic factors should be replaced with fresh frozen plasma, which is usually sufficient to correct the hypofibrinogenemia; cryoprecipitate, which also provides fibrinogen (250 mg/10 mL); and platelet concentrates. There is little evidence that this replacement therapy will "fuel the fire" and accelerate the pathophysiologic process. Most studies show that heparin is not helpful in acute forms of *DIC*, but the drug is indicated for purpura fulminans or venous thromboembolism. Fibrinolytic inhibitors such as *ε-aminocaproic acid (EACA)* may be used to block the accumulation of degradation products but are dangerous if the thrombotic process is still active. They should not be used without prior effective antithrombotic treatment with heparin.

Fibrinolysis

The acquired hypofibrinogenemic state in the surgical patient also can be due to pathologic fibrinolysis. This may occur in patients with metastatic prostatic carcinoma, shock, sepsis, hypoxia, neoplasia, cirrhosis, and portal hypertension and in those patients on extracorporeal bypass.

The pathogenesis of this bleeding disorder is complex. Secondary to shock or hypoxia, a release of excessive plasminogen activator into the circulation occurs. This is thought to consist of endogenous kinases that can be released from vascular endothelium and other tissues. Pharmacologic activation of plasminogen also occurs with pyrogens, epinephrine, nicotinic acid, and acetylcholine. Electric shock and pneumoencephalography have also been reported to cause activation. Patients with cirrhosis and portal hypertension have a diminished ability to clear normal amounts of plasminogen activator from the blood. Sufficient urokinase to cause fibrinolysis can be released during operations on the prostate. The administration of exogenous fibrinolytics can also result in diffuse bleeding.

In addition to the reduction in levels of plasma fibrinogen, diminution of factors V and VIII also occurs, since they also serve as substrates for the enzyme plasmin. Thrombocytopenia is not an accompaniment of the purely fibrinolytic state. Polymerization of fibrin monomers, a step in normal fibrin formation, is interfered with by the proteolytic residue of fibrinogen and fibrin. The fibrin and fibrinogen breakdown products usually disappear from the circulation in a matter of hours. The whole blood clot lysis time defines increased fibrinolytic activity if a non-anticoagulated blood sample lyses in a test tube in less than 8 h. A euglobulin lysis time of 20 min or less provides a more rapid assessment.

Treatment

The successful treatment of the underlying disorder usually is followed by rapid spontaneous recovery, since the severity of fibrinolytic bleeding is dependent upon the concentration of breakdown products in the circulation. *EACA*, which is a synthetic amino acid, interferes with fibrinolysis by inhibiting plasminogen activation. The drug may be administered intravenously or orally. An initial dose of 5 g for the average-sized adult is followed by another 1 g every 1 to 2 h until the hemorrhagic state subsides. Treatment rarely is required for more than 2 or 3 days. Just as the administration of *EACA* in a patient with consumptive coagulopathy is potentially dangerous, the administration of heparin in the patient who has a primary pathologic fibrinolysis is fraught with danger. Thus fine clinical judgment and reliable laboratories are needed to avoid therapeutic complications. Restraint is recommended in definitive treatment of fibrinolysis and consumptive coagulopathy, and measures designed to reverse the shock and stabilize the patient are emphasized.

Myeloproliferative Diseases

The polycythemic patient, particularly with marked thrombocytosis, is a major surgical risk. Operations should be considered only for the most grave surgical emergency. If possible, the operation should be deferred until medical management has effected normal blood volume, hematocrit level, and platelet count. Spontaneous thrombosis is a complication of polycythemia vera and can be explained in part by increased blood viscosity, increased platelet count, and increased tendency toward stasis. Paradoxically, a significant tendency toward spontaneous hemorrhage also is noted in these patients.

Myeloid metaplasia frequently represents part of the natural history of polycythemia vera. Approximately 50 percent of patients with myeloid metaplasia are postpolycythemic. There is evidence suggesting qualitative platelet abnormalities in these patients. Abnormalities in platelet factor 3 release and in platelet aggregation with *ADP* have been demonstrated.

Treatment

Thrombocytosis can be reduced by the careful administration of alkylating agents such as busulfan or chlorambucil. Elective surgical procedures should be delayed weeks to months after institution of treatment. Ideally the hematocrit level should be kept below 48 percent and the platelet count under 400,000/mm³. Before operation, a thorough laboratory examination of hemostatic function should be conducted. When an emergency procedure is required, the erythremic and thrombocytotic states should be reduced by phlebotomy and replacement of the blood removed with lactated Ringer's solution. The operation, at all times, must be performed fastidiously.

Other Diseases

Illnesses resulting in severe impairment of hepatic function may limit synthesis of plasma factors essential to normal coagulation. The patient with advanced cirrhosis may be lacking in factors of the prothrombin complex (II, V, VII, X), as well as factor XIII. In addition, there may be increased fibrinolysis as a result of failure of the liver to clear plasminogen activators.

Other diseases, such as macroglobulinemia, may be associated with the abnormal production of proteins that coat the platelets and interfere with their function. Multiple myeloma and the disorders associated with the excessive production of cryoglobulins may also bind certain blood-clotting factors.

Anticoagulation and Bleeding

Spontaneous bleeding may be a complication of anticoagulant therapy with either heparin or the coumarin and indanedione derivatives. The incidence of bleeding complications related to heparin is reduced with a continuous infusion technique, regulating the *PTT* between 60 and 100 s (control: 30 to 35 s). An exaggerated response to oral anticoagulants may occur if dietary vitamin K is inadequate. The anticoagulant effect of the coumarins is consistently reduced in patients receiving barbiturates, and increased coumarin requirements have also been documented in patients taking contraceptives, other estrogen-containing compounds, corticosteroids, and adrenocorticotrophic hormone (ACTH). Therefore, reduced anticoagulant dosage should be instituted after discontinuance of any of these drugs. Medications known to increase the effect of oral anticoagulants include phenylbutazone, the cholesterol-lowering agent clofibrate, anabolic steroids (norethandrolone), d-thyroxine, glucagon, quinidine, and a variety of antibiotics.

Unexplained bleeding in medical and paramedical personnel occasionally is due to self-induced anticoagulation. The onset of hematuria or melena in the patient receiving anticoagulants should be investigated, since it has been shown that anticoagulants may unmask underlying tumors. Patients with bleeding secondary to anticoagulation may present only with epistaxis, gastrointestinal hemorrhage, or hematuria. Physical examination, however, almost always reveals other signs of bleeding such as ecchymoses, petechiae, or hematoma. Bleeding secondary to anticoagulation therapy is not an uncommon cause of rectus sheath hematoma, simulating appendicitis, and intramural intestinal or

retroperitoneal hematoma.

Surgical intervention may prove necessary in patients receiving anticoagulation therapy. Increasing experience suggests that surgical treatment can be undertaken without discontinuing the anticoagulant program. The risk of thrombotic complications reportedly is increased when anticoagulation therapy is discontinued suddenly. If so, this may not be related to what has been called the “rebound phenomenon” but may represent an event in a patient who has an underlying thrombotic tendency. When the clotting time is less than 25 min in the heparinized patient or when the *PT* is greater than 20 percent of normal in a patient on a coumarin drug, reversal of anticoagulant therapy may not be necessary. Meticulous surgical technique is mandatory, and the patient must be observed closely.

Certain surgical procedures should not be performed in the face of anticoagulation. In sites where even minor bleeding can cause great morbidity, e.g., the central nervous system and the eye, anticoagulants should be discontinued and, if necessary, reversed. Because of the added problem of local fibrinolysis, prostatic surgical treatment should not be carried out in a patient on anticoagulants. Procedures requiring blind needle introduction should be avoided. Deaths have been reported after sympathetic block for peripheral vascular disease in patients receiving anticoagulation.

Emergency operation occasionally is necessary in patients who have been heparinized as treatment for deep venous thrombosis. The first step in managing these patients is discontinuation of heparin; this may be sufficient if the operation can be delayed for several hours. For more rapid reversal, 1 mg of protamine sulfate for every 100 units of heparin most recently administered is immediately effective. For each hour that has elapsed since the last heparin dose, the amount of protamine should be halved. The formation of both extrinsic and intrinsic prothrombinase can be retarded, prolonging the one-stage *PT* test and the *PTT* test. Some patients exhibit the phenomenon of “heparin rebound” after apparently adequate heparin neutralization with protamine; prolongation of the clotting time recurs after adequate postoperative antagonism of the heparin, which can contribute to postoperative bleeding. In the author's experience, this is the major cause of “unexplained” postoperative bleeding after cardiac and vascular surgical procedures. Activation of fibrinolysis and thrombocytopenia may also contribute to this problem.

Bleeding infrequently is related to hypoprothrombinemia if the prothrombin concentration is greater than 15 percent. In the elective surgical patient receiving coumarin-derivative therapy sufficient to effect anticoagulation, the drug can be discontinued several days before operation and the prothrombin concentration then checked. A level greater than 50 percent is considered safe. If emergency surgical treatment is required, parenteral injection of vitamin K can be used. Since the reversal effect may take 6 h, transfusion of whole blood or, preferably, freshly frozen plasma may be required. Parenteral administration of vitamin K also is indicated in elective surgical treatment of patients with biliary obstruction, malabsorption, and hypoprothrombinemia. The drug should result in a normal *PT*. By contrast, if the hypoprothrombinemia is related to hepatocellular dysfunction, vitamin K therapy is ineffective and should not be prolonged over 1 week if no response is noted. Vitamin K is an oxidant, and one must be aware that patients with red cell enzyme deficiencies may sustain hemolysis after its administration.

Cardiopulmonary Bypass

Overheparinization, heparin rebound, inadequate protamine neutralization, protamine excess, and thrombocytopenia all have been indicted as causes of excessive bleeding in patients undergoing cardiopulmonary bypass. *DIC* is difficult to document in most patients. The predisposing factors that seem to be associated with excessive bleeding are prolonged perfusion times, prior use of oral anticoagulants, cyanotic heart disease, hypothermia, and prior use of antiplatelet drugs. It is currently believed that the two factors most important in triggering excessive bleeding associated with cardiopulmonary bypass are excessive fibrinolysis and platelet function defects, with the latter the more important element.

The laboratory evaluation of patients with bleeding should include *PT*, *PTT*, complete blood count (CBC) and platelet count, peripheral blood smear examination, and measurement of fibrin degradation products. Heparin assay can indicate the heparin level; plasminogen and plasmin assays are also available.

The management of cardiopulmonary bypass hemorrhage should include the empiric administration of 6 to 8 units of platelet concentrates as rapidly as possible. If hyperheparinemia is believed to be the major factor, 25 percent of the calculated dose of protamine should be administered and repeated every 30 to 60 min until the bleeding ceases. If there is laboratory evidence of excess fibrinolysis, *EACA* should be given at an initial dose of 5 to 10 g followed by 1 to 2 g/h until bleeding ceases. *EACA* may be associated with ventricular arrhythmia, hypotension, and hypokalemia. Aprotinin, a protease inhibitor that acts as an antifibrinolytic agent, has been shown to reduce transfusion requirements associated with cardiac surgery and orthotopic liver transplantation. Desmopressin acetate is also effective in reducing blood loss during cardiac surgery.

TESTS OF HEMOSTASIS AND BLOOD COAGULATION (TABLE 3-3)

The most important assessment of hemostasis is a careful history and physical examination. Only the history can indicate whether the patient has a hemorrhagic diathesis. Rather than asking a patient if he or she is a “bleeder,” specific questions should be asked. These should include queries to determine whether there was untoward bleeding during a major surgical procedure, or if there was any bleeding after a minor operation such as tonsillectomy, circumcision, or dental extraction, or if spontaneous bleeding was ever experienced. If there is any suggestion of a bleeding diathesis, the age of onset and family history is helpful to determine whether a hereditary or acquired defect should be investigated. Questions should uncover a history of exposure to toxic agents, oral anticoagulants, and drugs that might interfere with hemostasis. Aspirin and ibuprofen are two of the more common medications in this category. A history of a recent regimen of broad-spectrum antibiotics should alert the physician to the possibility of a deficiency of vitamin K-dependent clotting factors. Patients with malignant disease may have a variety of abnormalities, such as compensated intravascular coagulation and increased circulating fibrin complexes. Complex hemostatic disorders may accompany liver and renal failure.

Platelet Count

Because thrombocytopenia is the most common abnormality of hemostasis in the surgical patient, determination of the level of circulating platelets is a critical screening test. Direct enumeration of blood platelets can be accomplished quite accurately. *Spontaneous* bleeding only rarely can be related to thrombocytopenia with platelet counts greater than 40,000/mm³. Platelet counts of 60,000 to 70,000/mm³ usually are sufficient to provide adequate hemostasis after trauma or surgical procedures if other hemostatic factors are normal. An abnormal count should be confirmed by inspection of the blood smear.

When an area is examined where the red blood cells display their customary central pallor and where few of the red blood cells overlap one another, 15 to 20 platelets per oil immersion field should be noted. If the blood is not anticoagulated before the smear is prepared, as many as half of these may be in clumps of three or four platelets. A well-stained blood smear that fails to display more than three or four platelets in at least every other oil immersion field can be considered significantly thrombocytopenic. In this situation, the patient's platelet count generally is less than 75,000/mm³. Blood smears that must be searched because platelets appear in only every four or five oil immersion fields usually represent platelet counts of fewer than 40,000/mm³. If coverslip smears have been prepared, the coverslips always should be mounted as matched pairs. Platelets occasionally stick to one of the coverslips, and examination of both will obviate a false impression of thrombocytopenia. Lightly stained blood smears may appear thrombocytopenic if the platelets are not prominent enough to attract the examiner's attention.

Inspection of the blood smear has the additional advantage of permitting the examiner to identify other pathologic features that may have meaning in the care of the patient. The presence of nucleated red blood cells or abnormal white cells can provide information important to the diagnosis. The presence of giant platelets or large fragments of megakaryocyte cytoplasm will alert the examiner to possible pathologic platelet function.

Bleeding Time

Bleeding time provides an assessment of both the interaction between platelets and a damaged blood vessel and the formation of the platelet plug. Bleeding time may be abnormal in patients with thrombocytopenia, qualitative platelet disorders, von Willebrand's disease, and also in some patients with factor V deficiency or hypofibrinogenemia. Aspirin ingested within 1 week will affect the results. The tests can be performed by a variety of techniques that do not have the same normal times or the same degree of accuracy. The Duke method of measuring bleeding time, performed by incising the most dependent portion of the earlobe and measuring the time lapse until the bleeding ceases, normally should not exceed 3½ min. The modified Ivy method has an upper limit of normal of 7 min.

Other Tests of Platelet Function

Platelet aggregation can be assessed with a variety of induction agents to uncover specific abnormalities. The results may be affected by venipuncture, blood pH, temperature, duration of storage, and the equipment itself. The degree of abnormality detected by the test is not correlated with the extent of untoward bleeding. Aspirin is the most common cause of platelet aggregation abnormality. Failure of platelets to aggregate with the addition of arachidonic acid indicates an aspirin effect. The failure of platelets to aggregate with *ADP*, epinephrine, and collagen is characteristic of Glanzmann's thrombasthenia. Abnormal platelet aggregation with ristocetin occurs in von Willebrand's disease and in Bernard-Soulier syndrome.

The ability of the platelets to liberate platelet factor 3 (phospholipid), essential in tiny amounts at several stages of the blood-clotting process (see Fig. 3-3), also can be measured. Impairment of platelet factor 3 release has been reported in conditions described as *thrombocytopathia*. This defect can represent a primary disease entity, but similar impairment has been described as a secondary phenomenon in uremia and liver disease. The inability of the platelet to make platelet factor 3 available for the clotting process may be a part of a more fundamental surface membrane abnormality. The ability of *ADP*, epinephrine, collagen, and arachidonic acid to liberate serotonin, b-thromboglobulin, or platelet factor 4 can be measured.

Prothrombin Time

This test measures the speed of the events described earlier as the extrinsic pathway of blood coagulation. A tissue source of procoagulant (thromboplastin), a lipoprotein, is added with calcium to an aliquot of citrated plasma and the clotting time determined. The laboratory should establish a normal dilution curve and normal values daily. The PT will be prolonged in the presence of even minute amounts of heparin. The presence of heparin, by its antithrombin action, will artificially prolong the clotting time of the mixture so that it appears that the prothrombin complex is low. Accordingly, an accurate prothrombin determination cannot be carried out in a patient receiving anticoagulation treatment with heparin until the heparin has disappeared from the plasma. This should be at least 5 h after the last intravenous dose. The amount of heparin used to maintain patency of an intravenous line is usually insufficient to alter the PT.

The use of tissue procoagulants in the test eliminates the roles of factors VIII, IX, XI, XII, and platelets. Properly done, the test will detect deficiencies of factors II, V, VII, X, and fibrinogen. The one-stage PT is the preferred method of controlling anticoagulation with the coumarin and indanedione drugs.

Partial Thromboplastin Time

The PTT is a screening test for the intrinsic clotting pathway. The in vitro clotting system now is sensitive to factors VIII, IX, XI, and XII, as well as the factors normally detected by

the one-stage *PT*. The range of normal with this test varies with the product used. The patient's plasma must be compared with a normal control sample.

The PTT, when used in conjunction with the one-stage PT, can help to place a clotting defect in the first or second stage of the clotting process. If the PTT is prolonged and the one-stage PT is normal, factors VIII, IX, XI, or XII may be deficient. If the PTT is normal and the one-stage PT is prolonged, a single or multiple deficiency of factors II, V, VII, or X or of fibrinogen may be present. The PTT is also abnormal in the presence of circulating anticoagulants or during heparin administration. It may be prolonged when heparin is used to maintain the patency of an intravenous line. The sensitivity of the test is such that only extremely mild cases of factor VIII or IX deficiency may be missed.

Thrombin Time

This test is of value in detecting qualitative abnormalities in fibrinogen and in detecting circulating anticoagulants and inhibitors of fibrin polymerization. The clotting time of the patient's plasma is measured after the addition of a standard amount of thrombin to a fixed volume of plasma. Control samples of normal plasma must be run in parallel. Failure of the clot to form, in the absence of circulating inhibitors such as heparin or the fibrinolytic degradation products of fibrin and fibrinogen, is consistent with severe diminution of fibrinogen, usually well below 100 mg/dL. It is also prolonged when fibrinolysis is taking place.

Other Tests of Coagulation

The fibrinogen level can be determined by clotting-time measurements or gravimetrically. Specific assays of coagulation factors are performed by measuring the clotting time of plasma from patients congenitally lacking in one of these factors and noting the effect of the addition of each factor. Relatively simple tests permit identification of circulating anticoagulants. The simplest of these are based on the retardation of clotting of normal recalcified plasma by varying mixtures of the test plasma. The sensitivity of such tests usually can be increased by incubating the test plasma with the normal plasma for 30 min at body temperature before recalcification. Detection of factor XIII deficiency requires a special test.

Tests of Fibrinolysis

Fibrin degradation products (FDP) can be measured by immunologic methods. Normally, dissolution of a recently formed blood clot will not occur for 48 h or more. When fibrinolysis is a significant factor in hemostatic failure, dissolution of the whole blood clot is observed in 2 h or less. The test has the disadvantage of being time-consuming in a circumstance where time may be of the essence. In addition, a false impression of increased fibrinolytic activity may be gained from clots formed in patients with high hematocrit levels or in thrombocytopenia, in which red cells may fall away from the clot. The euglobulin clot lysis time and dilute whole-blood or plasma-clot-lysis time are more sensitive indices and permit more rapid evaluation of fibrinolysis.

The *thromboelastogram* is a graphic representation of clotting. The record obtained provides information about the clotting time, the speed of fibrin polymerization, and the clot's strength and tendency toward dissolution.

EVALUATION OF THE SURGICAL PATIENT AS A HEMOSTATIC RISK

Preoperative Evaluation of Hemostasis

The patient's history provides meaningful clues to the presence of a bleeding tendency. It is reasonable to use a questionnaire on which the patient indicates: (1) prolonged bleeding or swelling after biting the lip or tongue, (2) bruises without apparent injury, (3) prolonged bleeding after dental extraction, (4) excessive menstrual bleeding, (5) bleeding problems associated with major and minor operations, (6) medical problems receiving a physician's attention within the past 5 years, (7) medications including aspirin or remedies for headache taken within the past 10 days, and (8) a relative with a bleeding problem.

Four levels of concern have been proposed on the basis of the history and surgical procedure being considered. At Level I, the history is negative and the procedure contemplated is relatively minor, e.g., breast biopsy or hernia repair: no screening tests are recommended. At Level II, the history is negative, screening tests may have been performed in the past, and a major operation is planned, but the procedure usually is not attended by significant bleeding: a platelet count and blood smear and *PTT* are recommended to detect any thrombocytopenia, circulating anticoagulant, or intravascular coagulation. Level III pertains to the patient whose history is suggestive of defective hemostasis and also to the patient who is to undergo an operative procedure in which hemostasis may be impaired, e.g., operating using pump oxygenation or cell savers, or procedures in which a large, raw surface is anticipated. Level III also pertains to situations in which minimal postoperative bleeding could be injurious, such as intracranial operations. At this level, a platelet count and bleeding time test should be performed to assess platelet function; a *PT* and *PTT* should be used to assess coagulation, and the fibrin clot should be incubated to screen for abnormal fibrinolysis. Level IV pertains to patients who present with a history highly suggestive of a hemostatic defect. A hematologist should be consulted, and, in addition to the tests prescribed for Level III patients, the bleeding time test should be repeated 4 h after the ingestion of 600 mg of aspirin, provided that the operation is scheduled to take place 10 or more days after this study. In the case of an emergency procedure, platelet aggregation tests using *ADP*, collagen, epinephrine, and ristocetin should be performed, and a *TT* is indicated to detect any dysfibrinogenemia or a circulating, weak, heparin-like anticoagulant. Patients with liver disease, renal failure, obstructive jaundice, and the possibility of disseminated malignant disease should have a platelet count, PT, and PTT performed preoperatively. In uremic patients the most common deficit is a qualitative platelet abnormality. This is best detected by the bleeding time test.

Evaluation of Excessive Intraoperative or Postoperative Bleeding

Excessive bleeding during or shortly after a surgical procedure may be due to one or more of the following factors: (1) ineffective local hemostasis, (2) complications of blood transfusion, (3) a previously undetected hemostatic defect, (4) consumptive coagulopathy, and/or (5) fibrinolysis. Excessive bleeding from the field of the procedure, unassociated with bleeding from other sites, e.g., central venous pressure line, intravenous line, or tracheostomy, usually suggests inadequate mechanical hemostasis rather than a defect in the biologic process. An exception to this rule applies to operations on the prostate, pancreas, and liver because operative trauma may stimulate local plasminogen activation and lead to increased fibrinolysis on the raw surface. In these circumstances 24 to 48 h interruption of plasminogen activation by the administration of *EACA* may prove effective.

Although one may be reasonably certain on clinical grounds that surgical bleeding is related to local problems, laboratory investigation must be confirmatory. Prompt examination should be made of the blood smear to determine the number of platelets, and an actual platelet count should be done if the smear is equivocal. A *PTT*, a one-stage *PT*, and a *TT* all can be determined within minutes. Correct interpretation of the results should confirm the clinical impression or identify the problem.

As pointed out previously, massive blood transfusion is a well-documented cause of thrombocytopenia. Although most patients who receive 10 units or more of banked blood within a period of 24 h will be measurably thrombocytopenic, this is usually *not* associated with hemostatic failure. Therefore, prophylactic administration of platelets is not indicated, but if there is evidence of diffuse bleeding, 8 to 10 packs of fresh platelet concentrates should be given empirically, because no clear association has been documented between the platelet count, bleeding time, and the occurrence of profuse bleeding.

Another cause of hemostatic failure related to the administration of blood is a hemolytic transfusion reaction. The first hint of a transfusion reaction in an anesthetized patient may be diffuse bleeding in an operative field that had previously been dry. The pathogenesis of this bleeding is thought to be related to the release of *ADP* from hemolysed red cells, resulting in diffuse platelet aggregation, after which the platelet clumps are swept out of the circulation. Release of procoagulants may result in progression of the clotting mechanism and intravascular defibrination. In addition, the fibrinolytic mechanism may be triggered.

Transfusion purpura is an uncommon cause of thrombocytopenia and associated bleeding after transfusion. When this occurs the donor platelets are of the uncommon PI_{A1} group. These platelets sensitize the recipient, who makes antibody to the foreign platelet antigen. The foreign platelet antigen does not completely disappear from the recipient circulation but seems to attach to the recipient's own platelets. The antibody, which attains a sufficient titer within 6 or 7 days after the sensitizing transfusion, then destroys the recipient's own platelets. The resultant thrombocytopenia and bleeding may continue for several weeks. This uncommon cause of thrombocytopenia should be considered if bleeding follows transfusion by 5 or 6 days. Platelet transfusions are of little help in the management of this syndrome, since the new donor platelets usually are subject to the binding of antigen and damage from the antibody. Corticosteroids may be of some help in reducing the bleeding tendency. Posttransfusion purpura is self- limited, and the passage of several weeks inevitably leads to subsidence of the problem.

DIC and disseminated fibrinolysis occur intraoperatively or postoperatively when control mechanisms fail to restrain the hemostatic process to the area of tissue damage. Either process can cause diffuse bleeding and can be caused by trauma, incompatible transfused blood, sepsis, necrotic tissue, fat emboli, retained products of conception, toxemia of pregnancy, large aneurysms, and liver diseases. It is important to distinguish between the two processes or the dominant element causing intraoperative or postoperative bleeding. No single test can confirm or exclude the diagnosis or distinguish between the two disorders. The combination of thrombocytopenia, defined by smear or platelet count, positive plasma protamine test for fibrin monomers, a low fibrinogen level, and an elevated level of *FDP* provides strong indications for DIC. The euglobulin lysis time provides a method of detecting diffuse fibrinolysis.

Diffuse intraoperative and postoperative bleeding is a complication of biliary tract surgery in cirrhotic patients. This has been related to portal hypertension and coagulopathy associated with chronic liver disease. The tests used to distinguish *DIC* from fibrinolysis pertain. The therapeutic approach includes the intravenous administration of vasopressin to effect a temporary reduction in portal hypertension, and *EACA* to correct the increased fibrinolysis.

An operation performed in a patient with sepsis sometimes is attended by continued bleeding. Severe hemorrhagic disorders due to thrombocytopenia have occurred as a result of gram-negative sepsis. The pathogenesis of endotoxin-induced thrombocytopenia has been studied in detail, and it has been suggested that a labile factor, possibly factor V, is necessary for this interaction. Defibrination and hemostatic failure also may occur with meningococcemia, *Clostridium perfringens* sepsis, and staphylococcal sepsis. Hemolysis appears to be one mechanism in sepsis leading to defibrination. Evaluation of these patients includes platelet count, *PT*, *PTT*, and *TT*.

LOCAL HEMOSTASIS

Surgical bleeding, even when alarmingly excessive, is usually caused by ineffective local hemostasis. The goal of local hemostasis is to prevent the flow of blood from incised or transected blood vessels. This may be accomplished by interrupting the flow of blood to the involved area or by direct closure of the blood vessel wall defect. The techniques may be classified as mechanical, thermal, or chemical.

Mechanical Procedures

The oldest mechanical method of effecting closure of a bleeding point or preventing blood from entering the area of disruption is digital pressure. When pressure is applied to an artery proximal to an area of bleeding, profuse bleeding is reduced, permitting more definitive action. A classic example is the Pringle maneuver of occluding the hepatic artery in the hepatoduodenal ligament as a method of controlling bleeding from a transected cystic artery or from the surface of the liver. Direct digital pressure over a bleeding site, such as

a lateral rent in the inferior vena cava, is also effective. The finger has the advantage of being the least traumatic vascular hemostat. All clamps, including the so-called atraumatic vascular clamps, do result in damage to the intimal wall of the blood vessel. The most obvious disadvantage of digital pressure is that it cannot be used permanently.

The hemostat also represents a temporary mechanical device to stem bleeding. In smaller and noncritical vessels, the trauma and adjacent tissue necrosis associated with the application of a hemostat are of little consequence. These minor disadvantages are outweighed by the mechanical advantage that the instrument offers to subsequent ligation. When bleeding occurs from a vessel that should be preserved, relatively atraumatic hemostats should be employed to limit the extent of intimal damage and subsequent thrombosis.

In general, a ligature replaces the hemostat as a permanent method of effecting hemostasis in a single vessel. When a vessel is transected, a simple ligature usually is sufficient. For large arteries with pulsation and longitudinal motion, transfixion suture to prevent slipping is indicated. When the bleeding site is from a lateral defect in the blood vessel wall, suture ligatures are required. The adventitia and media constitute the major holding forces within the walls of large vessels, and therefore multiple fine sutures are preferable to fewer larger sutures.

Historically, Aulus Cornelius Celsus devised the use of ligatures in the first century a.d. Because of the strong influence of Galen, who was inclined to cautery, this method did not gain popularity. Paré, in 1552, rediscovered the principle of ligature. In 1800 Physick used absorbable sutures of buckskin and parchment. In 1858 Simpson introduced the wire suture, and in 1881 Lister employed chromic catgut. Halsted, in the early 1900s, emphasized the importance of incorporating as little tissue as possible in the suture and indicated the advantages of silk. In 1911 Cushing reported on the use of silver clips to effect hemostasis in delicate vessels in critical areas. A wide variety of staples made of different metals, relatively inert in tissue, have been used.

All sutures represent foreign material, and their selection is based on the characteristics of the material and the state of the wound. Nonabsorbable sutures, such as silk, polyethylene, and wire, evoke less tissue reaction than absorbable materials, such as catgut, polyglycolic acid (Dexon), and polyglactin (Vicryl). The latter are preferable, however, in the face of overt infection. The presence of nonabsorbable material in an infected wound can lead to extrusion or sinus tract formation. Wire is the least reactive of the nonabsorbable sutures but the most difficult to handle. Monofilament wire and coated sutures have an advantage over multifilament sutures in the presence of infection. The latter tend to fragment and permit sinus formation.

Diffuse bleeding from multiple transected vessels may be controlled by mechanical techniques that employ pressure directly over the bleeding area, pressure at a distance, or generalized pressure. These techniques are based on the premise that as pressure and flow are decreased in the area of vascular disruption, a clot will develop. As a standard procedure of military surgeons in the seventeenth century, pressure at a distance was effected by application of tourniquets and other pressure devices at pressure points proximal to bleeding sites. Now it is generally believed that direct pressure is preferable and is not attended by the danger of tissue necrosis associated with prolonged use of tourniquets. Gravitational suits have been used to create generalized pressure and temporarily decrease bleeding from ruptured major intraabdominal vessels.

Direct pressure applied by means of packs affords the best method of controlling diffuse bleeding from large areas. Rarely is it necessary to leave a pack at the bleeding site and remove it at a second sitting. If this is done, several days should elapse before removal, and the possibility of recurrent bleeding should be anticipated. The question of whether hot wet packs or cold wet packs should be applied has been investigated. Unless the heat is so great as to denature protein, it may actually increase bleeding, whereas cold packs promote hemostasis by inducing vascular spasm and increasing endothelial adhesiveness. Bleeding from cut bone may be controlled by packing beeswax in the area. This material effects pressure and is relatively nonirritating to the body.

Thermal Agents

Galen's favoring of cautery influenced medicine for 1500 years, until the teachings of Paré were appreciated. The use of cautery was revitalized in 1928, when Cushing and Bovie applied this technique for effecting hemostasis of delicate vessels in recessed areas, such as the brain. Heat achieves hemostasis by denaturation of protein, which results in coagulation of large areas of tissue. With actual cautery, heat is transmitted from the instrument by conduction directly to the tissue; with electrocautery, heating occurs by induction from an alternating-current source.

When electrocautery is employed, the amplitude setting should be high enough to produce prompt coagulation but not so high as to set up an arc between the tissue and the cautery tip. This avoids burns outside the operative field and prevents the exit of current through electrocardiographic leads or other monitoring devices. A negative plate should be placed beneath the patient whenever cautery is employed to avoid severe skin burns. The advantage of cautery is that it saves time; its disadvantage is that more tissue is necrosed than with precise ligature. Certain anesthetic agents cannot be used with electrocautery because of the hazard of explosion.

A direct current can also result in electrical hemostasis. Since the protein moieties and cellular elements of blood have a negative surface charge, they are attracted to the positive pole, where a thrombus is formed. Direct currents in the 20- to 100-mA range have been applied to control diffuse bleeding from large serous surfaces. Argon gas has been applied successfully to the control of bleeding from superficial erosions.

At the other end of the thermal spectrum, cooling has been applied to control bleeding, particularly from the mucosa of the esophagus and stomach. Generalized hypothermia is of little avail, since in order to reduce the blood flow to visceral organs, the systemic temperature must be brought down to the level of 35°C. At this point shivering and ventricular fibrillation may occur. Thrombocytopenia may also be a consequence of generalized cooling. Direct cooling with iced saline is effective and acts by increasing the local intravascular hematocrit concentration and decreasing blood flow by vasoconstriction.

Extreme cooling, i.e., cryogenic surgery, has been applicable particularly in gynecology and neurosurgery. Temperatures ranging between -20 to- 180°C are used, and freezing occurs around the tip of the cannula within 5 s. At temperatures of -20°C or below, the tissue, capillaries, small arterioles, and venules undergo cryogenic necrosis. This is caused by dehydration and denaturation of lipid molecules. The muscular walls of large arteries are an exception. Although the major arteries and blood may be frozen solid, the blood contained in these vessels does not clot. When thawing occurs, normal circulation is resumed.

Chemical Agents

Chemical agents vary in their hemostatic action. Some are vasoconstrictive, while others have coagulant properties. Still others are relatively inert but possess hygroscopic properties which increase their bulk and aid in plugging disrupted blood vessels.

Epinephrine, applied topically, induces vasoconstriction, but extensive application can result in considerable absorption and systemic effects. The drug generally is used on oozing sites in mucosal areas, e.g., during tonsillectomy.

Historically, skeletal muscle was one of the first materials with locally hemostatic properties to be employed, its use having been introduced by Cushing in 1911. Shortly thereafter, hemostatic fibrin was manufactured. The properties required for local hemostatic materials include handling ease, rapid absorption, hemostatic action independent of the general clotting mechanism, and they should be nonirritating. The most widely used of the commercially available materials are gelatin foam (Gelfoam), oxidized cellulose (Oxycel), oxidized regenerated cellulose (Surgicel), and micronized collagen (Avitene). All these materials act, in part, by transmitting pressure against the wound surface, and the interstices provide a scaffold on which the clot can organize (Table [3-4](#)).

Gelfoam is made from animal skin gelatin that has been denatured. In itself Gelfoam has no intrinsic hemostatic action, but it can be used in combination with topical thrombin, for which it serves as an absorbable carrier. Its main hemostatic activity is related to the contact between blood and the large surface area of the sponge and to the pressure exerted by the weight of the sponge and absorbed blood. Before Gelfoam is applied, the sponge should be moistened in saline or thrombin solution and all the air should be removed from the interstices.

Oxycel and Surgicel are altered cellulose materials capable of reacting chemically with blood and producing a sticky mass that functions as an artificial clot. These substances are relatively inert and are removed by liquefaction in 1 to 4 weeks. They should be dry when they are applied. Like Gelfoam, these materials are nontoxic and relatively nonirritating but are somewhat detrimental to wound healing and require phagocytosis to be removed. Surgicel has been shown to have an antibacterial effect. Microcrystalline collagen has been shown to be as effective as other materials as a topical hemostatic agent for large oozing surfaces.

Fibrin glue is commercially available in Europe and Canada but not in the United States, because of the potential of disease transmission when fibrinogen is obtained from pooled plasma. Single-donor fibrinogen can be mixed with bovine thrombin to make the sealant. The glue is particularly effective in controlling surface bleeding from the liver and spleen.

TRANSFUSION

Background

In 1967, the tercentennial anniversary of the transfusion of blood into human beings was celebrated. In June 1667 Jean-Baptiste Denis and a surgeon, Emmerez, transfused blood from a sheep into a 15-year-old boy who had been bled many times as treatment for fever. The patient apparently improved, and a successful experience was reported simultaneously in another patient. Because of two subsequent deaths associated with transfusion from animals to humans, criminal charges were brought against Denis. In April 1668 further transfusions in humans were forbidden unless approved by the Faculty of Medicine in Paris. It was not until the nineteenth century that human blood was recognized as the only appropriate replacement. In 1900 Landsteiner and his associates introduced the concept of blood grouping and identified the major A, B, and O groups. In 1939 the Rh group was recognized. The introduction of various preservative solutions, such as *acid-citrate-dextrose (ACD)*, *citrate- phosphate-dextrose (CPD)*, and citrate-phosphate-double-dextrose adenine (CP2D-A) and newer additive solutions extended the shelf life of blood to up to 6 weeks.

Preservation of blood and its constituents has been achieved by freezing, and emphasis has been placed on the use of plasma expanders and component therapy.

Characteristics of Blood and Replacement Therapy

Blood

Blood has been described as a vehicular organ that perfuses all other organs. It provides transportation of oxygen to satisfy the body's metabolic demands and removes the by-product carbon dioxide. Blood also transports chemical nutrients for, and waste products from, metabolic activity. Homeostatic governors, including hormones, coagulation factors, and antibodies, are carried to and from appropriate sites within the fluid portion of the blood. Red blood cells, with their oxygen-carrying capacity; white blood cells, which function in body defense processes; and platelets, which contribute to the hemostatic process, comprise the formed elements.

Replacement Therapy Banked Whole Blood

Banked whole blood is now rarely indicated and rarely available. With the new preservatives, the shelf life has been extended to 40 ± 5 days. At least 70 percent of the transfused erythrocytes remain in the circulation for 24 h after transfusion and are viable. The changes in the red cell that occur during storage include reduction of intracellular adenosine triphosphate (ATP) and *2,3-diphosphoglycerate (2,3-DPG)*, which alters the curve of oxygen dissociation from hemoglobin, decreasing oxygen transport function. Banked blood is a poor source of platelets because platelets lose their ability to survive transfusion after 24 h of storage. Among the clotting factors, II, VII, IX, and XI are stable in banked blood. Within 21 days of storage, the pH decreases from 7.00 to 6.68, and the lactic acid level increases from 20 to 150 mg/dL. The potassium concentration rises steadily to 32 mEq/dL, and the ammonia concentration rises from 50 to 680 mg/dL at the end of 21 days for CPD whole blood. The hemolysis that occurs during storage is insignificant.

Typing and Crossmatching

In selecting blood for transfusion, serologic compatibility is established routinely for the recipients' and donors' A, B, O, and Rh groups. Crossmatching between the donors' red cells and the recipients' sera (the "major" crossmatch) is performed. As a rule, Rh-negative recipients should be transfused only with Rh-negative blood. Since this group represents 15 percent of the donor population, the supply may be limited. If the recipient is an elderly male who has not been transfused previously, the transfusion of Rh-positive blood is acceptable if Rh-negative blood is unavailable. Anti-Rh antibodies form within several weeks of transfusion. If further transfusions are needed within a few days, more Rh-positive blood can be used. Rh-positive blood should not be transfused to Rh-negative females who are capable of childbearing. Administration of hyperimmune anti-Rh globulin to Rh-negative women shortly before or after childbirth largely eliminates Rh disease in subsequent offspring.

In the patient who is receiving repeated transfusions, serum drawn not more than 72 h before crossmatching should be utilized for matching with cells of the donor. Emergency blood transfusion can be performed with type O blood. O-negative and type-specific red blood cells are equally safe for emergency transfusion. Problems are associated with the administration of 4 or more units of O-negative blood because there is a significant increase in the risk of a hemolytic reaction.

In patients with malignant lymphoma and leukemia, cryoglobulins may be present, and the blood should be administered through a blood warmer. If these antibodies are present in high titer, hypothermia may be contraindicated.

In patients with thalassemia who have been multiply transfused and, more particularly, with acquired hemolytic anemia, typing and crossmatching may be difficult, and sufficient time should be allotted during the preoperative period to accumulate blood that may be required during the operation. Crossmatching should always be carried out prior to the administration of dextran, since dextran interferes with the typing procedure.

Because banked blood may be stored for 40 ± 5 days, the use of autologous predeposit transfusion is growing. In otherwise healthy, nonanemic patients, up to 5 or 6 units of blood may be collected for use in elective surgical procedures. Patients may donate blood if the hemoglobin level exceeds 11 g/dL or if the hematocrit concentration is greater than 34 percent. The first procurement is performed 40 days before the planned operation and the last one, 3 days before the procedure. Donations can be scheduled at intervals of 4 to 5 days. Recombinant human erythropoietin (r-HuEPO) accelerates generation of red cells and allows for more frequent harvest for elective operative procedures.

Fresh Whole Blood

This term refers to blood that is administered within 24 h of its donation. It is rarely indicated. Because of the time requirements of testing for infectious diseases, fresh blood is only available untested. One unit of platelet concentrate has more viable platelets than 1 unit of fresh whole blood, which is also an inadequate source of factor VIII.

Packed Red Cells and Frozen Red Cells

Packed red cells is the product of choice for most clinical situations. Concentrated suspensions of cells can be prepared by removing most of the supernatant plasma after centrifugation. The preparation reduces but does not eliminate reaction caused by plasma components. It also reduces the amount of sodium, potassium, lactic acid, and citrate administered. Essentially it provides oxygen-carrying capacity.

Frozen red cells are not available for use in emergencies. They are often used for patients who have been previously sensitized because they have been selected for lack of certain antigens. The red cell viability is improved, and the ATP and *2,3-DPG* concentrations are maintained.

Leukocyte-Poor Washed Cells

This product is prepared by aspirating the buffy coat and supernatant plasma and passing them through a specific white-cell filter. The red cells then are washed with sterile isotonic solution. This should be done only for patients with demonstrated hypersensitivity to leukocytes or platelets (buffy coat reactions). Usually this syndrome is manifest by fever, chilly sensations, and urticaria due to plasma proteins in the absence of hemolysis.

Platelet Concentrates

The indications for platelet transfusion are as follows: thrombocytopenia due to massive blood loss and replacement with platelet-poor products, thrombocytopenia due to inadequate production, and qualitative platelet disorders. The preparations should be used within 120 h of blood donation. One unit of platelet concentrate has a volume of approximately 50 mL. Platelet preparations may transmit infectious diseases and account for allergic reactions similar to those caused by whole blood. When treating thrombocytopenic bleeding or preparing some thrombocytopenic patients for surgery, it is advisable to elevate the platelet levels to the range of 50,000 to 100,000/mm³ to provide continued protection. The development of isoimmunity remains one of the most important factors limiting the usefulness of platelet transfusion. Isoantibodies are demonstrable in about 5 percent of patients after 1 to 10 transfusions, in 20 percent after 10 to 20 transfusions, and in 80 percent after more than 100 transfusions. The use of HLA-compatible platelets addresses this problem.

Frozen Plasma and Volume Expanders

Frozen plasma prepared from freshly donated blood or fresh plasma is necessary to provide factors V and VIII. The other plasma clotting factors are present in banked preparations. The risk of infectious disease is the same whether fresh frozen plasma or whole blood/red cells is administered. Lactated Ringer's solution or buffered saline solution administered in amounts two to three times the estimated blood loss, is effective and is associated with fewer complications. Dextran or a combination of lactated Ringer's solution and normal human serum albumin are preferred for rapid plasma expansion. Commercially available dextran preparations probably should not be administered in amounts exceeding 1 L/day, since prolongation of bleeding time and hemorrhage can occur. Low-molecular-weight dextran, i.e., molecular weight of 30,000 to 40,000, has become popular because it possesses a higher colloidal pressure than plasma and effects some reversal of erythrocyte agglutination.

Concentrates

Antihemophilic concentrates are prepared from plasma and are available for the treatment of factor VIII deficiency. Some of these concentrates are twenty to thirty times as potent as an equal volume of fresh frozen plasma. The simplest factor VIII concentrate is the plasma cryoprecipitate. *Albumin* also has been concentrated, so 25 g may be administered and provide the osmotic equivalent of 500 mL of plasma. The advantage of albumin is that it is a hepatitis-free product.

Indications for Replacement of Blood or Its Elements

Improvement in Oxygen-Carrying Capacity

Oxygen-carrying capacity is primarily a function of the red cell. When anemia can be treated by specific therapy, such as erythropoietin, transfusion should be withheld. Acute anemias, such as hemolytic anemia, are more disabling physiologically than chronic anemia, since most patients with chronic anemia have undergone an adjustment to the condition. In pregnancy there is a moderate drop in hematocrit level, and transfusions are not indicated to correct the physiologic anemia of pregnancy before surgical treatment. The correction of chronic anemia before surgical treatment, though often performed, is difficult to justify. A 1988 National Institutes of Health Consensus Report challenged the dictum that a hemoglobin value of less than 10 g/dL or a hematocrit level of less than 30 percent indicates a need for a preoperative red cell transfusion. It is suggested that cardiac output does not increase significantly in healthy individuals until the hemoglobin value decreases to approximately 7 g/dL. Patients with chronic anemia and a hemoglobin value less than 7 g/dL in whom significant bleeding intraoperatively is not anticipated do not require a transfusion preoperatively. There is no correlation between anemia and dehiscence or severity of postoperative infection.

Blood volume may be replaced with dextran solution or lactated Ringer's solution with a reduction of the hemoglobin value to levels below 10 g and little demonstrable change in the effects of a reduction in oxygen-carrying capacity or the capacity to remove metabolic gaseous by-products. A stroma-free hemoglobin solution has been shown to have the ability to carry and exchange oxygen. Also, a whole blood substitute, Fluosol-DA, has been proposed as a solution with oxygen-handling capabilities.

Volume Replacement

The most common indication for blood transfusion in surgical patients is the replenishment of the circulating blood volume. It is difficult to evaluate the volume deficit accurately.

Values for "normal blood volume" are variable, and the techniques of measurement are relatively inaccurate when there is a rapidly changing situation, such as hemorrhage. Chronically ill and elderly patients may have a diminution of blood volume. In patients with cardiac decompensation, the blood volume may be greater than normal. Many patients with chronically reduced blood volume are well accommodated to that volume.

Measurement of hemoglobin or hematocrit levels is also used to interpret blood loss. These measurements are misleading in the face of acute blood loss, since the hematocrit level may be normal in spite of a severely contracted blood volume. It has been shown that, after a healthy adult male lost approximately 1000 mL of blood rapidly, the venous hematocrit fell only 3 percent during the first hour, 5 percent at 24 h, 6 percent at 48 h, and 8 percent at 72 h, thus indicating the time required for the body to restore blood volume.

Both the amount and the rate of bleeding are factors in the development of the signs and symptoms of blood loss. A healthy person can lose 500 mL in 15 min with only minor effects on the circulation and little change in blood pressure or pulse. Loss of 15 to 30 percent of blood volume (class II hemorrhage) is associated with tachycardia and decreased pulse pressure. Loss of 30 to 40 percent (class III hemorrhage) generally results in tachycardia, tachypnea, hypotension, oliguria, and changes in mental status.

Loss of blood may be evaluated in the operating room by estimating the amount of blood in the wound and on the drapes and by weighing sponges. The loss determined by weighing sponges is only about 70 percent of true loss. In patients who have normal preoperative blood values, blood loss up to 20 percent of total blood volume (TBV) is replaced with crystalloid solutions. Blood loss up to 50 percent of TBV is replaced with crystalloids and red blood cell concentrates. Blood loss above 50 percent of TBV is replaced with crystalloids, red blood cells, and albumin or plasma. Continued bleeding above 50 percent of TBV should receive the same components and fresh frozen plasma. If electrolyte solutions are used to replace blood volume, an amount three to four times the lost volume is required because of immediate diffusion into the interstitial space.

Replacement of Clotting Factors

Transfusion of platelets and/or proteins contributing to coagulation may be indicated in specific patients either before or during operation (Table 3-5). In the treatment of certain hemorrhagic conditions, it must be kept in mind that the clotting defects may be multiple. The efficacy of fresh frozen plasma in the management of coagulopathy in patients with liver disease and in patients receiving large amounts of volume replacement for acute blood loss is not well defined. There are insufficient data to specify criteria for transfusion of fresh frozen plasma. The initial volume of fresh frozen plasma needed for an effect on coagulation ranges from 600 to 2000 mL administered in 1 to 2 h. The rigid use of the *PT* and *PTT* to anticipate the effect of fresh frozen plasma is not justified.

Specific Indications Massive Transfusion

The term *massive transfusion* implies a single transfusion greater than 2500 mL, or 5000 mL transfused over a period of 24 h. The approximate percentages of *original* blood volume remaining after varying degrees of hemorrhage and transfusion are shown in Table 3-6. A variety of problems may attend the use of massive transfusion. Circulatory overload or *DIC* may occur. Dilutional thrombocytopenia, impaired platelet function, and deficiencies of factors V, VIII, and XI may occur. Routine alkalization is not advisable, since this could have an adverse effect on the oxyhemoglobin dissociation curve and presents an additional sodium load to a compromised patient. The increased potassium content of multiple units of stored blood does not provide clinical effects unless the patient is severely oliguric.

Citrate toxicity may be associated with massive transfusion, particularly in young children and patients with severe hypotension or liver disease. This toxicity is related to an excessive binding of ionized calcium and is usually corrected by spontaneous mobilization of calcium from bone. The physiologic consequences of citrate toxicity rarely have a significant effect. The function of hemoglobin is altered by storage, since the concentration of 2,3-DPG falls to a negligible level by the third week. This results in an increased affinity of the red blood cells for oxygen and a less efficient oxygen delivery system. In itself, reduction of 2,3-DPG may not have a significant effect, but when combined with acute anemia it may be an important factor.

When large transfusions are administered, a heat exchanger may be used to warm the blood, since hypothermia may cause a decrease in cardiac rate and output and a reduction in the blood pH. Warming the blood decreases significantly the frequency of intraoperative cardiac arrest.

The use of blood from many donors increases the possibility of hemolytic transfusion reaction due to incompatibility. This can be reduced by screening each potential donor in the pool and eliminating those who show possible incompatibility. Paradoxically, patients who survive a massive transfusion do not have a high probability of developing isoantibodies subsequently, and the risk is no greater than that from a single transfusion. The risk of infectious disease increases progressively with each succeeding unit. When administering massive transfusions, the pH, blood gases, and potassium should be measured regularly. Acidosis and abnormalities should be corrected. If diffuse bleeding occurs, coagulation screening tests and platelet counts should be performed and deficits corrected with frozen plasma and platelet concentrates.

Methods of Administering Blood

Routine Administration

The rate of transfusion depends on the patient's status. Usually 5 mL/min is administered for 1 min, after which 10 to 20 mL/min may be administered to complete routine transfusion. When marked oligemia is being treated, the first 500 mL may be given within 10 min, and the second 500 mL may be given equally rapidly in most cases. Cold blood may be used for this amount, but when larger amounts are administered, warm blood is desirable. As much as 1500 mL/min can be administered through two 7.5F catheters.

When large transfusions are administered, it is important not to overload the circulation, and the use of central venous pressure monitoring is particularly pertinent. There is no practical advantage in the use of intraarterial transfusion over the intravenous route in the treatment of oligemia. It has been shown that coronary flow and systemic arterial pressure respond as rapidly and to the same extent whether the blood is administered intravenously or intraarterially.

Other Methods

Blood may be instilled intraperitoneally or into the medullary cavity of the sternum and long bones. Intrasternal and intramedullary transfusion may be painful, and the rate of administration is limited. Approximately 90 percent of red cells injected intraperitoneally enter the circulation, but uptake is not complete for at least a week, and therefore the method is not suitable when immediate transfusion is required.

Intraoperative autotransfusion has become increasingly popular; it is a potentially life-saving adjunct to the management of trauma and is useful in elective operations in which multiple transfusions are likely to be required. Approximately 250 mL of blood can be retrieved, washed or filtered, and returned to the patient over a 5- to 6-min period. A comparison between cell washing and simple filtration revealed that filtration allowed a greater percentage of blood to be returned and was associated with less thrombocytopenia. Another approach to anticipated intraoperative large blood losses is hemodilution. At the onset of the procedure, red cells are removed while the intravascular volume is maintained with crystalloid or colloid. The reduced blood viscosity improves microcirculatory perfusion. The removed blood can then be retransfused during the operation to replace lost blood or be reinfused near the completion of the procedure.

Complications

Hemolytic Reactions

The incidence of nonfatal hemolytic transfusion reactions is approximately 1 per 6000 units of blood administered. Fatal hemolytic transfusion reactions occur once in every 100,000 units administered. Hemolytic reactions due to incompatibility of A, B, O, and Rh groups or many other independent systems may result from errors in the laboratory of a clerical or technical nature or the administration of the wrong blood at the time of transfusion. Hemolytic reactions are characterized by intravascular destruction of red blood cells and consequent hemoglobinemia and hemoglobinuria. Circulating haptoglobin is capable of binding 100 mg hemoglobin/dL plasma, and the complex is cleared by the reticuloendothelial system. When the binding capacity is exceeded, free hemoglobin circulates, and the heme is released and combines with albumin to form methemalbumin. Heme in plasma is detected by a positive Schumm's test. When free hemoglobin exceeds 25 mg/dL plasma, some is excreted in the urine, but in most subjects hemoglobinuria occurs when the total plasma level exceeds 150 mg/dL. The renal lesions that may occur consist of tubular necrosis and precipitation of hemoglobin within the tubules. Red cell stromal lipid is liberated, and this may initiate a disseminated intravascular coagulation. But *DIC* is more likely initiated by antibody-antigen complexes activating factor XII and complement, leading to activation of the coagulation cascade. The kallikrein-bradykinin system may be activated and affect the circulatory system. Minor incompatibilities may occur, causing hemolysis within the reticuloendothelial system manifested by fever, a mild decrease in hemoglobin, and an increase in bilirubin. If the recipient has a low antibody titer at the time of transfusion, reaction may be delayed for several days.

Clinical Manifestations

There is an increased hazard in patients who have had a previous transfusion reaction. If the patient is awake, the most common symptoms are the sensation of heat and pain along the vein into which the blood is being transfused, flushing of the face, pain in the lumbar region, and constricting pain in the chest. The patient may experience chills, fever, respiratory distress, hypotension, and tachycardia from amounts as small as 50 mL. In patients who are anesthetized and undergoing operation, the two signs that may draw attention are abnormal bleeding and continued hypotension despite adequate replacement. The mortality and morbidity resulting from hemolytic reactions is high if the patient receives a full unit of incompatible blood. Acute hemorrhagic diatheses occur in 8 to 30 percent of patients. There is a sudden fall in the platelet count, an increase in fibrinolytic activity, and consumption of coagulation factors, especially V and VIII, due to disseminated intravascular clotting.

Rudowski reported the following incidences of clinical manifestations in a large series with hemolytic posttransfusion reactions: oliguria, 58 percent; hemoglobinuria, 56 percent; arterial hypotension, 50 percent; jaundice, 40 percent; nausea and vomiting, 30 percent; flank pain, 25 percent; cyanosis and hypothermia, 22 percent; dyspnea, 20 percent; chills, 18 percent; diffuse bleeding, 16 percent; neurologic signs, 10 percent; and allergic reaction, 6 percent. The laboratory criteria are hemoglobinuria with a concentration of free hemoglobin over 5 mg/dL, a serum haptoglobin level below 50 mg/dL, and serologic criteria to show antigen incompatibility of the donor and recipient blood. The simplest clinical diagnostic test is insertion of a bladder catheter and evaluation of the color and volume of the excreted urine, since hemoglobinuria and oliguria are the most characteristic signs. A positive Coombs' test indicating transfused cells coated with patient antibody also provides evidence.

Treatment

If a transfusion reaction is suspected, the transfusion should be stopped immediately, and a sample of the recipient's blood should be drawn and sent along with the suspected unit to the blood bank for comparison with the pretransfusion samples. The serum bilirubin level should be determined in the recipient. Each gram of hemoglobin is converted to about 40 mg of bilirubin. The hemolytic reaction is characterized by an increase in the indirect reacting fraction.

A Foley catheter should be inserted and the hourly urine output recorded. Since renal toxicity is affected by the rate of urinary excretion and the pH, and since alkalizing the urine prevents precipitation of hemoglobin within the tubules, attempts are made to initiate diuresis and to alkalize the urine. This can be accomplished with mannitol or furosemide plus 45 mEq bicarbonate. If marked oliguria or anuria occurs, the fluid intake and potassium intake are restricted, and the patient is treated as a case of renal shutdown. In some instances, dialysis is required. After recovery from oliguria or anuria, diuresis is often copious and may be associated with significant losses of potassium and sodium, which require replacement.

Febrile and Allergic Reactions

These are relatively frequent, occurring in about 1 percent of transfusions. Reactions usually are mild and are manifested by urticaria and fever occurring within 60 to 90 min of the start of transfusion. In rare instances the reaction is severe enough to cause anaphylactic shock. Allergic reactions are caused by transfusion of antibodies from hypersensitive donors or the transfusion of antigens to which the recipient is hypersensitive. Reactions may occur after the administration of whole blood, packed red cells, plasma, and antihemophilic factor. Treatment consists of administration of antihistamines, epinephrine, and steroids, depending on the severity of the reaction. Repeated reactions can be prevented by the use of leukocyte-depleted or washed red cells.

Bacterial Sepsis

Bacterial contamination of infused blood is rare and may be acquired either from the contents of the container or the skin of the donor. Gram-negative organisms, especially coliform and *Pseudomonas* species, which are capable of growth at 4°C, are the most common cause. Clinical manifestations include fever, chills, abdominal cramps, vomiting, and diarrhea. There may be hemorrhagic manifestations and increased bleeding if the patient is undergoing surgical treatment. In some instances bacterial toxins can produce profound shock. If the diagnosis is suspected the transfusion should be discontinued and the blood cultured. Emergency treatment includes administration of adrenergic blocking agents, oxygen, antibiotics, and, in some cases, judicious transfusion.

Embolism

Although air embolism has been reported as a complication of intravenous transfusion, healthy animals tolerate large amounts of air injected intravenously at a rapid rate. It has been suggested that the normal adult generally can tolerate an embolism of 200 mL of air. Smaller amounts, however, can cause alarming signs and may be fatal. Manifestations of venous air embolism include a rise in venous pressure, cyanosis, a “mill wheel” murmur heard over the precordium, hypotension, tachycardia, and syncope. Death usually is related to primary respiratory failure. Treatment consists of placing the patient on the left side in a head-down position with the feet up. Arterial air embolism is manifested by dizziness and fainting, loss of consciousness, and convulsions. Air may be visible in the retinal arteries, and bubbles of air may flow from transected vessels.

Plastic tubes used for transfusion have also embolized after they have broken off within the vein. Plastic tubes have passed into the right atrium and the pulmonary artery, resulting in death. Embolized catheters have been removed successfully.

Thrombophlebitis

Prolonged infusions into peripheral veins using either needles, cannulae, or plastic tubes are associated with superficial venous thrombosis. Intravenous infusions that last more than 8 h are more likely to be followed by thrombophlebitis. There is an increased incidence in the lower limb as compared to upper limb infusions. Treatment consists of discontinuation of the infusion and local compression. Embolism from superficial thrombophlebitis of this nature is rare.

Overtransfusion and Pulmonary Edema

Overloading the circulation is an avoidable complication. It may occur with rapid infusion of blood, plasma expanders, and other fluids, particularly in patients with heart disease. In order to prevent this complication, the central venous pressure should be monitored in these patients and whenever large amounts of fluid are administered.

Circulatory overloading is manifested by a rise in the venous pressure, dyspnea, and cough. Rales generally can be heard at the base of the lungs. Treatment consists of stopping the infusion, placing the patient in a sitting position, and, occasionally, venous section for removal of blood.

Although acute pulmonary edema occurs more frequently after large transfusions, it has been reported in patients receiving small transfusions. A syndrome that can be confused with pulmonary edema consists of postoperative hypoxia, seen in patients who have undergone cardiac surgical treatment and extracorporeal bypass procedures. A damaging factor apparently is carried by the perfusing blood, and immature plasma cells are found in the interalveolar tissue. The illness represents an immune response to blood. The incidence is reduced by employing the hemodilution technique of pump priming.

Transmission of Disease

Malaria, Chagas' disease, brucellosis, and syphilis are among the diseases that can be transmitted by blood transfusion. Syphilis has been reported after the transfusion of platelets. The storage temperature used for all other blood components (4°C or lower) kills the spirochete. The incubation period ranges from 4 weeks to 4 months. The first manifestation is the skin rash of secondary syphilis. Cure is readily achieved with brief penicillin therapy. Malaria can be transmitted by all blood components, including platelets, fresh frozen plasma, and frozen or deglycerolized red cells. The species most commonly implicated is *Plasmodium malariae*. The incubation period ranges from 8 to 100 days; the initial clinical manifestation is shaking chill and spiking fever. Cytomegalovirus (CMV) infection, causing a syndrome resembling infectious mononucleosis, was commonly observed after open-heart surgery when large amounts of heparinized blood were used to prime the pump. The most significant morbidity and mortality occurs after transfusion of CMV-infected blood in low-birth-weight infants born of mothers who were CMV antibody–negative.

Posttransfusion viral hepatitis remains the most common fatal complication of blood transfusion. It is estimated that for every case of icteric posttransfusion viral hepatitis there are four anicteric cases, many of which are asymptomatic. Hepatitis is caused either by hepatitis B virus, or the non-A, non-B viruses, including C. The incubation period of the former is up to 6 months, the latter's may be as short as 2 weeks. Serologic markers for hepatitis B surface antigen (HB sAg) and hepatitis C are available, and collecting agencies are required to test all units of blood for these antigens. The risk of hepatitis transmission per unit of blood is 0.035 percent.

The clinical manifestations of hepatitis include lethargy and anorexia as part of anicteric disease, icterus, and chronic liver disease. HB sAg persists in about 35 percent of patients who develop serum hepatitis of type B. There is no risk from human serum albumin and other plasma protein fractions.

Immune globulin is effective in preventing type A hepatitis but is inconsistent in preventing type B hepatitis. Accidental self-inoculation with material that is definitely known to contain HBsAg, or transfusion of blood that is HB sAg-positive, constitutes an indication for immediate use of immunoglobulin (human) anti-HB sAg. The recommended dose is 0.02 to 0.06 mL/kg BW of IgG given as an intramuscular injection. A vaccine has been developed against HB sAg, and it is recommended that all surgeons undergo vaccination.

The incidence of AIDS following blood transfusion has been estimated to be one case per 225,000 patients transfused, and blood collecting agencies have taken measures to preclude donors in high-risk groups and to apply screening techniques.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 4 - Shock

Annabel Barber
G. Tom Shires III
G. Tom Shires

DEFINITION

Shock is a pathophysiologic condition clinically recognized as a state of inadequate tissue perfusion. A constant internal environment was first suggested by the French physiologist Claude Bernard. In the mid–nineteenth century Bernard proposed that higher animals lived in two very different environments—a *milieu int érieur*, in which the tissue elements live, and a *milieu ext érieur*, in which the body resides. He stated that “the stability of the *milieu int érieur* is the primary condition for freedom and independence of existence: the mechanism which allows this is that which insures in the *milieu int érieur* the maintenance of all the conditions necessary to the life of the elements.” Bernard went on to declare that “the circulation of the blood forms a true organic environment, intermediary between the external environment in which the individual as a whole lives and the molecules of the living cells which would otherwise not come into direct relationship with that external environment.” The notion that the constancy of the internal environment is protected by multiple intrinsic mechanisms, including renal, pulmonary, hepatic, and cell- membrane function, evolved over the next fifty years. Walter Cannon coined the term “homeostasis,” which led to the concept that an organism's fitness for survival is directly related to its capacity to maintain homeostasis. From this evolved the biologic precept that the extracellular fluid, including circulation, is the true milieu of life because it enables the cells of the body to function.

In the first part of this century a variety of theories addressed the cause of vascular collapse in injured patients. It was assumed that this vascular collapse was caused primarily by toxins. In a series of innovative experiments beginning with Blalock, researchers determined that almost all acute injuries are associated with changes in fluid and electrolyte metabolism. These studies showed that the alterations were primarily the result of reductions in the effective circulating blood volume and that this reduction may be the result of loss of blood as in hemorrhage, but also as a result of loss of vascular tone (e.g., in septic or neurogenic shock), pump failure (cardiac tamponade) or myocardial infarction, or loss of large volumes of extracellular fluid, which occurs in patients with diarrhea, vomiting, or fistula drainage. Blalock's studies showed that fluid loss in injured tissues was loss of extracellular fluid that was unavailable to the intravascular space for maintenance of circulation. The original concept of a “third space” in which fluid would be sequestered and thus unavailable to the intravascular space, evolved from those studies.

During World War II plasma became a favored resuscitative solution in addition to whole-blood replacement. However, the principle that a limited amount of salt and water should be given to the patient after surgical or other injury prevailed through the Korean War, largely because of the work of Collier and Moyer in experiments done at the University of Michigan. By the time of the Vietnam War, the provision of volume resuscitation in excess of replacement of shed blood became standard to maintain adequate homeostasis. During World War II acute tubular necrosis had been seen commonly after hypovolemic shock, but with the liberal use of fluid resuscitation during the Vietnam conflict, the incidence of acute tubular necrosis dramatically decreased (Table 4-1).

The etiologic classification offered by Blalock in 1934 remains a useful outline for a modern definition. Blalock suggested four categories: hematogenic, neurogenic, vasogenic, and cardiogenic. It now is clear that shock is a systemic disorder that disrupts vital organ function as the eventual result of a variety of causes. Whereas hemorrhagic or traumatic shock is characterized by global hypoperfusion, septic shock may be associated with hyperdynamic circulation resulting in a maldistribution of regional or intraorgan blood flow. Consequently, Cerra's description of shock as a “disordered response of organisms to an inappropriate balance of substrate supply and demand at a cellular level” may more accurately reflect the unifying functional abnormality at the metabolic level.

CIRCULATORY HOMEOSTASIS

Preload

The majority of the blood volume at rest is contained within the venous system. The effect of the return of this venous blood to the heart produces ventricular end-diastolic wall tension, a major determinant of cardiac output. Gravitational shifts in blood volume distribution are rapidly compensated for by active and passive alterations in venous capacity. The thin-walled systemic veins are highly compliant. As arteriolar inflow increases, the venous pressure rises and venous capacitance passively increases. With decreased arteriolar inflow, active contraction of the venous smooth muscle cells and passive elastic recoil combine to increase return of blood flow to the heart, maintaining adequate ventricular filling and supporting cardiac output.

In the normal heart, most changes in cardiac output are a reflection of alterations in preload. Changes in position, intrathoracic pressure, intrapericardial pressure, and circulating blood volume produce major changes in cardiac output. Different venous beds play different roles in regulating preload. Veins in the skeletal muscles show a minor response to sympathetic stimulation and respond more to external factors, predominantly the balance between gravitational forces and the muscle pump. Increases in sympathetic outflow to the splanchnic vascular bed produce a rapid and dramatic reduction in the splanchnic blood volume that normally contains about 20 percent of the total blood volume. Exercise and the response of central baroreceptors during hemorrhage reflexively decrease the splanchnic capacitance after these stimuli of sympathetic outflow. Cutaneous noradrenergic nerves respond to hypothalamic control and alter the venous tone of the skin to promote thermal regulation during resting heat stress, exercise demands, and the febrile response.

The normal circulating blood volume is maintained within narrow limits by balancing salt and water intake with external losses by the kidney's ability to respond to alterations in hemodynamics and the hormonal effects of renin, angiotensin, and antidiuretic hormone. The summation of these relatively slow responses, which maintain adequate preload by altering the circulating blood volume, are overshadowed in the acute setting by the changes in the venous tone, systemic vascular resistance, and intrathoracic pressure. In addition, the net effect of preload on the ventricle also responds to the cardiac determinants of ventricular function, including coordinated atrial contraction, which augments ventricular diastolic filling, and tachycardia, which drops the effect of preload on the ventricle by compromising diastolic filling time.

Ventricular Contraction

The Frank-Starling curve describes the varying force of ventricular contraction as a function of its preload. The changes in force development are explained by the ultrastructural property of the myocardium, which generates a force of contraction dependent on initial muscle length. A variety of disease states, including myocardial injury, valve dysfunction, and cardiac hypertrophy, may alter the mechanical performance of the heart. Experimental studies in burn, septic, hemorrhagic, and traumatic shock have documented deteriorating intrinsic cardiac function during these injury states. While the mechanisms of these alterations in myocardial performance are unclear, their effect on the evaluation and management of global perfusion in clinical shock may be assessed by Swan-Ganz catheterization that measures preload indirectly as end-diastolic pressure, thermodilution cardiac output, and estimations of calculated vascular resistance.

Afterload

Afterload is the force acting to resist myocardial work during contraction. Arterial pressure is the major component of afterload that influences the ejection fraction. This vascular resistance is primarily determined by precapillary smooth muscle sphincters in conjunction with other rheologic factors such as blood viscosity. If afterload increases, stroke volume can be maintained in the presence of an increase in preload. Unlike in normal heart function, in which stroke volume can be maintained in the face of increased vascular resistance by increasing preload, the decreased effective circulating volume in shock states prevents this compensatory maintenance of cardiac output. This imbalance of preload-after-load effects overwhelms the normal increase in inotropic state produced by increased sympathetic nerve activity in the heart and by increased circulating catecholamines released by the stress response.

PATHOPHYSIOLOGY OF HYPOVOLEMIC SHOCK

Hypovolemic shock results from a decrease in the circulating or effective intravascular volume. Consequently, most of the signs of clinical shock are characteristic of peripheral hypoperfusion and increased adrenergic activity. Young, healthy patients in shock initially appear anxious and exhibit restlessness. This behavior gives way to apathy and lethargy after initiation of treatment. Frank coma rarely results from blood loss alone; usually it is a sign of concomitant direct brain injury, or it is coincident with complete cardiovascular collapse.

As intravascular volume is lost, an increase in peripheral vascular resistance occurs to defend the blood pressure in compensation for falling cardiac output. Differential increases in peripheral resistance in regional arteriolar beds, particularly in the skin, gut, and kidney, further defends pressure at the cost of further decreasing organ flow. The pale, cool skin noted on examination and the blanching of the bowel with decreased pulses in the mesentery are gross signs seen at the bedside and at laparotomy.

A decrease in circulating blood volume also results in tachycardia in response to decreased stroke volume from inadequate preload. The tachycardic response depends on the rate of blood loss and the position of the patient; orthostatic testing may unmask cardiovascular instability with tachycardia and hypotension in a patient who appears stable when examined in the supine position. Significant orthostasis reflects a 30 percent reduction in circulating blood volume in young patients.

Compensatory Responses

The following compensatory responses occur during shock:

- (1) Loss of circulating intravascular volume results in increased vascular tone, which elevates peripheral vascular resistance, resulting in a redistribution of blood flow among the organ systems of the body. Blood flow to those organs that are “autoregulated, ” such as the heart and the brain, is maintained at the expense of cutaneous, splanchnic, and renal circulatory beds, which depend on sympathetic tone for blood flow.
- (2) Decreases in intravascular volume stimulate increased sympathetic activity, which diminishes vagal inhibition of the rate and the force of cardiac contraction. Greater myocardial contractility and enhanced venous return, which help to improve stroke volume, accompany the tachycardia present during shock. Cardiac output is increased by these responses, as is myocardial oxygen consumption. Blood pressure is maintained by increases in total peripheral resistance and cardiac output.
- (3) Loss of circulating intravascular volume leads to decreased capillary hydrostatic pressure. Transcapillary influx of extravascular extracellular fluid then occurs from the interstitial space as a result of this alteration in Starling forces. This mobilization of the interstitial fluid pool into the intravascular space has two major effects: circulating intravascular volume is increased, and blood viscosity is decreased secondary to dilution.
- (4) In addition to the increased systemic oxygen-carrying capacity produced by hemodilution, tissue extraction of oxygen is enhanced in hemorrhagic shock by the presence of acidosis and elevated levels of erythrocyte 2,3-diphosphoglycerate (2,3-DPG). Decreased delivery of cellular substrates leads to increased anaerobic metabolism of glucose and accumulation of lactic acid. The resultant tissue acidosis produces a rightward shift in the oxyhemoglobin dissociation curve, decreasing the affinity of hemoglobin for oxygen,

thereby making more oxygen available to the tissues. Hypoxia also stimulates respiratory centers, leading to hyperventilation and respiratory alkalosis and a subsequent increased rate of erythrocyte synthesis of 2,3-DPG. This produces a further, more prolonged rightward shift of the oxyhemoglobin dissociation curve (Fig. 4-1).

(5) Arteriolar constriction and loss of circulating volume diminish renal blood flow. Both afferent and efferent arterioles are then stimulated, with resultant corticomedullary shunting of the remaining flow in an attempt to maintain an effective glomerular filtration rate. Urine output subsequently decreases as water and sodium are retained. Clearance of urea and acids, as well as buffering capacity, are diminished, leading to a loss of control of acid-base balance.

(6) Changes in blood volume in association with afferent sensory impulses lead to marked release of epinephrine and norepinephrine early in the course of hemorrhagic shock. The increased secretion of these catecholamines by the adrenal glands is an acute, short-lived response, usually limited to the day of injury unless complications occur. Epinephrine and norepinephrine produce vasoconstriction and tachycardia, resulting in increased cardiac output and blood pressure. Glycogenolysis, lipolysis, and skeletal muscle breakdown are stimulated, and insulin release is inhibited, promoting glucose mobilization, protein catabolism, and negative nitrogen balance. Epinephrine incites insulin resistance in skeletal muscle, and perhaps in other tissues, thus favoring glucose utilization by insulin-independent tissues such as the heart and the brain. The acute catecholamine response also results in retention of sodium and water in the proximal tubule of the nephron.

(7) Pituitary *adrenocorticotrophic hormone (ACTH)* release is stimulated in hemorrhagic shock by decreased blood volume, decreased arterial pressure, pain, hypoxemia, and hypothermia. After severe hemorrhage, circulating cortisol provides no feedback inhibition on ACTH release, but feedback is restored once blood volume has been reexpanded. Cortisol potentiates the actions of epinephrine and glucagon on glucose metabolism and insulin resistance and further stimulates mobilization of amino acids from skeletal muscle. Increased cortisol secretion also results in renal sodium and water retention.

(8) Unlike the counterregulatory hormones, insulin secretion is diminished in shock. This relative hypoinsulinemia augments the mobilization of glucose, amino acids, and fat stores stimulated by epinephrine, glucagon, cortisol, and growth hormone.

(9) Arginine vasopressin (AVP), also known as antidiuretic hormone (ADH), is secreted in response to increased serum osmolarity and hypovolemia (Fig. 4-2). It appears, however, that hypovolemia is the more potent stimulus, producing a picture similar to the syndrome of inappropriate antidiuretic hormone secretion in response to hemorrhage and shock. Arginine vasopressin increases water permeability and passive sodium transport in the distal tubule of the nephron, allowing increased water resorption. Arginine vasopressin is also a potent splanchnic vasoconstrictor. Activation of the renin-angiotensin system occurs in shock in response to increased sympathetic stimulation of the juxtaglomerular cells via a beta-adrenergic mechanism, decreased renal perfusion pressure, and compositional changes in tubular fluid. Renin, released from the juxtaglomerular apparatus, results in increased production of angiotensin I, which is rapidly converted to angiotensin II in the lung. Angiotensin II is a powerful arterial and arteriolar vasoconstrictor and stimulates renal prostaglandin production as well as the release of aldosterone and *ACTH*. Increased aldosterone secretion during shock occurs in response to increased circulating levels of angiotensin II and ACTH. Aldosterone increases sodium resorption in the distal nephron in exchange for potassium and hydrogen ions and represents the principal mechanism by which the kidney may excrete the accumulated by-products of anaerobic metabolism and cellular damage. Prostaglandins, particularly PGE₂, and kallikreins, produced in the kidney, function locally to dilate renal vessels and increase renal blood flow. Thromboxane A₂ results in splanchnic and cutaneous vasoconstriction and may promote cardiovascular dysfunction. Platelet-activating factor, produced by stimulated macrophages, results in coronary vasoconstriction and cardiac depression and increases platelet aggregation. The leukotrienes, produced by activated mast cells, also are potent vasoconstrictors and appear to promote muscle catabolism and amino acid release.

(10) The response of the extracellular fluid to acute hemorrhagic shock has been demonstrated experimentally in animal models. Using different isotopes to measure simultaneously the total-body red blood cell mass, plasma volume, and interstitial fluid volume, researchers can determine the distribution of these components of the extracellular fluid after shock. A 10 percent blood loss in splenectomized dogs produces no evidence of clinical shock. Red blood cell and plasma loss in this model were determined to be equal to the volume of shed blood, with no evidence of additional extracellular fluid loss. However, a 25 percent blood loss in the same model results in hypotension and an 18 to 26 percent reduction of functional extracellular fluid volume in addition to the measured losses of red blood cells and plasma. Further losses of extracellular fluid volume in addition to red blood cells and plasma can be demonstrated when the magnitude of hemorrhage is increased to 35, 45, or even 50 percent of circulating blood volume. With less severe shock models, there remains a reduction in the early equilibrating extracellular fluid available for intravascular influx, while total anatomic extracellular fluid may be normal. As there is no measured external loss of this functional extracellular fluid volume in any of these models, it is presumed that these changes represent an internal redistribution of the extracellular fluid in response to hemorrhagic shock (Fig. 4-3). It has been demonstrated that blood reinfusion after hemorrhagic shock restores the measured deficit in red blood cell mass and plasma volume, but not the deficit in extracellular fluid volume. However, the addition of a balanced salt solution or extracellular fluid "mimic," such as lactated Ringer's solution, to the shed blood infusion results in return of extracellular fluid volumes to control levels. Mortality in a model of "irreversible" shock was reduced from 80 percent in animals given only blood to 30 percent by restoration of the functional extracellular fluid volume with balanced salt solutions in addition to return of shed blood. This loss of functional extracellular fluid volume during shock is partially explained by the transcapillary influx of interstitial fluid into the intravascular space in response to decreased capillary hydrostatic pressures. However, the magnitude of intravascular refilling is inadequate to explain the total reduction of extracellular fluid observed. The isotonic movement of interstitial water and sodium into the cellular mass represents the most likely mechanism for the additional reduction of extracellular fluid volume, and isotonic movement of water and sodium into muscle cells has been demonstrated during hemorrhagic shock (Fig. 4-4).

(11) A semipermeable cell membrane functions through active transport mechanisms to maintain the ionic differences between intracellular and extracellular fluid. A negative cellular membrane potential difference ensues and is important for cell homeostasis and control of volumes and concentrations in the fluid compartments. This transcellular membrane potential difference may be measured serially in vivo using Ling-Gerard ultramicroelectrodes. In muscle and liver tissue this membrane potential difference serves as a reliable indicator of cellular dysfunction during hemorrhagic shock. During profound acute hemorrhagic shock, the transcellular membrane potential difference in skeletal muscle falls from -90 to -60 mV. This fall in potential difference is specific to the state of shock and independent of acid-base status. Alterations in membrane potential difference in primate models of acute hemorrhagic shock are reversible, with recovery of the normal difference after adequate fluid resuscitation, including replacement of losses from the extracellular fluid compartment (Fig. 4-5). Muscle biopsies obtained concomitantly with measurements of transcellular membrane potential allow measurements of intracellular water and electrolytes. These studies demonstrate the correlation of altered membrane potential difference and cellular swelling, marked by increased intracellular water, an influx of extracellular sodium and chloride, and an efflux of intracellular potassium. There appears to be little change in intracellular sodium activity, indicating that the extra sodium that diffuses into the cell during membrane dysfunction is bound to fixed charges or compartmentalized into the organelles. Decreased production of ATP in association with anaerobic metabolism results in a diminished ability to control sodium flux and has been proposed as a cause of cellular dysfunction during hemorrhagic shock. However, levels of high-energy phosphate compounds are maintained in liver and skeletal muscle early in hemorrhagic shock at a time when alterations in membrane potential difference have already occurred. Cellular membrane dysfunction is not prevented by administration of such high-energy phosphates as ATP-MgCl₂. The fall in concentration of intracellular ATP appears to be the effect of cellular dysfunction and not its primary cause. Other possible explanations for cellular dysfunction during hemorrhagic shock are decreased activity of the sodium-potassium pump responsible for maintenance of membrane potential difference independent of ATP content, or changes in membrane permeability. These effects might occur through the actions of inflammatory mediators generated in response to injury and hemorrhage. Tumor necrosis factor-alpha (TNF-α), a cytokine produced by activated macrophages, is implicated in the mediation of alterations in skeletal muscle membrane dysfunction and the hemodynamic consequences of sepsis and may play a similar role in hemorrhagic shock. Other factors, such as platelet-activating factor, leukotrienes, thromboxane A₂, and complement activation, have also been implicated.

(12) Nitric oxide is a potent regulator of basal blood vessel tone. It is a ubiquitous free radical produced as a result of two forms —constitutive and inducible —of nitric oxide synthase. The inducible form (iNOS) is up-regulated by endotoxin and proinflammatory cytokines, such as interleukin-1 (IL-1) and *tumor necrosis factor (TNF)* in macrophages, Kupffer cells, vascular smooth muscle, and endothelium. It is likely that nitric oxide plays vital regulatory roles in the cardiovascular, pulmonary, gastrointestinal, immune, and central nervous systems.

Pulmonary Derangements in Shock

Accompanying successful fluid resuscitation is the emergence of pulmonary dysfunction in 1 to 2 percent of the survivors of shock. This occurs in some patients without lung injury per se. Originally referred to as shock lung, adult respiratory distress syndrome was first described in 1967. This disorder, now referred to as *acute respiratory distress syndrome (ARDS)*, is characterized by hypoxia (despite oxygen therapy), decreased pulmonary compliance, diffuse or patchy infiltrates on chest x-ray, and noncardiac pulmonary edema.

Etiology

The pulmonary system exhibits a stereotypic response to insult. That is, a variety of injuries can trigger a final common pathway, resulting in the symptom complex known as *ARDS*. These include direct pulmonary injury, such as that seen in aspiration, inhalation injury, pulmonary contusion, and near drowning, and seemingly unrelated disorders, such as multiple transfusions and trauma such as fractures. Common to all these disorders is the initiation of inflammatory mediators. These result in increases in microvascular permeability and subsequent proteinaceous fluid deposition in the alveolar epithelial and pulmonary capillary endothelial interface. Resulting from this disruption are abnormal ventilation and perfusion relationships and subsequent hypoxia (Fig. 4-6).

The phenomenon of noncardiac edema occurs as a result of derangements of lung microvascular permeability. Under normal conditions, a small amount of fluid moves out of the pulmonary capillaries and is cleared from the interstitium by pulmonary lymphatics (Fig. 4-7). Flux is governed by Starling forces described by the equation:

$$Q = K_f(P_{mv} - P_t) - s(P_{mv} - P_t)$$

where: Q = transcapillary exchange

K_f = filtration coefficient of water

P_{mv} = capillary hydrostatic pressure

P_t = tissue interstitial pressure

s = osmotic reflection coefficient

P_{mv} = capillary colloid osmotic pressure

P_t = tissue interstitial colloid osmotic pressure

In the pathologic state, the alveolar–capillary interface is disrupted and the resultant fluid overwhelms pulmonary lymphatic clearance. These abnormalities have been documented to occur prior to, or without, associated chest-x-ray abnormalities. Diuretics and fluid restriction have no impact on this pathophysiology and are not useful. Likewise, colloid administration has not been shown to effectively decrease extravascular lung water, because the normal barrier is disrupted and is permeable to large molecules such as albumin.

Pulmonary failure associated with sepsis has been studied extensively, and it is likely that similar mechanisms are operative in other conditions resulting in ARDS. Endotoxin or lipopolysaccharides (LPS) have a direct effect on pulmonary endothelial cells, increasing cellular permeability. This has been observed in in-vitro and in-vivo models. Other active mediators include complement, eicosanoids, platelet-activating factor, leukotrienes, and thromboxane A₂.

Diagnosis

The diagnosis of ARDS begins with clinical suspicion and is based on documentation of hypoxia, abnormal chest x-ray, and a measured decreased lung compliance. Multiple scales for grading the severity of lung injury are in use, most of which include percentage of abnormal lung seen on x-ray and the amount of *positive end-expiratory pressure (PEEP)* required to maintain adequate oxygenation. These scales have been criticized because of the subjective nature of grading of x-rays and practice variations in the use of PEEP.

Therapy for ARDS

The therapeutic goal is to maintain tissue oxygenation. Supplemental oxygen is supplied to maintain Pa_{o2} of 65 mmHg or more. Hemoglobin concentration should be maintained at 12 g/dL or higher, with buffering of pH to allow optimal oxygen transport. A pulmonary artery catheter is desirable to monitor central volumes and mixed venous saturations. Fluid overload or underresuscitation can adversely effect the patient with ARDS. Standard pulmonary management includes the use of a volume ventilator in the mandatory mode with tidal volume and rate set to allow adequate carbon dioxide exchange. This usually can be accomplished with rates of 10 to 12 breaths per minute and tidal volumes of 10 to 12 mL/kg dry weight. PEEP is initiated at 5 cmH₂O to approximate glottic pressure. Sedation and paralysis may be necessary. PEEP is used in order to maintain oxygenation at nontoxic (50 percent or less) levels of oxygen. In managing the patient with ARDS, the ventilator is set at 100% oxygen and the optimal level of PEEP is identified. This level is found by increasing PEEP by increments of 2.5 cmH₂O, allowing at least 30 min for equilibration and measuring arterial and mixed venous blood gases, pulmonary capillary wedge pressure, and cardiac output. PEEP is increased to as much as 20 cmH₂O, and optimal settings (highest oxygenation without compromise of cardiac output) are identified. PEEP is then set at this level and oxygen decreased incrementally to maintain a Pa_{o2} of 65 mmHg, with a goal of 50% inspired oxygen or less. PEEP can then be decreased, if oxygenation is maintained, by increments of 2.5 cmH₂O every 12 h.

The potential risks of PEEP are exacerbated by hypovolemia. Volume loading to assure adequate filling pressures before PEEP is applied is beneficial in interpreting changes in wedge pressure and cardiac output that may occur after use of PEEP. Increased intrathoracic pressure and decreased venous return can cause depression of cardiac output. Lowering of PEEP is necessary if cardiac output becomes compromised. Wedge pressure measurements may be corrected in the patient with noncompliant lungs by subtracting one-fourth of the applied PEEP from the measured wedge pressure.

Pneumothorax can occur at high pressures (>20 cmH₂O) and can be catastrophic. Peak airway pressures should be carefully monitored. Patients with chronic obstructive lung disease characterized by preexisting increases in functional residual capacity may not benefit from PEEP therapy.

Trials of early application of PEEP in patients at high risk for ARDS failed to show any benefit in overall mortality or complications. The course of ARDS has been relatively unaffected in trials of anti-inflammatory drugs such as ibuprofen and sepsis trials using anticytokine therapy (interleukin-1 receptor antagonist, monoclonal antibodies to TNF).

THERAPY FOR SHOCK

Hypovolemic Shock

Initial care of the injured patient should follow the guidelines from the advanced trauma life support procedures of the American College of Surgeons Committee on Trauma. In a patient who has undergone trauma, more than one causative factor may be operating. Once the diagnosis of shock has been made and supportive therapy begun, a diligent search can be made for the causative factor or factors. Deficits of total body water and electrolytes usually is subtle, and correction requires specific therapy with crystalloid solutions. Reductions in the extracellular fluid volume (plasma and interstitial fluids) as a result of burns, peritonitis, and some forms of crush injury are more easily recognized. Specific therapy should be started with electrolyte solutions; occasionally plasma or some source of protein is required as well. External blood loss should be corrected immediately with appropriate fluid therapy.

Fluid Resuscitation Composition of Resuscitation Fluids

Lactated Ringer's Solution

Lactated Ringer's solution is the most widely available and most frequently used balanced salt solution (crystalloid) for fluid resuscitation during shock. It is safe and inexpensive, and it equilibrates rapidly throughout the extracellular compartment, restoring the extracellular fluid deficit associated with blood loss. Concern that the lactate content of Ringer's solution might aggravate the lactic acidosis coexisting with hemorrhagic shock is unwarranted. Studies in animal models and human beings demonstrate that the use of lactated Ringer's solution in addition to blood replacement results in more rapid return of serum lactate and pH to normal levels than does replacement with blood alone. Lactate is rapidly converted to bicarbonate in the liver.

Because of the rapid equilibration of balanced salt solutions into the extracellular space, larger volumes may be required for adequate resuscitation, resulting in decreased intravascular oncotic pressure. Although there has been no documented increase in morbidity or mortality secondary to the appropriate use of balanced salt solutions, it has been speculated that loss of large volumes of balanced salt solution into the interstitial space during resuscitation may subsequently contribute to postresuscitation organ dysfunction, particularly pulmonary edema and respiratory failure. These potentially deleterious effects of balanced salt solution appear to be offset by the active lymphatic circulation, which buffers against fluid overload and helps to maintain normal oncotic gradients between the intravascular and interstitial spaces.

Colloid Solutions

The use of colloidal substances that tend to remain intravascular continues to be advocated by some, in lieu of balanced salt solution, for resuscitation. Administration of fluid preparations containing colloidal substances such as albumin raise the intravascular colloidal pressure, leading to intravascular influx of interstitial fluid. Because colloids remain briefly in the intravascular space, a lower total volume of resuscitative fluid is required to attain hemodynamic stability than when crystalloid solutions are used; it has been theorized this serves to prevent postresuscitation fluid overload. However, colloid solutions are more expensive, may bind and decrease the ionized fraction of serum calcium, decrease circulating levels of immunoglobulins, decrease the immune reaction to tetanus toxoid, and decrease endogenous production of albumin. More important, use of colloid- containing solutions as resuscitative therapy during hemorrhagic shock further compromises the extracellular fluid volume deficit rather than restoring it. This was confirmed in a study by Greenhalgh and colleagues in which serum albumin levels were maintained by exogenous administration of albumin in pediatric burn patients. No differences were demonstrated in terms of resuscitation, maintenance fluid requirements, subsequent complications, length of stay, or mortality.

Numerous experimental and clinical studies have examined the issue of the superiority of crystalloid over colloid resuscitation. Moss and associates demonstrated that adequate resuscitation of primates from potentially lethal hemorrhage with balanced salt solution or balanced salt solution plus 5% albumin resulted in restoration of circulation parameters and perfusion to normal. Although the volume of balanced salt solution required was three times greater than that of the colloid solution, no measurable differences were observed between the two groups in pulmonary compliance or postresuscitation lung water content.

In another study primates were subjected to plasmapheresis sufficient to decrease the serum oncotic pressure significantly, similar to the situation occurring after crystalloid resuscitation. Zarins and coworkers demonstrated that extravascular lung water, pulmonary compliance, oxygenation, and shunt fraction did not change. This was true when pulmonary capillary wedge pressure was not elevated. Holcroft and colleagues also reported that pulmonary edema after resuscitation for hemorrhagic shock with balanced salt solution could be produced only by sustained elevations of pulmonary artery pressure. Guyton and Lindsey found that lowering intravascular colloid oncotic pressure alone does not result in pulmonary edema in dogs, but does lower the level of left atrial pressure necessary to produce pulmonary dysfunction. In a model similar to that of Zarins, Demling and associates further demonstrated a marked increase in pulmonary lymph flow without development of pulmonary edema, apparently compensating for the altered oncotic pressure gradient.

In a clinical study, Horovitz and Shires reported an incidence of pulmonary dysfunction in 2.1 percent of 978 patients undergoing operative procedures after severe trauma, despite receiving large volumes of crystalloid. Carey reported no evidence of acute pulmonary edema in 56 injured patients in Vietnam, despite an average resuscitation volume of 12 L of balanced salt solution.

As it is well established that stabilization of hemodynamic parameters after hemorrhagic shock requires a greater volume of crystalloid than colloid solution, meaningful prospective clinical trials comparing crystalloid and colloid require resuscitation to equal end points and not to equal volumes. Using these guidelines, Virgilio and colleagues found no differences in pulmonary function or shunt fraction among 29 patients undergoing aortic surgery randomized to balanced salt solution or colloid solution resuscitation. In addition, Shires III and coworkers found no differences in extravascular lung water either immediately after operation or 24 to 48 h later in 19 aortic surgery patients randomized to balanced salt solution or colloid resuscitation, despite a markedly lower intravascular oncotic pressure in the balanced salt solution group (Fig. 4-8).

Lowe and associates found no differences in survival rates, incidence of pulmonary failure, or postoperative pulmonary dysfunction in stable and unstable trauma patients undergoing laparotomy who received balanced salt solution or colloid solution resuscitation. Lucas found that the use of colloid not only prolonged the resuscitation phase but also delayed postresuscitation diuresis, perhaps through failure of restoration of the interstitial volume deficit. Further, the incidence of postresuscitation hypertension was higher in patients given albumin, suggesting that renal function may best be protected during resuscitation from shock by rapid replacement of the intravascular and total extracellular fluid deficits. A meta-analysis of colloid versus crystalloid fluid resuscitation that included many of the studies cited here concluded that crystalloid is superior to colloid for resuscitation after trauma in human beings, with a 12 percent reduction in mortality after crystalloid infusion.

There is therefore no clinical evidence that appropriate resuscitation with balanced salt solution is associated with any harmful effects on pulmonary function when guided by hemodynamic parameters. No protective effect of colloid solutions on postresuscitation pulmonary function can be demonstrated, even though colloid solutions do produce transiently greater intravascular expansion per unit volume given than do crystalloid solutions. As the volume expansion with colloid occurs to some extent by further compromise of the extracellular fluid volume, renal function during shock may best be preserved by crystalloid resuscitation.

Hypertonic Saline

Clinical and experimental studies have demonstrated that a small volume of hypertonic saline can be an effective initial resuscitative solution. Hypertonic saline resuscitation results

in a lower water load than equivalent resuscitation with balanced salt solutions. However, patients resuscitated with hypertonic saline solution require close monitoring of electrolytes to prevent hyponatremia and hyperosmolar coma. Recent studies in patients have shown that while blood pressure may be elevated more rapidly in the first few minutes after shock and resuscitation, no changes in survival rates occurred. In view of the need for electrolyte monitoring and lack of the definition of the volumes appropriate for infusion, long-term benefits have not been established. In animal studies showing deterioration of cellular function there was a higher mortality 24 h after hypertonic saline/dextran resuscitation.

Hetastarch

Hydroxyethyl starch (hetastarch) is an artificial colloid derived from amylopectin, with colloidal properties similar to those of albumin. It is less expensive than albumin, and, because of its larger molecular weight and need for enzymatic degradation, it has a longer plasma half-life than albumin. As with any colloidal solution, hetastarch restores intravascular volume at the further expense of the already compromised interstitial space when used in resuscitation during shock. Resuscitation with hetastarch also may be difficult to control, because the slow equilibration of these large molecules can lead to rapid fluctuations in central venous pressure. Mild and transient coagulopathies have been noted in patients resuscitated with hetastarch, and a role in depression of the reticuloendothelial system has been postulated.

Dextran

Dextran, in 40 kD and 70 kD solutions, has also been used as a plasma expander. Although dextran has a shorter half-life than hetastarch, it also approximates the colloidal activity of albumin when given by intravenous infusion. Clinical studies have demonstrated no differences in rates of organ dysfunction or mortality when resuscitation with dextran is compared to that with balanced salt solutions. However, dextran use is associated with a greater risk of anaphylaxis than is hetastarch or albumin and has produced coagulation defects and immunoglobulin depression.

Blood Substitutes

Periodic shortages of blood products and the infectious risks associated with transfusions have fueled the search for an efficacious artificial blood substitute. Early studies using free hemoglobin obtained from outdated blood resulted in allergic reactions, renal failure, coagulopathies, and immune dysfunction secondary to reaction to retained erythrocyte stromal elements. The subsequent purification of stroma-free hemoglobin (SFH) has eliminated these side effects, but problems with the use of SFH for resuscitation remain. SFH has an abnormally high affinity for oxygen, has a short plasma half-life, and is available only from human sources. Although binding with other molecules or incorporation into liposomes may reduce oxygen affinity and improve plasma retention time, SFH will probably not become a practical substitute for blood until an artificial or animal donor source becomes available. Bovine SFH appears to be one useful approach for treating hemorrhage (Fig. 4-9).

Perfluorochemical compounds have uniquely enhanced abilities to dissolve gases, particularly oxygen and CO₂, but require emulsification to be water soluble. Perfluorodecalin (Fluorosol-DA) also contains electrolytes, bicarbonate, and starch to obtain osmotic and pH balance with plasma. The oxygen-carrying capacity of this emulsion is lower than that of hemoglobin, and thus its use requires higher inspired oxygen concentrations (Fig. 4-10). Fluorosol-DA has been extensively used in Japan and in clinical trials has proved an effective substitute for hemoglobin. Potential adverse effects include acute pulmonary edema, activation of complement and the coagulation cascade, acute respiratory failure, and depression of the reticuloendothelial system. It is expensive and requires special storage to prevent gelatinization.

Volumes Appropriate for Resuscitation

Models of “controlled hemorrhage” involve instrumentation of animals and withdrawing blood to various end points to simulate shock. Many of these experiments have been criticized for their failure to reproduce clinical circumstances. In models of uncontrolled hemorrhage, vascular injuries are created to allow free hemorrhage. A “hybrid model” of controlled and uncontrolled hemorrhage was created by Stern and associates. Pigs were subjected to controlled hemorrhage by rapidly withdrawing blood from a femoral catheter to a significant level of shock (30 mmHg) over 30 min, removing an average of 40 mL/kg. Subsequently, uncontrolled hemorrhage was induced by creating a tear in the aorta. Resuscitation volumes of 55.8 mL/kg and 90 mL/kg of saline were used in a short period of time, after which shed blood was reinfused. These were the volumes necessary to attempt to achieve a preselected mean arterial pressure (MAP). The high volumes used in these experiments are roughly equivalent to 6.3L in a 70 kg human being. Despite aggressive fluid and blood administration, the goal MAP was never achieved in the Stern study. It might be argued that the two insults induced in this swine model are too severe, that the model is not clinically relevant, and that the volumes used for resuscitation were excessive. Nonetheless, fluid resuscitation using varying amounts of volume is associated with a lower mortality than when no resuscitation is administered (Fig. 4-11). In the Stern study, the survival rate fell precipitously in animals receiving a large volume in a short amount of time in an attempt to restore a MAP of 80 mmHg. Although significantly increased blood loss also was seen in the large-volume (90 mL/kg) resuscitation group in the Stern study, the mechanism of death was not studied. In the Soucy study, uncontrolled hemorrhage treated with a moderate volume (40 mL/kg) of isotonic saline solution was associated with increased survival times and decreased mortality.

Current studies in unanesthetized animals imply that early resuscitation of hemorrhagic shock with a moderate volume of isotonic solution is superior to delayed resuscitation with isotonic or hypertonic solutions. Volume replacement appears to result in a centrally mediated reflex relaxation of peripheral vascular beds and increased perfusion with little or no associated increase in mean arterial pressure. The existence of this postulated “shock set-point” of autoregulated low blood pressure could account for the poor results in the studies outlined above in which the resuscitative goal was to achieve a predetermined “normal” blood pressure.

Timing of Resuscitation

In the first part of the twentieth century, Cannon stated, “Hemorrhage in the case of shock *may* not have occurred to a marked degree because blood pressure has been low and flow too scant to overcome the obstacle offered by a clot.” Cannon also stated, however, “The low blood pressure of shock has been met by the injection of normal or hypertonic salt.” Experience gained in the treatment of soldiers during the Korean and Vietnam wars established that early resuscitation was associated with a lower incidence of renal failure and with increased survival (see Table 4-1).

Current clinical protocols for resuscitation from hemorrhagic shock are based on numerous studies in the basic and clinical sciences. The practice of early administration of fluids to victims of hemorrhagic shock has been the dominant method of treatment for twenty years, but this method has been challenged.

Animal studies of “delayed resuscitation” in rats after uncontrolled hemorrhage implied that there was no beneficial effect of volume resuscitation. In a swine model, Bruttig and colleagues showed that early resuscitation was associated with higher mortality when compared to animals with “delayed resuscitation,” which was *no* resuscitation. This “early resuscitation” necessitated giving the entire resuscitation volume in 6 min, beginning 4 min after the aortic tear. This resulted in little clot formation, probably because of extremely rapid dilution of clotting factors, and it resulted in three times the bleeding volume compared to controls without fluids. The Bruttig study is at variance with that of Gross and associates, in which delayed resuscitation for uncontrolled hemorrhage was associated with increased mortality when compared to animals that were not resuscitated.

A droperidol-ketamine (DK) mixture was used for anesthesia by many investigators. To examine the mechanism responsible for these increases in mortality associated with resuscitation, Bilytskyj and coworkers demonstrated that DK anesthesia was associated with increases in blood loss and mortality. The poor outcomes were thought to be from the vasodilatory effects of the anesthetic drugs. In the same study, pentobarbital was found not to cause vasodilatation was not associated with excessive blood loss. This effect of anesthesia was confirmed in a study of rats observed after receiving pentobarbital anesthesia.

DK anesthesia was associated with ongoing hemorrhage and increased mortality, but in contrast to several previous studies, all animals resuscitated in the Shires study fared better, regardless of the anesthetic used (Fig. 4-12).

Two large clinical trials were performed to test the “scoop and run” philosophy, delaying resuscitation until the patient reached the hospital. In both studies, delayed resuscitation was not associated with decreased survival. The groups with delayed resuscitation did not progress more favorably than the groups with early resuscitation. Additionally, these studies were performed in selected patients in an urban setting with relatively short prehospital time; therefore, the “delay” was only approximately 30 min. Modern urban emergency medical services usually can be relied on to provide transport to a Level I or II trauma center within 30 min. It should be noted that in both clinical trials fluids were not withheld. The Kaweski study was a retrospective review of data obtained from trauma patients. Patients who were less severely injured received less fluid. In the Mattox study, the “delayed resuscitation” group received 771 ± 1228 mL of fluid preoperatively and were given comparable amounts of crystalloid and blood in the operating room. In neither study did patients who were given fluid have a worse outcome.

Bickell and colleagues reported a randomized clinical trial on the effect of presurgical infusion of isotonic fluids in penetrating trauma to the torso. The study has several methodological problems, including the inclusion of resuscitated patients with the nonresuscitated patients prior to statistical analysis, precluding validation of results. In analyzing the data with appropriate statistical methods, no differences can be found in blood pressure at admission, incidence of complications, length of hospitalization, or survival.

Adjuvant Therapy Vasopressors

The clinical state of hemorrhagic shock is defined by the presence of inadequate tissue perfusion resulting from hypovolemia. Treatment with vasopressors during shock may elevate blood pressure, but at the expense of further increased peripheral resistance and diminished tissue perfusion. Vasopressor therapy also may worsen the plasma volume deficit associated with hemorrhage, and the use of such agents in place of adequate fluid resuscitation is inadvisable.

Adenosine Triphosphate (ATP)

The demonstration of decreased high-energy phosphate levels by some investigators during hemorrhagic shock, as well as a decreased rate of conversion from adenosine diphosphate (ADP) to ATP, led to an interest in ATP replacement as an adjuvant therapy for hemorrhagic shock. In experimental models, ATP-MgCl₂ has been found to improve survival rates in potentially lethal shock when combined with adequate fluid resuscitation. However, because ATP is rapidly degraded in plasma and tissue and cannot cross intact cellular membranes because of its charge, it is unlikely that the improved survival is a result of restoration of cellular ATP content. The improvements may be from the reversal of hypoperfusion secondary to local vasodilatory effects, degradation and provision of phosphate precursors that may cross the cell membrane, or blockade of cellular uptake of calcium. Infusion of ATP-MgCl₂ may result in marked hemodynamic instability in the hypovolemic patient, which limits its usefulness in the clinical treatment of hemorrhagic shock.

Positioning

Most first-aid courses teach that the patient in shock should be placed in the head-down position. It is true that some forms of shock, particularly neurogenic shock, respond to the head-down position, but the effect of posture on the cerebral circulation when confronted with true hypovolemia has not been defined. The patient with multiple trauma often has injuries in the abdomen and chest, so that the routine use of the Trendelenburg, or head-down, position may interfere with respiratory exchange far more than when the patient is left supine. The beneficial effect of the head-down position probably is the result of transient autotransfusion of pooled blood in the capacity or venous side of the peripheral circulation. This beneficial effect can be obtained easily by elevating both legs while maintaining the head, trunk, and arms in the supine position. This is the preferred position for the treatment of hypovolemic shock.

MAST Garment

There has been enthusiasm for the in-the-field application of *military antishock trousers*, the MAST garment. When applied to the extremities with modest pressures, the garment functions well as a splint and may control some venous bleeding. When applied at high pressures, the resultant increase in total peripheral resistance may elevate the systemic pressure while decreasing cardiac output and peripheral perfusion. Additionally, inflation of the abdominal bolster may compress the inferior vena cava, impairing venous return to the heart by further increasing the venous resistance. Several reports of reperfusion injury with compartment syndromes in uninjured limbs have appeared. The MAST device may be of value when used strictly as a temporizing device or occasionally as specific treatment of bleeding pelvic fractures. Its use must not delay the immediate repletion of intravascular and extravascular volume by fluid therapy or interfere with rapid transport of the injured patient.

Pulmonary Support

In the past, most writing on the treatment of hypovolemic shock stated that breathing high oxygen concentrations probably is of little value during a period of hypotension. These conclusions were based on the notion that the principal defects are in volume flow to tissues and decreased cardiac output. The oxygen saturation in the majority of patients with uncomplicated hypovolemic shock generally is normal, and the small increase in dissolved oxygen in the blood contributed by raising the P_{O_2} above this level is insignificant, particularly with a markedly decreased cardiac output. This concept continues to be valid in terms of improvement of the shock state or tissue oxygenation. Nevertheless, in the small but significant group of patients in hypovolemic shock in whom the oxygen saturation is not normal, the *initial* use of increased oxygen concentrations may be extremely important, because the fall in cardiac output accompanying hemorrhagic shock has been shown to compound existing defects in oxygenation. This can occur in patients with preexisting defects, such as chronic obstructive lung disease. More frequently, problems in oxygenation arise directly from the patient's injuries and may include a coexisting pneumothorax, pulmonary contusion, aspiration of gastric contents or blood, or airway obstruction. Although oxygen is not routinely administered to patients in shock, if any doubt exists as to the possibility of one of these circumstances or the adequacy of oxygenation of arterial blood, the initial administration of oxygen until the injuries to the patient have been diligently assessed is certainly justified. If oxygen is administered to patients under these circumstances, it should be delivered through a loose-fitting face mask designed for this purpose. If a controlled airway is indicated for other reasons, an endotracheal tube is ideal. The use of nasal catheters, particularly those passed into the nasopharynx, is avoided because of potential complications of pharyngeal lacerations and gastric distention. Gastric rupture has been recorded secondary to a nasal catheter's being inadvertently placed in the esophagus.

Antibiotics

Antibiotics were used in the treatment of hypovolemic shock for many years and were thought to exert a protective mechanism against the ravages of hypovolemia. Subsequent data failed to support this hypothesis. The use of antibiotics in patients who have open or potentially contaminated wounds continues to be sound practice when combined with good surgical debridement and care. The use of wide-spectrum antibiotics is advisable as a preventive measure in the severely injured patient. Cefoxitin 2 g I.V. has proved to be a safe and effective single agent in multiorgan abdominal injuries.

Analgesics

Treatment of pain in the patient with hypovolemic shock is rarely a problem. However, if the causative injury produces severe pain, e.g., fracture, peritonitis, or injury to the chest wall, control of pain becomes mandatory. When the patient is moved to an emergency facility where physicians and care are available, simple supportive measures (administration of intravenous fluids, passing of catheters) will give reassurance. The need for analgesics is greatly reduced, because the need to allay fear and anxiety is decreased. If the patient continues to have severe pain, the observations made by Beecher in World War II are critical. Many battle casualties received morphine or other narcotic agents by subcutaneous administration soon after wounding. Because these analgesics did not enter the circulation immediately, the pain continued and the patient ultimately received several doses that were poorly absorbed. Once effective therapy was begun for shock, the doses previously administered were absorbed, and profound sedation resulted. As a result, the recommendation was made that small doses of narcotics be given *intravenously* for the management of pain in the patient with shock.

Steroids

Adrenocorticoid depletion was regarded as a contributory factor in shock after it was learned that the presence of hypovolemic shock could deplete the adrenal cortex of adrenocortical steroids. Subsequent studies have shown that adrenocortical steroid production is stimulated maximally by the presence of hypovolemic shock. Steroid depletion with hypovolemic shock may occur in the elderly patient or in patients with specific adrenocortical diseases, such as incipient Addison's disease, postadrenalectomy patients, or patients who have had adrenal suppression with exogenous adrenocortical steroids. In these specific instances, the intravenous administration of hydrocortisone is desirable. In the trauma patient with hypovolemic shock, administration of adrenocorticoids is not indicated.

Monitoring

Continuous bedside monitoring of circulatory efficacy, including assessment of the heart rate, arterial blood pressure, urinary output, and peripheral perfusion, remains the cornerstone for resuscitation. Adequate resuscitation is indicated when adequate cerebral function and urinary output are restored. In the patient with multiple injuries, central venous pressure (CVP) monitoring is useful. Although left ventricular overload can occur while right ventricular function and CVP remain normal, this is not often the case in the absence of myocardial injury. Changes over time in the CVP with fluid infusion do indicate the ability of the myocardium to pump the volume presented to it. A normal-to-depressed CVP that does not rise with rapid administration of crystalloid fluid usually indicates continuing hypovolemia. The presence of an elevated CVP or its rapid rise in response to fluid administration is indicative of impairment of the pumping mechanism. Although this usually represents primary myocardial deficiency and should be treated as outlined in the following section on cardiogenic shock, any mechanical obstruction to venous return with cardiac tamponade or mediastinal compression by intrapleural air or blood must be immediately attended to in the injured patient. The use of a balloon- tipped Swan-Ganz catheter allows measurement of pulmonary artery and pulmonary wedge pressures as well as thermodilution cardiac output determinations. The early use of the Swan-Ganz catheter rarely is necessary in the initial emergency department treatment for hemorrhagic shock.

Hemorrhagic shock may become refractory to the therapies described above and become irreversible. Complete vascular collapse with hypotension unresponsive to volume or drug intervention eventually leads to lethal central nervous system and cardiac dysfunction. Irreversibility is difficult to define but has been related to the duration and volume of hemorrhage, the age and preexisting cardiovascular fitness of the patient, and the coexistence of massive trauma with multiple direct organ derangement. Before the physician concludes that refractory shock has occurred, the multiple causes of failure to respond to therapy should be resolved. These include continuing unsuspected blood loss into the chest or abdomen, inadequate volume replacement, inadequate clotting, multisystem trauma with occult thoracic injuries, including cardiac tamponade and hemopneumothorax, and acute myocardial insufficiency from direct injury or secondary to prolonged coronary hypoperfusion.

Cardiogenic Shock

Cardiogenic shock occurs when the heart is unable to generate sufficient cardiac output to maintain adequate tissue perfusion. Unlike hypovolemic shock, cardiogenic shock is manifested by hypotension in the face of adequate intravascular volume. Cardiogenic shock that is unresponsive is associated with significant mortality and morbidity, particularly in conjunction with myocardial infarction and the secondary end-organ injuries of pulmonary edema, oliguric renal failure, and coma.

Pathophysiology

Myocardial failure may result from a variety of diseases, including valvular heart disease, cardiomyopathy, and direct myocardial contusion. Acute myocardial infarction is the most frequent cause of cardiogenic shock, which is often fatal when 40 percent of the left ventricular mass has been lost. Papillary muscle dysfunction, ischemic ventricular septal defects, massive left ventricular infarction, and arrhythmias are complications of acute myocardial infarction that may lead to cardiogenic shock.

The initial compensatory response to diminished myocardial contraction is tachycardia, in an attempt to maintain cardiac output, despite a decreased left ventricular ejection fraction, at the expense of increasing myocardial oxygen consumption. As cardiac index falls below 2 L/min/m² hypotension produces reflex sympathetic vasoconstriction. This attempt to maintain central pressure by increasing peripheral vascular resistance leads to decreasing organ perfusion. An increase in afterload further impairs left ventricular function and increases myocardial work. The combination of increased myocardial oxygen demand, hypotension, and shortened diastole amplifies the mismatch between coronary arterial oxygen delivery and myocardial oxygen demand, extending the zone of infarction in the patient who does not receive prompt intervention.

Treatment

Although the goal of medical management of cardiogenic shock has been to enhance ventricular performance and improve global perfusion, the traditional management with fluids and inotropic drugs continues to yield a mortality of 80 to 90 percent. The techniques that maximize ventricular performance paradoxically increase myocardial oxygen demand at a time when therapy, to limit infarct size and salvage reversibly ischemic myocardium, should include minimizing myocardial demand and attempting to provide early reperfusion. Initial therapy includes optimizing ventricular preload by manipulating filling pressure, decreasing afterload in the patient with adequate systolic pressure, correcting arrhythmias, and improving contractility to sustain vital organ perfusion.

Monitoring and Volume Management

Supplemental oxygen, pain relief and sedation, and continuous electrocardiographic monitoring should be initiated early. A Foley catheter is inserted for monitoring urine output. Cutaneous oximetry and automated arterial blood pressure cuff measurements can be used in place of an intraarterial catheter for continuous arterial pressure monitoring and blood gas determinations. Placement of a Swan-Ganz catheter for measurement of cardiac output and pulmonary artery wedge pressure is crucial to therapeutic decision-making in these critically ill patients. Cardiogenic shock with low cardiac output and arterial hypotension can occur in some patients with normal to slightly elevated pulmonary artery wedge pressures. A small increase in left ventricular filling pressure by volume infusion may maximize cardiac output via the Frank-Starling mechanism. It should be emphasized that although hemodynamic measurements suggest myocardial insufficiency, mechanical obstruction, such as cardiac tamponade in the injured patient or pulmonary embolism in the postoperative patient, may be present. Although these diagnoses are made largely on clinical grounds in the emergency setting, volume infusion usually is of some benefit while echocardiography, pericardiocentesis, or thoracentesis are performed quickly. Constant vigilance is required during volume challenge in this setting; increased filling pressures may lead to further myocardial ischemia and acute pulmonary edema. If any pulmonary complications evolve, early intubation and mechanical ventilation will decrease the myocardial oxygen demand as a consequence of the increased work of breathing and correct arterial hypoxemia that may further impair cardiac performance.

Inotropic Agents

The beta₁-adrenergic receptors of the myocardium respond to exogenous sympathomimetic drugs by increasing contractility and improving cardiac output. These effects are obtained at the cost of increasing myocardial oxygen demand in the setting of already-compromised myocardial perfusion, but intravenous infusion of dopamine may promptly reverse life- threatening hypotension and restore mean arterial pressure to about 80 mmHg. The dopaminergic effects of splanchnic, coronary, and renal vasodilatation at low (2 to 5 mg/kg/min) doses are augmented by adrenergic-mediated increases in contractility and heart rate as dosages rise to 5 to 8 mg/kg/min. At higher doses, alpha-adrenergic receptor effects predominate, and central arterial pressure can increase while coronary artery constriction further decreases coronary blood flow. Dopamine also causes a variable increase in heart rate and can precipitate other arrhythmias, which underscores the need to titrate the lowest acceptable dose. Dobutamine, a synthetic catecholamine with predominantly inotropic effect, appears to be less arrhythmogenic and may redistribute cardiac output to the coronary circulation. Studies appear to favor dobutamine over dopamine for treating cardiogenic shock after cardiopulmonary bypass or myocardial infarction. Digitalis remains a controversial drug in acute pump failure. Although very useful in the treatment of

supraventricular arrhythmias, digitalis increases myocardial oxygen consumption and adds very little hemodynamic benefit relative to therapy with sympathomimetic agents.

Vasodilator Agents

Some patients with low cardiac output and high filling pressures have near- normal arterial blood pressure in the setting of profoundly decreased perfusion by clinical assessment. In these circumstances, systolic ventricular wall stress is high, and reducing afterload should increase cardiac output and decrease myocardial work. An agent such as sodium nitroprusside should be used with extreme caution in hypotensive patients because redistribution of an already depressed cardiac output away from the coronary and cerebral circulation can occur, and any decrease in systemic diastolic pressure would further depress coronary perfusion pressures.

Mechanical Support

As the role of early reperfusion strategies evolves (including thrombolysis, percutaneous transluminal coronary angioplasty, and emergency bypass), mechanical therapy can temporarily support the failing myocardium until these modalities are initiated or some myocardial recovery occurs. Despite significant associated morbidity, successful mechanical cardiac support will maintain organ perfusion while decreasing myocardial oxygen demand by unloading the left ventricle and reducing myocardial work. The intraaortic balloon pulsation device has been used most widely. It can be inserted at the bedside and fulfills the criteria of elevating diastolic blood pressure, which increases pulmonary perfusion, while decreasing myocardial work, by increasing cardiac output distal to the ventricle. It is unclear whether this device improves long-term survival, but it clearly supports the failing myocardium while recovery or other interventions proceed. Those patients with surgically correctable problems after acute myocardial infarction appear to respond better than patients who were not operative candidates after temporary mechanical support.

Left heart bypass by left atrium–to–femoral artery circulatory bypass and temporary implantation of left ventricular assist devices may be even more effective in assuming cardiac work. These techniques usually have been limited to patients with cardiogenic shock after cardiac surgery, because operative placement is required. The role of these techniques may broaden if ongoing studies support improved survival rates from early revascularization by thrombolytic therapy, coronary angioplasty, and emergency surgery.

Arrhythmias

Rapid ventricular rates can depress cardiac output to shock levels. Ventricular end-diastolic pressure decreases as a result of shortened filling time, and ventricular relaxation is incomplete by the end of the abbreviated diastolic period. Cardiac output falls because stroke volume cannot be compensated for by the rapid heart rate. Digoxin is the drug of choice for atrial fibrillation or atrial flutter, but electrical cardioversion should be promptly undertaken for tachycardia that produces hypotension and hypoperfusion. Resistant sinus tachycardia, while well tolerated by the normal heart, may produce a low flow state in the diseased heart. Verapamil has been useful in treating tachyarrhythmias of atrial origin, and propranolol slows sinus tachycardia. Beta blockade can further decrease cardiac output in this setting, and a careful search for the cause of the sinus tachycardia, including fever, hypovolemia, and drug effect, should be undertaken. Immediate nonsynchronized direct-current electrical shock is mandatory treatment for ventricular fibrillation or ventricular flutter that has caused cardiogenic shock with loss of consciousness. Ventricular fibrillation rarely converts spontaneously, and the subsequent rapid development of cardiovascular collapse demands prompt therapy. In the patient with acute myocardial injury, premature ventricular complexes may lead to ventricular tachyarrhythmias. Intravenous lidocaine usually is the initial treatment and also is given after cardioversion to prevent recurrent ventricular fibrillation. Bretylium tosylate has been useful in treating life- threatening ventricular tachyarrhythmias that are unresponsive to lidocaine or class Ia agents, such as procainamide.

Low cardiac output with ventricular rates less than 70 beats/min may occur in patients with impaired cardiac performance. Stroke volume cannot increase to compensate for the pathologic bradycardia. Electrical pacing of the heart at a rate of 80 to 100 beats/min can restore sufficient cardiac output whether the underlying mechanism is sinus bradycardia, atrial fibrillation with slow ventricular rate (e.g., digitalis toxicity), or atrioventricular dissociation.

Neurogenic Shock

Neurogenic shock (“primary shock” in the older classification) is the form of shock that occurs after serious interference with the balance of vasodilator and vasoconstrictor influences to the arterioles and venules. This is the shock that is seen with clinical syncope—the sudden exposure to unpleasant events, such as the sight of blood, the hearing of bad tidings, or the sudden onset of pain. Similarly, neurogenic shock often is observed with serious paralysis of vasomotor influences, as in high spinal anesthesia or injury to the spinal cord. The reflex interruption of nerve impulses also occurs with acute gastric dilatation.

The clinical picture of neurogenic shock is quite different from that classically seen in hypovolemic shock. While the blood pressure may be extremely low, the pulse rate usually is slower than normal and is accompanied by dry, warm, and even flushed skin. Measurements made during neurogenic shock indicate a reduction in cardiac output, but this is accompanied by a decrease in resistance of arteriolar vessels and a decrease in the venous tone. Consequently, there appears to be a normovolemic state with a greatly increased reservoir capacity in the arterioles and venules, thereby inducing a decreased venous return to the right side of the heart and hence a reduction in cardiac output.

If neurogenic shock is not corrected, a reduction of blood flow to the kidneys and damage to the brain result, and the ravages of hypovolemic shock appear. Treatment of neurogenic shock usually is obvious. Gastric dilatation can be treated rapidly with nasogastric suction. Shock due to high spinal anesthesia can be treated effectively with administration of fluids and a vasopressor, such as ephedrine or phenylephrine (Neo- Synephrine). These drugs will increase cardiac output, restore venous tone, and elevate systemic blood pressure by arteriolar constriction. With the milder forms of neurogenic shock, such as fainting, simply removing the patient from the stimulus, relieving the pain, and elevating the legs is adequate therapy while the vasoconstrictor nerves regain the ability to maintain normal arteriolar and venous resistance.

There is rarely a need for hemodynamic measurement in this usually self- limited form of hypotension. An exception is when this form of shock results from injury, as with spinal cord transection from trauma. In this instance there may be significant loss of blood and extracellular fluid into the area of injury surrounding the cord and vertebral column. Considerable confusion can arise as to the relative need for fluid replacement or for vasopressor drugs under these circumstances. Similarly, if surgical intervention for any reason becomes necessary, hemodynamic measurements may be of great value in the management of these patients. In uncomplicated neurogenic shock, central venous pressure should be slightly low, with a near-normal cardiac output. If hypovolemia ensues, central venous pressure and cardiac output decrease. Careful monitoring of central venous pressure may be necessary. Fluid administration without vasopressors in this form of hypotension may produce a gradually rising arterial pressure and cardiac output without elevation of central venous pressure by gradually “filling” the expanded vascular pool; caution must be used during fluid administration.

In managing these patients, slight volume overextension is much less deleterious than excessive vasopressor administration. The latter decreases organ perfusion in the presence of inadequate fluid replacement, particularly in the body proximal to the cord injury. Balance is best obtained by maintaining a normal central venous pressure that rises slightly with rapid fluid administration (ensuring adequate volume) and using a vasopressor such as phenylephrine judiciously to support arterial pressure.

Septic Shock

Sepsis, the sepsis syndrome, and septic shock define the continuum of human response to infection. Although any agent capable of producing infection, including viruses, parasites, and fungi, may generate septic shock, the most frequent causative organisms in the antibiotic era are gram-negative bacteria, and occasionally gram-positive bacteria. The initial infectious process appears to be only a stimulus for a series of host responses that may culminate in death, even in the absence of infection at the time of death. Over the past few decades, the incidence of gram- negative sepsis has risen dramatically, from fewer than 100 reported cases in the early 1920s to an estimate of 400,000 cases per year, of which approximately 100,000 episodes of septic shock are treated in the United States. Even in the most recent series, overall mortality exceeds 30 percent, with mortalities over 80 percent in complicated cases with associated multiple organ system failure.

Gram-negative organisms supplanted gram-positive organisms as the predominant cause of septic shock after the widespread application of effective antibiotics for gram-positive infections. Despite increasingly powerful gram-negative antibiotics, the incidence of gram-negative sepsis continues to rise. Proposed causes for this increasing incidence include a developing reservoir of resistant and virulent organisms, concentration of infected patients in critical-care settings, more extensive operations in elderly and poor-risk patients, initial salvage of the severely injured, and a growing population of patients immunosuppressed by organ transplant protocols, radiotherapy, and chemotherapy.

The most common source of gram-negative infection is the genitourinary system. This frequently follows instrumentation of the urinary tract, which is performed in up to one-third of hospitalized patients. The second most frequent site of origin is the respiratory system, followed by the alimentary system, including the biliary tract. Increasing and prolonged use of indwelling catheters for monitoring and hyperalimentation is responsible for many bloodstream infections. The early and aggressive use of appropriate antibiotic therapy with other treatments outlined below have a crucial role in favorable outcome. Patients with a surgically correctable focus of infection have a more favorable prognosis.

Clinical Manifestations

Gram-negative infections frequently are heralded by the onset of chills and temperature elevations above 38°C (101°F). The patient may rapidly progress to evidence of altered organ function, most often renal and pulmonary in nature. This clinical situation plus the development of hypotension completes the picture of septic shock. Unlike most other forms of shock, the patient who is normovolemic has hypotension despite an increased cardiac output and a reasonable filling pressure. The peripheral resistance is low and produces the paradoxical “warm shock” with pink, dry extremities. The high cardiac output often is associated with a decrease in oxygen utilization and a narrowed arteriovenous oxygen difference.

In a patient who is initially hypovolemic or persists in the shock state, a hypodynamic pattern emerges that is characterized by a falling cardiac output, low central pressures, and increased peripheral resistance with more typical cold, pale extremities consistent with global hypoperfusion. Early volume replacement frequently increases cardiac output and produces a hyperdynamic circulation, while the patient later in shock is unresponsive to volume replacement and has a low cardiac output with increasing metabolic acidosis.

Concomitant laboratory tests usually show an elevation in the white blood cell count, but leukopenia may be present in immunosuppressed and debilitated patients or those with overwhelming white cell consumption from sepsis. Thrombocytopenia may be an early indicator of gram-negative sepsis, particularly in pediatric and burn patients. Mild hypoxia with compensatory hyperventilation and respiratory alkalosis are common early findings, despite clinical or radiological evidence of intrinsic pulmonary disease. Although the onset of hypotension may be coincident with these clinical signs of infection, a patient can have relatively subtle findings of hyperventilation, respiratory alkalosis, and altered sensorium for a prolonged period before shock begins.

Septic shock is the result of numerous complex interactions between exogenous and endogenous mediators and host responses to these stimuli. The wide individual variation in septic shock and in response to various interventions in human and experimental studies serves to underscore this intricate and poorly understood pathogenesis. Necessary host responses to local injury and infection form the local defenses against progression to systemic illness. When the ability to contain local infection is overwhelmed, systemic illness may result from the inappropriate systemic effects of these mediators.

At the organ level, cardiovascular response to systemic infection, in the absence of hypovolemia, is the development of a hyperdynamic state. A number of vasoregulatory mediators combine to produce a net decrease in systemic vascular resistance. This is quite distinct from the increased vascular resistance seen in responses to hypoperfusion in

other forms of shock. As cardiac index increases, the arterial-venous oxygen difference narrows. An apparent effect in peripheral oxygen extraction was originally ascribed to pathologic arteriovenous shunting. However, microvascular blood flow in many capillary beds does not appear to be altered in septic shock. Additionally, neither cellular hypoxia nor any defects in the energy- producing metabolic pathways have been documented by studies using in vivo nuclear magnetic resonance spectroscopy. Despite increased cardiac index and decreased oxygen extraction, no direct evidence for cellular hypoxia was detected.

Although patients with hyperdynamic septic shock have an increased cardiac output, detailed studies in patients and animals report a depression in myocardial function. The long-postulated myocardial depressant factor, although poorly characterized biochemically, appears to be a reasonable explanation for documented decreases in left ventricular ejection fraction despite acceptable filling pressures. The pathophysiologic mechanisms that produce organ dysfunction in the septic state, prior to the onset of hypotension, are compounded by the development of refractory hypotension, with tissue ischemia probably contributing a component of cell death in end-stage hypodynamic septic shock.

Pathophysiology

Many studies suggested that the agents responsible for the induction of fever were endogenous products, and this led to the concept of endogenous mediators of the action of endotoxin. While early studies demonstrated that interleukin-1 (IL-1) was an endogenous mediator of infection, recent studies by Beutler and coworkers and by Tracey and coworkers demonstrated that cachectin- *TNF* is a central and proximal mediator of the host response to endotoxemia and bacteremia.

Cachectin is a cytokine secreted by activated macrophages; it was purified to homogeneity by Beutler and Cerami while they were looking for a factor that mediated cachexia. This protein, which is produced predominantly by cells of macrophage lineage, was purified and characterized by Aggarwal and associates. It is a 17-kD protein with 50 percent homology with lymphotoxin, a lymphokine with which it shares many properties. Because of the similarity of function, lymphotoxin has been called TNF-b, and TNF- cachectin is referred to as TNF-a. After protein and complementary DNA (cDNA) sequence analysis was completed, it became apparent that cachectin was identical to a factor called tumor necrosis factor, isolated for its ability to mediate endotoxin-induced tumor cytotoxicity.

Early studies by Shires and colleagues recognized that cellular dysfunction accompanying acute hemorrhagic shock is associated with a reduction in the transcellular membrane potential. Subsequent studies in animal models of septic shock by Illner and associates and a primate model of septic shock by Trunkey and associates demonstrated that similar changes in the transcellular membrane potential occurred during septic shock and that the reduction in the membrane potential preceded the onset of hypotension. A similar decrease in the transmembrane potential occurred in human volunteers in response to endotoxin infusion.

To understand the etiology of the reduction in cellular membrane potential during sepsis, the effects of TNF-a on skeletal muscle membrane potentials were studied. It was found that TNF-a was capable of inducing membrane depolarization in vitro. Using the in vitro measurement of skeletal muscle membrane potentials as a bioassay, Tracey and colleagues were able to demonstrate that plasma from critically ill patients with septicemia in a surgical intensive care unit contained a circulating factor that caused muscle membrane depolarization. Preincubation of the plasma samples with a monoclonal antibody to human TNF-a was able to abrogate the in vitro decrease in the membrane potential. Infusion of TNF-a intraarterially in dogs induced a rapid decrease in skeletal muscle membrane potential and extremity lactate efflux, similar to that seen in sepsis before the onset of shock. These changes preceded the onset of hypertension by 2 to 3 h and were associated with an increased fluid requirement secondary to third-space losses. The time course of these changes, i.e., skeletal muscle membrane depolarization and lactic acidemia preceding circulatory compromise, is consistent with clinical observations of the development of septic shock. These initial studies, confirmed by Schirmer and associates, suggested that TNF-a was an important mediator of septic shock.

To assess the role of TNF-a in bacteremic shock in primates, monoclonal antibodies [F(ab')₂ fragments] to human TNF-a were administered to baboons 2 h before challenge with a lethal dose of *Escherichia coli*. The administration of anti-TNF-a antibodies prevented the development of subsequent septic shock and associated organ dysfunction. Subsequent analysis of plasma samples from these baboons revealed that anti-TNF-a monoclonal antibody pretreatment before the infusion of *E. coli* abrogated the increases in circulating interleukin-1 (IL-1) and IL-6 seen in the untreated animals (Fig. 4-13). In a similar study, Mathison and colleagues demonstrated that administration of anti-TNF-a antibodies protected rabbits from lethal shock induced by endotoxemia. These studies established the role of TNF-a as a central and proximal mediator of experimental septic shock.

Circulating TNF-a can be found in response to endotoxin administration in human beings and peak levels may correlate with sepsis and overall mortality when detected in human disease states. Although endotoxin or gram- negative bacteria infusions may not accurately reproduce the mode of onset of clinical septicemia, the fact that TNF-a is present in human critical disease states strongly supports an important role of TNF-a in the pathophysiology of septic shock in human beings.

The mechanisms through which TNF-a induces the pathophysiologic changes associated with shock are being studied. TNF-a induces the synthesis and secretion of a variety of secondary mediators, including other cytokines, prostaglandins, leukotrienes, platelet-activating factor, complement components, and activation of the clotting cascade, that possess toxic properties capable of causing widespread tissue damage if liberated systemically. In addition, *lipopolysaccharide (LPS)* may synergize with TNF-a to induce many of the toxic effects mediated by TNF-a. The TNF-a-induced release of these factors may be responsible for pathologic changes seen in the lungs, liver, bowel, and kidneys in response to sepsis and septic shock.

Discrepancies in TNF-a detection in the circulation may be partially explained by the presence of circulating TNF receptors that may interfere with the measurement of TNF-a in the plasma. These receptors can neutralize TNF-a cytotoxicity, measured by bioassay, and interfere with the detection of TNF-a by enzyme-linked immunosorbent assay (ELISA) if the antibodies used do not recognize TNF-a-soluble receptor complexes. It also has been noted that competitive radioimmunoassays (RIAs) may overestimate TNF-a concentrations in the presence of soluble TNF receptors. RIA binding of the labeled antigen to the soluble receptors reduces binding to the antibody. This results in a low measured value and a correspondingly high calculated value for the concentration of unlabeled TNF-a.

Concentrations of soluble receptors found in endotoxemic volunteers or critically ill patients are sufficient to neutralize or attenuate the cytotoxicity associated with TNF-a concentrations observed in mild inflammation. Such levels apparently are inadequate to neutralize the toxicity associated with the excessive or persistent TNF-a activity found during overwhelming sepsis.

Evidence of the therapeutic potential of TNF-blockade was obtained by Tracey, who administered monoclonal murine antihuman TNF-a antibodies to baboons before administering a lethal dose of live *E. coli* and demonstrated survival for at least 48 h. In addition to the difference in survival, anti-TNF-a antibodies attenuated the leukopenia and the release of catabolic stress hormones (epinephrine, norepinephrine, and glucagon) that normally accompany such a lethal bacteremia. The appearance of other cytokines known to be produced during sepsis (IL-1b and IL-6) also was significantly reduced.

A different approach to achieving blockade of excessive TNF-a has been provided by the recent identification of naturally occurring inhibitors of TNF-a activity in human serum and urine. The isolation and characterization of these inhibitors revealed them to be the extracellular domains of the type I and type II TNF receptors, which are shed from the cell surface in response to some of the same inflammatory stimuli known to induce TNF-a production. Both naturally occurring soluble TNF receptors (sTNFR-I and sTNFR-II) can be found in human volunteers after endotoxin administration.

To investigate the therapeutic potential of exogenous soluble TNF receptors, recombinant soluble TNFR-I (rsTNFR-I) was administered to baboons with a lethal bacteremia. The baboons received an intravenous bolus of live *E. coli* (LD100), followed by a 3-h primed continuous infusion of rsTNFR-I sufficient to provide a 300-fold molar excess of sTNFR-I over the maximum TNF-a plasma concentration. Administration of rsTNFR-I significantly reduced the volume of resuscitation fluid required to maintain hemodynamic stability and the maximum decline in mean arterial pressure (MAP) (Fig. 4-14).

The term interleukin-1 applies to two different polypeptides, IL-1a and IL- 1b. These biochemically distinct forms share only 26 percent amino acid homology, but both varieties bind to the same cell-surface receptors and induce the same biologic responses. IL-1, like TNF-a, is one of the key mediators of the host response to infection, inflammation, and injury. The two cytokines have been found to have many overlapping biologic functions.

Because much of the circulating IL-1 is thought to represent excess local tissue production, the infrequent detection of IL-1 in the circulation during clinical sepsis is not surprising. There is evidence that local production of IL- 1 and TNF-a is increased in inflammation even in the absence of circulating cytokine, suggesting that assays of local tissue levels may be more appropriate than measurement of circulating concentrations.

A naturally occurring inhibitor of IL-1 was identified in the urine of febrile patients as well as in the plasma of endotoxemic volunteers and critically ill patients. This inhibitor is a 17-kD polypeptide with 26 percent and 19 percent sequence homology to IL-1b and IL-1a, respectively, which binds to IL-1 cell-surface receptors. Despite the sequence homology and the protein's ability to bind to both types of IL-1 receptors, it has no agonist activity, and it was termed interleukin-1 receptor antagonist (IL-1ra).

IL-1ra blocks many functions of IL-1 in vitro, including: lymphocyte proliferation; increased adhesion of endothelial cells for neutrophils and eosinophils; synthesis of IL-1, TNF-a, IL-6, and IL-8 by monocytes; and nitric oxide production in human smooth muscle cells.

Fischer and associates investigated the potential benefit of exogenous IL- 1ra in severe sepsis by administering a lethal bacteremia to nonhuman primates. Administration of IL-1ra significantly decreased the hypotension and fall in cardiac output observed in controls and improved survival. In addition, circulating levels of IL-1b and IL-6 were significantly reduced. TNF- a levels were unchanged by IL-1ra, leading to the conclusion that the control mechanisms of these two cytokines are independent (Fig. 4-15).

These animal models provided the basis for initiating clinical trials of IL-1ra in the treatment of sepsis. In patients in whom IL-1 activity is adequately balanced by endogenous IL-1ra, administration of exogenous IL-1ra may not be beneficial and may potentially exacerbate certain types of infection.

Since endotoxin is the bacterial component largely responsible for the toxic effects of gram-negative infection, methods of blocking endotoxin are being investigated. One approach is the administration of antiendotoxin antibodies. Ziegler and coworkers reported an improvement in survival in a subset of patients with sepsis who had documented gram-negative bacteremia and were treated with HA-1A, a human monoclonal IgM antibody, against the lipid A portion of endotoxin. There was no improvement in those patients without positive blood cultures. Another report showed improved survival in patients with gram-negative sepsis (with positive cultures) but without shock who were treated with E5, a murine monoclonal IgM antibody also against lipid A. A follow-up study failed to confirm the beneficial effects of E5.

Endogenous *LPS*-binding proteins may be used to block the effects of endotoxin. One such protein, first isolated from neutrophil granules, is bactericidal/permeability-increasing protein (BPI). This peptide specifically binds to LPS and has been shown to inhibit endotoxin-mediated TNF-a release in vitro by peripheral blood mononuclear cells and in vivo in the murine lung. BPI, when preincubated with LPS, blocks LPS-mediated pyrogenicity in rabbits. Early data suggest that BPI may have therapeutic potential in the treatment of gram-negative sepsis.

The above factors are only some of the likely mediators of the host response to overwhelming sepsis and septic shock. Classic hormonal responses in septic shock are similar to those observed in hemorrhagic shock. Other potential mediators of septic shock include the kinins, endogenous opiates, and a number of recently purified cytokines believed to mediate host inflammatory responses.

Therapy

The control of infection by antibiotic treatment and early surgical debridement or radiologically guided drainage represent definitive therapy. Other recommended measures, including fluid therapy and the use of vasoactive drugs, represent adjunctive forms of therapy. Adjunctive therapies are useful in preparing patients for surgical or radiologic intervention or supporting patients until the infectious process can be controlled, but if the infection cannot be adequately controlled, the death of the patient is inevitable. It is essential that a prompt search for the source of infection be made as soon as infection becomes evident. It is advisable to institute supportive measures so that arterial and central venous or pulmonary capillary wedge pressures can be measured directly, and urine output and arterial and central venous blood gases should be measured, if indicated. If the infectious process requires drainage, operation should be performed as soon as possible after the patient has been stabilized, because some conditions, such as septic shock secondary to ascending cholangitis, will respond only briefly to adjunctive measures.

Antibiotic treatment should be based on the results of cultures and sensitivity tests when possible, but in the absence of these data broad- spectrum antibiotics should be started, including coverage for anaerobic organisms such as *Bacteroides* species or fungi, if clinically indicated. Antibiotic therapy should be adjusted when culture and sensitivity reports become available.

Correction of preexisting fluid deficits is essential in the septic patient and should proceed rapidly, though carefully, using pulmonary capillary wedge pressure and cardiac output as a guide to target appropriate replacement volumes, so that the detrimental pulmonary effects of fluid overload can be avoided. Monitoring is essential, because fluid requirements may be massive in these patients. Resuscitation requirements in excess of 10L of lactated Ringer's solution are common.

The use of corticosteroids in the treatment of septic shock was controversial until data from two well-controlled, prospective clinical trials demonstrated that the use of high-dose corticosteroids did not confer a survival advantage over non-steroid-treated controls (Fig. 4-16). The only indications for steroid treatment in patients with septic shock are hypoadrenalism and for stress coverage in patients taking steroids (or who recently completed a course of steroids) for immunosuppression or anti- inflammatory purposes.

Future clinical use of anti-TNF antibodies, protein C, or other antimediator treatment regimens probably will depend on early recognition of the sepsis syndrome for success unless used prophylactically in a population of patients at high risk.

Pharmacologic Support

Although fluid resuscitation remains the initial therapy for hypotension in sepsis, it is frequently necessary to administer drugs with inotropic or vasopressor activity. Dopamine is the initial inotropic agent used. As discussed above in the section on cardiogenic shock, dopamine has some vasopressor activity at higher doses that may be required in the volume-loaded patient with persistent profound hypotension. Augmentation of the impaired myocardial performance in septic shock is a reasonable goal. Dobutamine often increases cardiac input with less tachycardia and arrhythmia than dopamine. The beta-adrenergic vasodilatation from dobutamine infusion may not be tolerated by these hypotensive patients. Vasodilator drugs have been shown to improve cardiac output and oxygen delivery in normotensive septic patients. Their use in septic shock is limited by low systemic pressure or decreased cardiac filling pressures. More potent vasopressors, despite their obvious detrimental effect on peripheral perfusion, may be transiently unavoidable in patients who have persistent life-threatening hypotension, despite optimal fluid and dopamine infusions. Norepinephrine is a potent alpha-receptor agonist that usually is effective in raising pressure in patients for whom the measures described above have failed. Its use is sometimes preferred over high-dose dopamine because the cardiac effects are less. Epinephrine, a catecholamine with potent alpha- and beta-adrenergic activity, may support the blood pressure in patients who do not respond to norepinephrine. Polypharmacy frequently is ineffective and may be harmful in patients with complex cardiovascular alterations. Use of pressors is primarily for transient support while primary definitive therapy with antibiotics and drainage of surgical infection are being instituted.

Manipulations of Humoral Responses

Given the obviously complex and ill-defined interactions among a large number of mediators, therapy directed at any single agent is probably ineffective. Carefully tailored multidrug or serial antimediator therapy eventually may allow modulation of the deleterious systemic effects of the necessary host responses to injury and infection.

Initial trials of steroids, fibronectin, and naloxone were disappointing. More specific immunotherapy using monoclonal IgM antibodies to core lipopolysaccharide have been completed. Only small subsets of patients from each study showed statistical benefit. Treatment with HA-1A improved survival and organ function in the presence of gram-negative bacteremia with or without shock. The E5 trial was beneficial in patients with gram-negative bacteremia only in the absence of shock. None of the antibodies directed at lipid A or other epitopes of the core lipopolysaccharide are established therapeutic modalities.

The naturally occurring IL-1 receptor antagonist has been manufactured by recombinant technology. IL-1ra showed benefit in several animal models but failed to show efficacy in human sepsis. Monoclonal antibodies to *TNF* also are available and appear promising in patients with septic shock. The use of anticoagulants such as anti-thrombin III in the treatment of sepsis that is unassociated with shock is currently under investigation. The persistent significant morbidity and mortality from sepsis, despite powerful antibiotic therapy, prompt surgery, and carefully titrated fluid and drug support suggest that the future for improved therapeutic results lies in these exciting new treatments within the next decade.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 5 - Surgical Infections

Richard J. Howard

HISTORICAL BACKGROUND

Infection is encountered by all surgeons: by the nature of their craft, they invariably impair the first line of host defenses—the cutaneous or mucosal barrier. The entrance of microbes into host tissues is the initial requirement for infection. Preventing microbial penetration, reducing the microbial inoculum, and treating established infection have been important developments in reducing the mortality associated with surgery.

For most of surgical history, death from infection was common, although it was not until the end of the nineteenth century that the bacterial cause of surgical infection was appreciated. Death from infection was so common after compound fracture or fracture caused by gunshot wound that amputation was the standard treatment. Before antiseptic practices were instituted, mortality rates for amputation in times of war between 1745 and 1865 were between 25 and 90 percent. Mortality rates for amputation in civilian practice during the same period ranged from 5 to 50 percent.

The introduction of anesthesia by Long in 1842 and by Morton in 1846 increased the scope of surgery by permitting operations on body cavities and allowing surgeons to operate more slowly and deliberately, so that death from blood loss was diminished. Infection remained a great problem, however. Hospital gangrene (or *erysipelas*, a term popularized during the Civil War describing a necrotizing infection presumably caused by *Streptococcus*) and tetanus continued to plague surgeons and their patients. Many surgeons realized that a more favorable prognosis was associated with an infection that developed “laudable pus” rather than a more serious infection that was not associated with purulence. Surgeons did not yet understand the cause of infection.

Joseph Lister (1827–1912) made one of the great contributions to surgery by demonstrating that antisepsis could prevent infection and hence that compound fractures did not have to be treated by amputation. In March 1865 he began placing pure carbolic acid into wounds. Later he reduced the concentration to 10, 5, and 2.5 percent. In 1867 he published his initial series of papers on antisepsis, reporting, among other things, that compound fractures healed without infection when the wounds were treated with carbolic acid.

Wound antisepsis was not new with Lister. More than 20 articles appeared in British medical publications between 1859 and 1865 describing antiseptic treatment of wounds. Numerous agents had been placed in wounds since ancient times in an attempt to foster healing and prevent death—turpentine, pitch and tar, balsams and balms, myrrh and frankincense, honey, alcohol, glycerin, mercuric chloride, silver nitrate, iodine, hypochlorites, creosote, ferric chloride, zinc chloride, and carbolic acid. In 1871 Lister began to use a carbolic acid spray to reduce contamination of the operating room atmosphere, a practice he abandoned in 1887.

The “antiseptic principle” or “Listerian method” emphasized antiseptic treatment of wounds after the operation. Although initially resisted by many surgeons (more by British and American surgeons than by surgeons on the continent), they were gradually adopted.

Even late in the nineteenth century aseptic surgery was not generally practiced. Surgeons washed their hands after, but seldom before, operations. When asked what was new in surgery in 1882, Ernst Bergmann said, “Today we wash our hands before an operation.” Gloves were not worn routinely until the early part of the twentieth century. Only gradually and with much opposition was aseptic surgery adopted. Sterilization of instruments, first by chemicals and then by steam, came into practice in the 1880s and 1890s. Hand washing and the wearing of masks, caps, gowns, and gloves were also introduced about this time.

William Stewart Halsted (1852–1922) introduced rubber gloves for his scrub nurse (and future Mrs. Halsted), Caroline Hampton, because the corrosive sublimate used to sterilize instruments, mercuric chloride, irritated her skin. One of Halsted's students, Joseph Bloodgood, introduced their routine use by the entire operating team.

The introduction of antibiotics was a major step in the treatment of infections. Although the discovery of penicillin was first reported by Alexander Fleming in 1928, it was not used clinically until administered by Howard Florey in the 1940s. Penicillin was then rapidly introduced into general clinical medicine and was followed by streptomycin and numerous other antibiotics. It was hoped that antibiotics would eliminate the risk of infection as a surgical complication and would enable established infection to be cured easily, but this has not been the case. Wound infection and other postoperative infections continue to be a problem even though antibiotics have reduced their risk. The widespread use of antibiotics has even led to the emergence of strains of antibiotic-resistant bacteria. The nature of postoperative infections has also changed because of the many patients (debilitated, elderly, cancer patients) being operated on who have compromised host defenses or who are given drugs that inhibit host defenses (cancer chemotherapy agents, immunosuppressants to prevent organ transplant rejection).

It was also hoped that antibiotics would cure most infections even without operation. While the introduction of antibiotic therapy was a giant step in the treatment of nonsurgical infections, it had a much smaller impact in the treatment of surgical infections. The mortality rate of acute appendicitis was approximately 50 percent in the latter part of the nineteenth century. The recognition in the 1890s that a person with acute appendicitis required an immediate operation led to a dramatic decrease in mortality rates in the early part of the twentieth century. Until that time surgeons believed that waiting a few days would allow the omentum and intestines to isolate the appendix, after which an abscess could be safely drained. The development of intravenous fluid therapy and blood transfusion led to another decrease in mortality rates in the early part of the twentieth century. By the time penicillin became available, the mortality rate of acute appendicitis had already decreased to 5 percent. Continuing improvement in anesthesia, surgical technique, and postoperative care have also contributed to the continuing decline in mortality—as has antibiotic therapy.

Although antibiotic therapy was a monumental advance in the treatment of infections, for patients with surgical infection it constitutes only a part of the treatment. Surgical infections generally require an operative procedure (or radiology-assisted percutaneous drainage) for a successful outcome. In the future, continued improvement in the treatment outcome of surgical infection is more likely to stem from such factors as earlier and better means of diagnosis, improved patient care, and therapy directed against bacterial products or host responses than from improvements in antimicrobial therapy.

GENERAL CONSIDERATIONS

Surgical infections can conveniently be defined as infections that require operative treatment or result from operative treatment. Infections that require operative treatment include (1) necrotizing soft tissue infections; (2) body cavity infections such as peritonitis, suppurative pericarditis, and empyema; (3) confined tissue, organ, and joint infection such as abscess and septic arthritis; and (4) prosthetic device–associated infections. With the possibility of patient-to-surgeon and surgeon-to-patient spread of viral infections such as from the human immunodeficiency virus (HIV) and hepatitis viruses, infections in health care workers have also become of interest to surgeons.

Infections that result from operative treatment include wound infection, postoperative abscess, postoperative (tertiary) peritonitis, other postoperative body cavity infection, prosthetic device–related infection, and other hospital-acquired infections among which are pneumonias, urinary tract infection, and vascular catheter–related infection. Immunocompromised patients are subject to viral and fungal infections that seldom cause infection in the normal host.

Principles of Therapy

The patient's own host defenses and antibiotic therapy are adequate to overcome most infections. Nonoperative treatments can assist recovery from some infections. Chest physiotherapy is useful in patients with pneumonia, especially those with thickened secretions. Increasing fluid intake and thus increasing urine flow is helpful in patients with urinary tract infections. Immobilization and elevation can relieve pain and reduce the swelling of an extremity afflicted with cellulitis or lymphangitis.

Operative treatment is generally required when host defenses cannot function properly or when there is continuing contamination with microorganisms: infected fluid collections must be drained, infected necrotic tissue must be debrided, and infected foreign bodies must be removed. Infected fluid collections such as abscesses must be drained because phagocytic cells cannot function properly with the metabolic conditions usually present. Antibiotics are not very effective against bacteria in abscesses because they penetrate abscesses poorly and because antibiotics work best on actively dividing bacteria—and most bacteria in abscesses are not actively dividing. Drainage is also salutary because necrotic tissue and foreign bodies inhibit the proper functioning of host defenses.

Defects in the gastrointestinal tract provide a continuing source of bacteria that rapidly overwhelms host defenses. Operation is required to end this source by closing the defect in the gastrointestinal tract or by bringing the defect to the outside as an ileostomy or colostomy.

Determinants of Infection

The development of surgical infection depends on several factors: (1) microbial pathogenicity and number, (2) host defenses, (3) the local environment, and (4) surgical technique (for postoperative infection).

Microbial Pathogenicity

The ability of a microbe to cause infection is a balance between host defenses and microbial pathogenicity. Some microbes that have virtually no ability to cause infection in the normal host can cause lethal infection in an individual with compromised host defenses.

Many bacteria (*S. pneumoniae*, *Klebsiella pneumoniae*, *Streptococcus pyogenes*, *Staphylococcus aureus*, *Salmonella typhi*) and fungi (*Histoplasma capsulatum*, *Candida albicans*, *Cryptococcus neoformans*) have thick capsules that make them resistant to phagocytosis (see section [Surgical Microbiology](#) below). Other microbes (*Mycobacterium tuberculosis*, *Aspergillus flavus*, and *Toxoplasma gondii*) resist intracellular killing after they have been phagocytosed when lysosomes that contain enzymes that digest microbes do not fuse with the phagosome. Other microbes successfully resist digestion by lysosomal enzymes.

Some bacteria can elaborate toxins, many of which are enzymes that injure or kill cells or promote spread within tissues. Exotoxins play an important role in the pathogenicity of *Clostridium* species, *Staph. aureus*, and *Strep. pyogenes*. Other bacteria (*Clostridium tetani*, *Clostridium botulinum*) elaborate neurotoxins that alter normal neural transmission.

Endotoxins are lipopolysaccharide-protein complexes that are normal constituents of the cell wall of gram-negative bacteria. These molecules activate many biological pathways, including the complement and coagulation systems, and cause release of cytokines and other biologic mediators from macrophages, release of hormones, and alteration in metabolism.

Host Defenses

Local host defenses are important in preventing microbial penetration into the tissues. Systemic host defenses are needed to rid the tissues of microbes once penetration has occurred.

Local Host Defenses

Tissues are protected from microbial invasion by a layer of epithelium. The epithelium of the skin is multilayered, and the superficial layers are keratinized. The epithelium also is multilayered in the nasopharynx, oral cavity, esophagus, and genitourinary tract. At other sites (the tracheobronchial tree, gastrointestinal tract, and eye) a single layer of epithelium protects the underlying tissues. Each site also provides a local environment that is not conducive to microbial attachment and growth. Among these local environmental features may be lack of moisture (skin), the flushing action of tears and urine, cilia (trachea, bronchi), peristalsis, mucus, pH (gastrointestinal tract), and local immunity (IgA).

Systemic Host Defenses

A complex system of defense mechanisms exists throughout the body that can inactivate and kill microbial agents. These host defenses consist of phagocytic cells, the immune system, and other molecular cascades such as the complement system, the coagulation system, and the kinin system. Phagocytic cells that can ingest and kill microbes include polymorphonuclear leukocytes (PMNs) and tissue macrophages (monocytes in the blood). Through a complex set of interactions of microbes with complement and other activation molecules, PMNs adhere to vascular endothelium, migrate across the endothelium and move in the direction of the microbes (chemotaxis), attach to the microbes (which may involve immunoglobulins or other opsonins), and phagocytose the microbes. Finally, lysosomes containing a variety of enzymes fuse with the phagosome, and the microbe is rapidly digested. The initiation of this process and its attendant chemical, cellular, and physiologic changes result in inflammation.

Macrophages are phagocytic cells found throughout the body tissues: in liver (Küpf­fer cells), spleen, lymphoid tissue, lung (alveolar microphages), brain (glial cells), connective tissue (histiocytes), and pleura and peritoneum. Macrophages can also move toward microbes in response to chemotactic agents and phagocytose and kill them. In addition, macrophages are important in initiating the immune response and can elaborate cytokines, tissue necrosis factor, interferon, and other biologically active molecules. Humoral and cellular immunity are important systemic host defense mechanisms for many microbial agents. The complement system, clotting system, kinin system, leukotrienes, cytokines, and other biologically active molecules are also activated by microbial agents and play an important role in host defenses.

Host defenses are altered in malnourished individuals, trauma patients, postoperative patients, burn patients, patients with malignant neoplasms, and patients receiving drugs such as cancer chemotherapeutic agents, immunosuppressive agents to prevent transplant rejection, or steroids or other agents that have immunosuppressive effects.

Local Environmental Factors

Local factors may permit an infection to occur in a person with minimal microbial contamination and with otherwise adequate host defenses. These environmental factors inhibit systemic host defenses from being fully effective. A traumatic wound that normally would heal without infection has a greatly increased likelihood of becoming infected if the trauma has resulted in devitalization of tissue or if foreign bodies have been deposited in the wound. Phagocytic cells do not function effectively in the presence of devitalized tissue or foreign bodies. A suture can reduce the number of *Staph. aureus* required to produce a subcutaneous infection by a factor of 100,000. Fluid collections and edema also increase the likelihood of infection because they inhibit phagocytosis.

Peripheral vascular disease contributes to soft tissue infection by preventing blood and the systemic host defenses that it contains (phagocytic cells, immune globulins, and other mediators) from reaching the site of microbial contamination. Shock also decreases the amount of blood that reaches these sites.

These environmental factors can prevent phagocytic cells from functioning efficiently by lowering tissue oxygen tension (P_{O_2}). The lowered P_{O_2} inhibits function of phagocytic cells and promotes the growth of anaerobes.

Surgical Technique

Surgical technique is an important determinant of postoperative wound infection and other postoperative infections. Surgeons can decrease the likelihood of postoperative infection by handling tissues gently; removing devitalized tissues, blood, and other substances that promote the growth of microbes; using drains appropriately (and avoiding inappropriate use); avoiding excessive cautery; and not performing intestinal anastomoses under tension or when there is any question of inadequate blood supply.

TYPES OF SURGICAL INFECTIONS

Soft Tissue Infections

Infection of the soft tissues (Table 5-1)—skin, subcutaneous fat, fascia, and muscle—usually can be treated by antibiotics unless an abscess has formed or tissue necrosis has developed.

Cellulitis and Lymphangitis

Cellulitis is a spreading infection of the skin and subcutaneous tissues. There may or may not be evidence of injury to the skin. It is characterized by local pain and tenderness, edema, and erythema. Usually the border between infected and uninvolved skin is indistinct, with the region of erythema gradually fading into normal-appearing skin. Erysipelas, which is caused by *Strep. pyogenes*, is characterized by intense erythema with a sharp line of demarcation between involved and uninvolved skin. Cellulitis may be accompanied by systemic manifestations such as fever, chills, malaise, and toxic reaction.

Cellulitis can be caused by numerous bacteria in addition to *Strep. pyogenes*, such as *Staph. aureus*, *Strep. pneumoniae*, other streptococci, *Haemophilus influenzae*, and aerobic and anaerobic gram-negative bacteria. Lymphangitis, inflammation of the lymphatic channels in the subcutaneous tissues, presents as visible red streaks. Bacteria may reach the lymph nodes and cause lymphadenitis.

Cellulitis and lymphangitis can be treated by antibiotics alone, but surgery may be needed to treat the source, such as when an abscess has formed. Treatment includes immobilization and elevation to reduce pain and swelling. Failure to achieve prompt clinical response should suggest that suppuration may have occurred and that surgical drainage may be required.

Other skin infections that can be treated by local cleansing and local or systemic antibiotics include impetigo (*Staph. aureus*), erysipeloid (*Erysipelothrix rhusiopathiae [insidiosa]*), folliculitis, and furunculosis.

Some microbial factors that cause granulomatous infections produce ulcers, nodules, sinuses, or infiltrated plaques. Biopsy with culture and histologic examination of tissue with special stains may be required for such lesions. Mycobacterial and fungal infections can manifest themselves in this way.

Soft Tissue Abscess

Surgical treatment is usually required when soft tissue infection results in abscess or tissue necrosis. Furuncles and carbuncles (boils), breast abscess, and perirectal abscess require surgical incision and drainage and usually antibiotic therapy. A *carbuncle* is a subcutaneous abscess usually formed by a confluent infection of multiple contiguous hair follicles. They are most frequently found on the back of the neck and on the upper back. The most common cause is *Staph. aureus*. Overlying erythema may lead to the mistaken diagnosis of cellulitis, but the presence of a fluctuant mass usually leads to the correct diagnosis. A *felon* is a purulent collection in the distal phalanx of the fingers that causes intense pain and pressure in that compartment. Swelling may be minimal because of the fibrous bands between the skin and bone. Treatment requires incision and drainage. A lateral incision is used to avoid a painful scar on the finger top. *Breast abscess* is usually caused by *Staph. aureus* but can be due to gram- negative bacteria as well. It frequently occurs in nursing mothers. Treatment consists of incision and drainage and antibiotics (see Chap. 14). *Perirectal abscess* begins as an infection of one of the crypt glands that then extends into the perirectal space and may present subcutaneously near the anus. It is caused by aerobic and anaerobic gram-negative bacteria that are normal residents of the colon. Incision and drainage and antibiotic therapy are the appropriate initial treatment. Up to 50 percent of perirectal abscesses may result in a fistula communicating with the anal crypt and may require later treatment. The fistula may be difficult to identify because of the intense inflammation of the abscess, and it is usually best to drain the abscess rather than risk making a passage into the anus looking for a fistula where none existed previously (see Chapter 26).

Necrotizing Soft Tissue Infections

Soft tissue infections that cause necrosis are more serious because of their propensity for extensive tissue destruction and high mortality rates. The nomenclature for necrotizing soft tissue infections is confusing. Terms such as necrotizing fasciitis, streptococcal gangrene, gas gangrene, bacterial synergistic gangrene, clostridial myonecrosis, and Fournier's gangrene are commonly used. Attempts to differentiate these infections on the basis of predisposing conditions, presence of pain, toxic condition, fever, presence of crepitus, appearance of the skin and subcutaneous tissues, and presence of bullae are of little help in diagnosis or initial treatment. Bacteria seldom respect anatomic barriers, and hence necrotizing fasciitis is rarely limited to fascia and myonecrosis is frequently not limited to muscle.

Most necrotizing soft tissue infections are caused by mixed aerobic and anaerobic gram-negative and gram-positive bacteria. *Clostridium* species, of which *C. perfringens*, *C. novyi*, and *C.septicum* are the most common, cause the most dramatic infections, with rapid progression, early toxic condition, and high mortality rates. The term *gas gangrene* has become synonymous with clostridial infection. But the presence of gas in tissue simply means that anaerobic bacterial metabolism has produced insoluble gases such as hydrogen, nitrogen, and methane. Both facultative and obligate anaerobes are capable of such metabolic activity. Aerobic bacteria can also produce gas. Gas in tissues is much more likely to be caused by bacteria other than *Clostridium* species. *Strep. pyogenes* can also cause extensive tissue necrosis. Halophilic marine *Vibrio* species can cause rapidly progressive necrotizing soft tissue infections, especially in individuals with liver disease. Fungi too can cause necrotizing cutaneous and subcutaneous infection, but these infections progress much more slowly than do bacterial infections.

Necrotizing soft tissue infections must be recognized early and treated promptly. Diagnosis is not difficult when skin necrosis or bullae are present, but occasionally the clinical findings are subtle until extensive necrosis has occurred. The overlying skin may appear to be normal or afflicted only with cellulitis. Early confusion, toxic condition, and failure to respond to nonoperative therapy may be the earliest clues to the presence of a necrotizing infection. The presence of cutaneous necrosis, bullae, or crepitus strongly suggests a necrotizing infection, and surgical exploration is warranted.

Surgical treatment requires debridement of all necrotic tissue. Computed tomography is a sensitive method of detecting soft tissue gas and may allow a better appreciation of the extent of tissue necrosis. Amputation may be required for myonecrosis of the extremities. It may be difficult to evaluate fully the extent of necrosis at the initial operation, or viable tissue may become necrotic after the initial debridement. Therefore, the wound must be inspected daily with adequate (usually general) anesthesia either in the operating room or in the intensive care unit in the case of small wounds until the surgeon can be sure that there is no further necrosis. Extensive debridement, which can leave the patient with large

tissue defects and extensive wounds, may be required. The goal of treatment is to remove all necrotic tissue; the surgeon should worry about reconstruction later. Initially, broad-spectrum antibiotics including penicillin should be administered. A Gram stain of the tissue and fluid should be done to look for gram-positive rods (*Clostridium* species) or cocci (*Streptococcus* species).

The use of hyperbaric oxygen to treat necrotizing soft tissue infections is controversial. Patients are placed in a chamber at three times atmospheric pressure absolute. Hyperbaric oxygen inhibits production of alpha toxin by *Clostridium* species. Proponents of hyperbaric oxygen treatment claim that it makes the patient's condition less toxic and diminishes the amount of tissue requiring excision. Although hyperbaric oxygen can decrease mortality due to clostridial infection in experimental animals, there are no controlled clinical trials of hyperbaric oxygen in humans. Even proponents, however, note hyperbaric oxygen does not improve outcome for patients with necrotizing soft tissue infections caused by nonclostridial organisms. Hyperbaric oxygen should not be used before surgical debridement. Some centers favor oxygen therapy at atmospheric pressure because a high arterial oxygen pressure can still be achieved without requiring a hyperbaric chamber. If adequate debridement is carried out, the patient usually improves rapidly and hyperbaric oxygen usually is not required. Since clostridial infections account for only a small proportion of necrotizing soft tissue infections, the number of patients who can benefit from this form of therapy is small.

Hyperbaric oxygen therapy has possible complications. Barotrauma can cause injury to the middle ear if the eustachian tube is blocked, trauma to a sinus, pneumothorax, and air embolism. Oxygen toxicity can cause neurotoxicity resulting in reduced seizure threshold and pulmonary toxicity if treatment is prolonged. A feeling of claustrophobia and reversible visual changes are other potential problems associated with hyperbaric oxygen therapy.

Tetanus

Tetanus is caused by *Clostridium tetani*, a large gram-positive spore- forming bacillus. In recent years the number of cases in the United States has decreased sharply from over 450 in 1955 to fewer than 100 in 1975. Currently there are approximately 50 cases of tetanus reported per year.

C. tetani is usually acquired by implantation of the organisms into tissues by means of breaks in the mucosal or skin barriers. Although it is frequently said that tetanus occurs in dirty, necrotic, and neglected wounds, the majority of cases in the United States appear after punctures, lacerations, and abrasions. Tetanus can appear after surgical wounds, injections, and in patients who have no apparent injury at all. Organisms proliferate at the site of inoculation and have virtually no capacity for causing an invasive infection. Clinical tetanus is as much an intoxication as an infection.

C. tetani elaborates two toxins, tetanospasmin and tetanolysin. Tetanospasmin acts on the anterior horn cells of the spinal cord and on the brainstem. It blocks inhibitor synapses at these sites, leading to muscle spasms and hyperreflexia. These physiologic effects are similar to those of strychnine poisoning. Tetanolysin is cardiotoxic and causes hemolysis, but it is not thought to be of major clinical importance.

The median incubation period for both fatal and nonfatal cases of tetanus is 7 to 8 days. Tetanus usually appears in generalized form but occasionally appears as localized tetanus with increased muscle tone and spasms confined to muscles near the wound and without systemic signs. Neonatal tetanus is recognized as difficulty in sucking beginning at 3 to 10 days of age and progressing to generalized tetanus.

In generalized tetanus the initial symptoms are variable. Some patients have symptoms of restlessness and headache. In other patients the first symptoms are muscle spasms with vague discomfort in the neck, lumbar region, and jaws. Spasm of the pharyngeal muscles makes swallowing difficult. A stiff neck is one of the early signs. Progressively, other muscle groups become involved until the spasms become generalized. Orthotonos, opisthotonos, and emprosthotonos can develop. Generalized toxic convulsions are frequent, exhausting, and unpredictable. Any slight external stimulus (a breeze, sudden movement, noise, or light) and internal stimuli (cough, swallow, distended bladder) may trigger generalized convulsions. These convulsions may involve the laryngeal and respiratory muscles and result in fatal acute asphyxia.

Throughout these spasms, which can be extremely painful and even cause fractures, the patient remains mentally alert. The pulse is elevated and there is profuse perspiration. Fever may or may not be present.

Diagnosis of tetanus is based on the clinical picture associated with no prior history of immunization. Although laboratory studies may show an elevated white blood cell count, they are not helpful in making the diagnosis. The demonstration of gram-positive organisms in the wound does not establish the diagnosis of tetanus; failure to demonstrate that the bacillus is in the wound does not eliminate the diagnosis. Consequently, the diagnosis can be difficult in early tetanus. Even with adequate treatment, the mortality rate can exceed 50 percent.

Treatment

Patients with clinical tetanus require exquisite nursing care to avoid complications. Initially therapy consists of administration of tetanus immune globulin (TIG) 500 to 10,000 units as soon as the diagnosis is made. The precise effective dosage of TIG has not been established. Routine laboratory tests should be obtained and the patient should be monitored. Nursing care must be provided constantly in an intensive care unit setting. Formerly patients with tetanus were cared for in quiet rooms that provided a minimum of stimulation. Currently most are treated in an intensive care unit on a respirator with paralytic drugs given to prevent muscle spasms.

Mild cases can be treated with sedation, but most physicians administer muscle relaxants. Adequate doses of analgesics are required because of the pain associated with muscle spasms. Detailed attention must be given to caring for a paralyzed patient who is on a respirator. Adequate nutrition must be provided. Laxatives are generally indicated so that gastrointestinal elimination can be facilitated. A urinary catheter should be provided. Eye protection to prevent desiccation should be provided. Pressure sores can occur rapidly and must be prevented with appropriate skin protection, hygiene, and cushions. Patients may require tracheostomy if they need a respirator for a prolonged period. Roentgenographic studies are used to monitor the development of fractures. Pulmonary emboli can be a problem in patients who have minimal movement. Cardiac exhaustion and circulatory disruption can occur from sympathetic overstimulation. Hyperbaric oxygen treatment is not recommended because it is ineffective. Oxygen has no effect on the toxemia.

The wound must be treated to remove as much of the *C.tetani* and nonviable tissue as possible. Debridement of all necrotic tissue should be done. Penicillin G should be administered to treat any bacteria that remain behind, but antibiotics are no substitute for good wound care.

Prevention

Active immunization with tetanus toxoid is a safe and effective way of preventing tetanus (Table [5-2](#)). Unfortunately many children in the United States are not adequately vaccinated; immunization is also inadequate in many developing countries.

One month after the diagnosis of tetanus is made, tetanus toxoid immunization should be begun. The amount of tetanus toxin released during an infection is so small that the patient does not make antibody.

Body Cavity Infections

Peritonitis and Intraabdominal Abscess

Primary peritonitis is caused by a single organism and occurs most commonly in young children and in adults with ascites or with renal failure that is being treated by peritoneal dialysis. Primary peritonitis can be treated with antibiotics and other medical measures.

Secondary bacterial peritonitis is usually the result of a defect in the gastrointestinal tract and requires operative intervention. The goals of surgery are to control the source of contamination, to remove bacteria and adjuvant materials from the peritoneal cavity, and to prevent postoperative abscess or recurrent peritonitis. Antibiotic therapy that is effective against aerobic and anaerobic enteric bacteria has an important role in treating patients with secondary bacterial peritonitis, but it should never serve to delay or replace operative intervention. Peritonitis occurring (or persisting) after initial operation for secondary peritonitis is *persistent peritonitis*. *Tertiary peritonitis* is a peritonitis-like syndrome occurring late as a result of a disturbance in the host's immune response and is characterized by peritonitis without evidence of pathogens or peritonitis caused by fungi or low-grade pathogenic bacteria.

Percutaneous or operative drainage along with antibiotic therapy is necessary for the treatment of intraabdominal abscesses. The etiology, clinical presentation, diagnosis, and treatment of peritonitis and intraabdominal abscess are discussed in Chap. [32](#).

Empyema

Empyema is usually due to pneumonia (see Chap. [16](#)). Other causes are pulmonary infarct, septic emboli to the lung, tracheal or bronchial fistula, leaking esophageal anastomosis, hepatic abscess, subphrenic abscess, trauma, leaking bronchial closure, infected hemothorax, and paravertebral abscess.

Empyema may be encapsulated and localized or may involve the entire pleural cavity. Initially the fluid in the chest is thin, but with increasing numbers of polymorphonuclear leukocytes and fibrin deposition the fluid becomes thicker and the visceral and parietal peritoneum adhere to each other.

The clinical manifestations of empyema initially resemble those of pneumonia, with pleuritic chest pain and fever, but unresponsiveness to antibiotic therapy may suggest the diagnosis. Chronic empyema can be manifested by dyspnea, fatigue, anemia, debility, and clubbing of the fingers.

Treatment of empyema is aimed at evacuation of the empyema contents and restoration of normal pulmonary function by expansion of the lung. Most empyemas can be treated by tube thoracostomy, especially in early empyema when the fluid is thin, and antibiotic therapy. The course of the disease is observed by the patient's clinical response and chest roentgenograms. The tube may be converted to open drainage after 2 to 3 weeks when the visceral and parietal pleura have become adherent so that the lung does not collapse.

Open drainage should be used if there are multiple pus pockets, if the pus is very thick, or if the empyema is inadequately drained by tube thoracostomy. In some cases a decortication procedure may be necessary to reexpand the lung, or, if a bronchopleural fistula is present, a thoracoplasty may be required.

Other Closed-Spaced Infections

Purulence in closed spaces usually requires drainage and tetanus toxoid immunization and antibiotic therapy. If the diagnosis of septic arthritis is made promptly, antibiotic therapy alone may be sufficient to treat the infection. If the diagnosis is delayed, surgical treatment is required to preserve joint function and to eradicate the infection.

Suppurative pericarditis generally requires operative intervention. Although antibiotic therapy alone may be sufficient to treat some early infections, operative therapy is usually required once suppuration has occurred.

Prosthetic Device—Associated Infections

Infections in prosthetic devices, such as cardiac valves, pacemakers, vascular grafts, and artificial joints, are associated with great morbidity and the defeat of the goals of the operation and too frequently end with the death of the patient. Although intensive antibiotic therapy alone can occasionally cure the infection, frequently it can be eradicated only by complete removal of all foreign material and antibiotic therapy. Replacement of cardiac valves with new porcine or homograft valves and antibiotic therapy has met with some success. Vascular grafts have occasionally been salvaged without graft removal by treatment with debridement, povidone- iodine–soaked dressings, and antibiotic therapy when the suture line has not been infected. Infected prosthetic joints and pacemakers have occasionally been salvaged by antibiotic irrigation of the joint or pacemaker. Usually infected prosthetic devices require complete removal.

Hospital-Acquired (Nosocomial) Infections

Hospital-acquired infections are infections that develop within a hospital or are acquired within a hospital. These infections are costly in terms of the suffering and death they can cause as well as in terms of the cost of increased hospital stays, time lost from work, and legal liability. Each year in the United States there are an estimated 2 million hospital-acquired infections that result in 150,000 deaths. Hospital-acquired infections add an average of 1.5 days to the hospital stay of patients who develop lymphangitis, 14.8 days for patients with septicemia, and 16.6 days for patients who have infections at multiple sites. The total yearly cost of these infections is estimated at several billion dollars.

The Centers for Disease Control and Prevention (CDC) examines hospital- acquired infections through the National Nosocomial Infections Surveillance System (Table 5-3). Infection rates were greatest on the surgical service, at 44.3 per 1000 discharges. On surgical services urinary tract infections are most common, followed by wound infection, lower respiratory infection, bacteremia, and cutaneous infection. Vascular catheter–related infections are frequently classified under bacteremia or cutaneous infection.

Wound Infections Classification

For many years wounds have been classified into four categories according to the theoretical number of bacteria that contaminate wounds: clean, clean-contaminated, contaminated, and dirty (Table 5-4). Wound infection rates in large series are approximately 1.5 to 3.9 percent for clean wounds, 3.0 to 4.0 percent for clean-contaminated wounds, and approximately 8.5 percent for contaminated wounds. Dirty wounds generally are left open, but wound infection rates for dirty wounds of 28 and 40 percent have been reported (Table 5-5). With over half of all operations currently being performed as outpatient procedures and early discharge of hospitalized patients, most wound infections are recognized when the patient is outside the hospital. Because of the expense and difficulty of detecting wound infections that occur outside the hospital and because surgeons may be penalized for reporting too many wound infections, accurate wound infection data may become more difficult to collect in the future.

Wound infections encompass infections of the wound that occur above the fascia (superficial wound infection) and those that occur below the fascia (deep wound infection). Some authors have proposed more inclusive terms, e.g., “surgical field” or “surgical site infection,” that would include all operative sites potentially exposed to bacteria. These more inclusive terms would include superficial and deep wound infections and infections that do not occur in direct proximity to the surgical incision (e.g., postoperative intraabdominal abscess).

Definition of Surgical Wound Infection

An incisional (superficial) wound infection must meet the following criteria: Infection occurs at an incision site within 30 days after operation, and involves skin or subcutaneous tissue above the fascial layer, and any of the following:

- (1) There is purulent drainage from the incision or a drain located above the fascial layer.
- (2) An organism is isolated from culture of fluid that has been aseptically obtained from a wound that was closed primarily.
- (3) The wound is opened deliberately by the surgeon, unless the wound is culture-negative.

Deep surgical wound infection must meet the following criteria: Infection occurs at the operative site within 30 days after operation if no prosthesis was permanently placed and within 1year if an implant was placed, and infection involves tissues or spaces at or beneath the fascial layer, and any of the following:

- (1) The wound spontaneously dehisces or is deliberately opened by the surgeon when the patient has a fever (>38 °C) and/or there is localized pain or tenderness, unless the wound is culture-negative.
- (2) An abscess or other evidence of infection directly under the incision is seen on direct examination, during operation, or by histopathologic examination.
- (3) The surgeon diagnoses infection.

Bacteria can gain entrance to the wound from endogenous or exogenous sources. Virtually all infections in clean-contaminated and contaminated wounds and also in the majority of clean wounds are caused by endogenous bacteria present on the skin or mucosal surfaces.

Prophylaxis Operating Room Environment

Air-handling systems are designed to reduce the number of airborne microbes. Filtration of air can reduce the number of dust particles to which microbes can adsorb. Operating room air should have a positive pressure relative to air in the corridors so that unfiltered air does not enter the operating room. Special laminar flow systems with high efficiency particulate air (HEPA) filters are frequently used when prosthetic joints are implanted to reduce the likelihood of airborne contamination. Reducing the number of people in the operating room and limiting talking are also advocated by some experts to reduce the number of airborne microbes.

Instruments and Drapes

Properly sterilized instruments should never be a source of infection. If drapes become wet bacteria can move from underneath the drapes to the surgical field by capillary movement. These bacteria theoretically can then enter the wound and cause a wound infection. Disposable drapes with plastic liners and cloth drapes with tighter weaves are designed to minimize this type of bacterial contamination. It is extremely difficult to establish whether the type of drape affects wound infection rates. The choice of drapes should be based on other considerations, such as cost and ease of disposal.

Adhesive plastic drapes do not lower the incidence of wound infection. Cruse found that using adhesive plastic drapes in addition to routine draping was associated with a wound infection rate of 2.3 percent (214 of 9252) compared to 1.5 percent (405 of 26,303) when plastic drapes were not used. Skin bacteria may actually proliferate under the warm, moist environment provided by the plastic drape, and these bacteria may enter the wound if the edge of the plastic drape lifts off the wound margin. Plastic drapes can be helpful, however, in isolating potential sources of contamination such as ostomies or fistulae near the incision.

Hand Washing

Hand washing with soap and an antiseptic agent removes dirt and desquamated skin and reduces the number of microbes on the skin. Although tradition calls for scrubbing for 10 min and using two brushes, washing for 5 min and using one brush accomplishes equal reduction in skin bacterial counts. In practice many surgeons scrub for a shorter time, especially after the day's the first operation, when most dirt and desquamated skin have already been removed.

Hexachlorophene, povidone-iodine, and chlorhexidine are the antiseptics most commonly used for hand washing. Hexachlorophene has the disadvantage of acting slowly. It should be used daily to achieve maximal reduction of skin bacteria. It has been replaced in some hospitals because of its slow action and because it can be absorbed through the skin. Both povidone-iodine and chlorhexidine result in prompt reduction of skin microbes.

Gloves

Gloves should fit snugly over the fingers and hands and over the cuff of the surgical gown. Thirty percent of gloves have defects in them by the end of the operation. Surgeons are potentially exposed to infectious agents harbored by their patients when blood enters through these holes and gets onto their skin. Glove perforations are more likely to occur during long operations, during operations for trauma, and when the patient's blood loss is great. The CDC and some experts advocate wearing two pairs of gloves to reduce the likelihood of exposure to patient's blood.

Other Barriers

Caps prevent hair and skin scales (and adherent bacteria) from falling into the patient's wound, masks prevent droplets produced during speaking or coughing from entering the patient's wound, and gowns prevent desquamated skin and other particles from entering the patient's wound. There are no data that demonstrate unequivocally that wearing these barriers lowers the wound infection rate. In two recent studies the wound infection rate was found not to increase when surgeons did not wear masks. In one study surgeons did not wear masks for a 6-month period, and that period's wound infection rate was compared with that of the preceding 5 years. The wound infection rate actually fell to 1.8 percent from 5.7 percent. In another study more than 3000 patients were randomized to be operated on by surgeons wearing masks or not wearing masks. There was no difference in wound infection rate. But these barriers should still be worn if for no other reason than to prevent blood of the patient from coming in contact with members of the operating room team.

Preoperative Stay

Patients who have longer preoperative hospitalizations are more likely to develop postoperative wound infections. These patients may acquire more virulent or antibiotic-resistant hospital bacteria. Since patients who have long preoperative hospital stays are likely to have compromised host defenses, it is not surprising that they are more likely to develop wound infections.

Preoperative Shower

A shower with an antiseptic soap such as chlorhexidine or povidone-iodine can reduce the resident skin bacteria. Cruse reported that the infection rate was 1.3 percent for patients who took a preoperative shower with soap containing hexachlorophene, 2.1 percent for those who took a shower with ordinary soap, and 2.3 percent for those who did not shower.

But another study of 5536 patients found no reduction in wound infection rates in patients who had a preoperative shower with 4% chlorhexidine detergent.

Remote Infections

Remote infections can triple the rate of wound infection. Elective operations should generally be delayed until the infection has been eliminated. Areas of dermatitis are generally moist, and bacterial growth at these sites increases dramatically. Elective operations should be delayed until the dermatitis is treated, especially if the skin incision is near or through such regions.

Hair Removal

Shaving, clipping, and depilatory agents are used to remove hair. Shaving remains the most commonly used method of hair removal. But nicks and cuts caused by shaving are sites where bacteria can proliferate.

When shaving is done the night before operation, there is ample time for bacterial proliferation in any nicks or cuts, and the wound infection rate is higher than when shaving is done in the operating room immediately before operation. When hair is removed by clipping with an electric clipper, the wound infection rate can be reduced further.

Skin Preparation

Degerming of the operative site usually entails washing the site with a germicidal soap solution for 5 to 10 min followed by painting the site with an antimicrobial solution such as chlorhexidine or povidone-iodine. Painting the operation site with an alcohol solution of povidone-iodine, which can be accomplished in less than 1 min, is as effective as a 5-min scrub with povidone-iodine followed by painting with povidone-iodine solution.

Reduction of Colonic Bacteria

There are approximately 10^{10} to 10^{11} bacteria per gram feces. Colon procedures therefore potentially expose the wound to numerous bacteria. Colonic bacteria can be greatly reduced by cleansing the colon of feces. A variety of enemas or cathartics such as magnesium citrate solution or electrolyte solutions in polyethylene glycol can be used. These agents should be used before all elective colon surgery. Oral antibiotics can further reduce the number of colonic bacteria. A combination of neomycin and erythromycin base is used most commonly, but other antibiotics are also effective.

Improving Host Defenses

Any malnutrition should be corrected to restore the patient's resistance to infection toward normal. Obesity should be corrected since it is associated with an increased rate of wound infection. Weight reduction also lowers the risk of pulmonary complications. Abnormal physiologic states that result from cirrhosis, uremia, and diabetes should be corrected as far as possible. Patients with pulmonary disease should have therapy before elective surgery to optimize their pulmonary status. Patients who smoke should cease smoking before the operation. Since smoking inhibits ciliary movement, people who smoke might not be able to clear tracheal secretions as well as nonsmokers. Although it is desirable to have patients lose weight or stop smoking before elective operations, it is extremely difficult to accomplish in practice.

Surgical Technique

Every surgical incision injures tissues. Bacteria contaminate the wounds of virtually all clean-contaminated and contaminated procedures and probably of most clean operations as well. The surgeon's goal should be to make the local wound environment as unfavorable to the growth of these bacteria as possible.

The incision should be made in such a way to injure as little tissue as possible and to prevent the accumulation of agents that facilitate bacterial growth or inhibit host defense such as devitalized tissue, foreign bodies, blood, and serum. The initial skin incision should be made with a scalpel through the entire skin layer. The subcutaneous fat should then be divided with a single incision down to the fascia. This may not be possible in obese patients, but the number of scalpel passes should be kept to a minimum. It is important to begin each new pass of the scalpel in the depths of the wound so that tissue is not devitalized. Some surgeons prefer to use the laser or electrocautery for the incision. These techniques may result in less bleeding but can cause more tissue destruction. There are no definitive studies showing that one technique results in fewer wound infections than another.

The surgeon should be fastidious in ascertaining that bleeding has been stopped before closure. Blood in the incision provides a good environment for bacterial growth. The surgeon should not rely on drains to remove blood. Blood is more likely to clot and form a hematoma than to be removed by a drain.

The wound edges can become desiccated, leading to tissue necrosis at the wound margins. Desiccation can be prevented by placing moist laparotomy pads over the edges of the wound and keeping them wet. Some surgeons place antibiotics into the irrigant. There is no solid evidence that local antibiotics lessen the likelihood of infection. When the wound is closed there is a potential space where a seroma can collect. There are no definitive studies that provide data on whether subcutaneous sutures affect the risk of wound infection, but it seems advisable to place as few foreign bodies into the wound as possible.

If the surgeon is concerned about the possibility of a wound seroma such as might occur in the subcutaneous tissue of an extremely obese patient, a closed-suction drain should be used. Latex rubber (Penrose) drains should not be used because bacteria can actually enter the wound through the drain tract. Use of latex rubber drains leads to a higher wound infection rate than not using a drain.

All devitalized tissue and foreign bodies should be removed from traumatic wounds. Irrigation with saline solution can facilitate the removal of small particles, especially if the irrigation is performed with the saline solution under pressure. When complete removal of devitalized tissue and foreign bodies cannot be assured, or when the wound is heavily contaminated with bacteria, it can be left open and closed secondarily. If the wound is left open, saline-soaked gauze should be placed in the depths of the wound to keep the edges apart. There are no definitive studies that demonstrate that using antibiotic or antiseptic solutions inhibits infection or improves healing.

Prophylactic Antibiotics

Antibiotic administration can reduce the incidence of postoperative wound infection in patients having certain operations. There are certain principles that guide antibiotic prophylaxis use (Table 5-6). Prophylactic antibiotic therapy should be directed against the bacteria likely to contaminate the wound (Table 5-7). For clean operations for which antibiotic prophylaxis is appropriate, *Staph. aureus*, *Staphylococcus epidermidis*, and gram-negative enteric bacteria are the most likely bacteria to cause wound infections. Gram-negative enteric bacteria are the most likely causes of wound infection following gastroduodenal and biliary tract procedures, colorectal surgery, appendectomy, and gynecologic surgery.

The antibiotics should generally be given intravenously 30 to 60 minutes before operation so that adequate blood and tissue levels are present at the time that the skin incision is made. The antibiotic dose should be repeated if the operation lasts longer than 4 h or twice the half-life of the antibiotic or if blood loss has been great. With many operations currently being performed on patients who are not in a hospital before surgery and with newer antibiotics available, oral antibiotic prophylaxis may also be suitable. Bowel preparation with cleansing of the colon, and oral antibiotics (e.g., neomycin and erythromycin base), can be given to nonhospitalized patients. Prophylactic antibiotics should not be continued beyond the day of operation. The most commonly violated principle is giving the antibiotic longer than is actually needed, which increases costs and the likelihood of antibiotic resistance among hospital strains of bacteria.

Cephalosporins are the most commonly used antibiotics for prophylaxis because of their broad antibacterial spectrum, which provides activity against gram-positive pyogenic cocci and gram-negative enteric bacteria (some cephalosporins) and because of their low toxicity. But despite their safety profile, allergic reactions can occur with these antibiotics, so they should not be used indiscriminately. Cefazolin, a first-generation cephalosporin, is an effective antibiotic prophylaxis for indicated clean gastroduodenal, biliary tract, and head and neck operations and traumatic wounds. Vancomycin can be substituted in hospitals where methicillin-resistant *Staph. aureus* or *S. epidermidis* is a problem and in patients who are allergic to penicillins or cephalosporins. For colorectal procedures oral neomycin plus erythromycin base and/or cefoxitin or cefotetan provide effective coverage.

First- or second-generation cephalosporins provide effective prophylaxis for gynecologic surgery and cesarean section. Third-generation cephalosporins are no more effective than first- or second-generation agents and are more expensive. Many other antibiotic classes are also effective, but none has gained the popularity of the cephalosporins.

Indications

Prophylactic antibiotics are indicated when bacterial contamination of the wound is likely or for patients having clean operations in which a prosthetic device is placed where infection could lead to disastrous results, such as in a cardiac valve, a vascular graft, or a prosthetic joint. Bacterial contamination is likely in traumatic wounds, when the intestinal tract has been entered as a result of trauma, in elective operations on the intestine or colon, in gastroduodenal operations in which the patient has increased gastric flora, in high-risk biliary tract operations, and in gynecologic operations. Studies indicate that prophylactic antibiotics can lower the incidence of all infectious complications in clean surgery (hernia and breast surgery) but the incidence of wound infection is not reduced.

The bacteria in the stomach are increased in patients who have gastric outlet obstruction, decreased gastric acidity (achlorhydria, antacid or H₂-receptor blocker therapy, gastric cancer), and normal or high acidity if bleeding has occurred. High-risk biliary tract operations include the presence of jaundice, bile duct obstruction, stones in the common bile duct, reoperative biliary tract operation, acute cholecystitis, and age greater than 70 years.

Infection Surveillance

An infection surveillance program can help to reduce the rate of wound infections and is required by the Joint Commission on Accreditation of Health Organizations. Large studies have shown the usefulness of regular wound surveillance. The introduction of a good wound surveillance program lowered the wound infection rate of more than 20,000 wounds to 1.9 percent from 4.9 percent over a 5-year period.

Other Hospital-Acquired Infections Urinary Tract Infection

Urinary tract infection accounts for 40 percent of hospital-acquired infections. Two-thirds of patients with hospital-acquired urinary tract infection have had operation on the lower urinary tract, instrumentation of the bladder, or catheterization. Because of the large number of patients who fit one of these categories there are an estimated 400,000 urinary tract infections per year in hospitalized patients. Catheter-associated urinary infections cause bacteremia in 2 to 4 percent of patients and are associated with a case fatality rate three times as high as that of nonbacteremic patients.

Bacteriuria occurs in 1 to 5 percent of patients after a single short-term catheterization. The risk of infection is higher in pregnant patients, in elderly or debilitated patients, and in patients with urologic abnormalities. The risk of bacteriuria in patients with long-term indwelling catheters is approximately 5 to 10 percent for each day the catheter is in place. Therefore, urinary catheters should be placed only when necessary and should be removed as soon as possible. If prolonged urinary tract catheterization is required, as in

comatose patients, incontinent men, or spine-injured patients, suprapubic or condom catheters can be used to reduce the risk of infection.

Appropriate catheter insertion after careful cleansing of the urethral meatus and postinsertion care of the catheter can reduce the risk of infection. Irrigation of the catheter (i.e., to remove blood clots) should be done only by trained personnel.

Lower Respiratory Tract Infection

Lower respiratory tract infections are the third most common hospital- acquired infection, according to the National Nosocomial Infections Surveillance System. Anesthesia, operations on the head and neck, and postoperative endotracheal intubation interfere with the normal protective cough reflex and may permit aspiration of contaminated material. Pain associated with thoracic or upper abdominal operations and trauma interferes with coughing and deep breathing, which promotes collection of material in the tracheobronchial tree and atelectasis, which in turn predispose to infection. Pulmonary edema or adult respiratory distress syndrome resulting from injudicious use of intravenous fluids, cardiac failure, trauma, sepsis, renal failure, or inhalation of hot gases by burn patients also predispose to pulmonary infection. Fluid that accumulates in alveoli inhibits the phagocytic capacity of pulmonary macrophages.

Hospitalized patients may have gram-negative bacteria as part of their oral flora. These bacteria may be aspirated into the lungs during the postoperative period. Tracheostomies and respiratory care devices also predispose to the entry of bacteria into the lower respiratory tract.

The most common causative organisms of lower respiratory tract infection in hospitalized patients are *Staph. aureus*, *Pseudomonas aeruginosa*, *Klebsiella* species, *Escherichia coli*, and *Enterobacter* species. These bacteria, especially in the intensive care unit setting, may be resistant to commonly used antibiotics. Pulmonary infections in hospitalized patients are frequently not associated with sputum production, and culture of sputum specimens may not reflect the cause of pneumonia. Specially protected specimen swabs can be introduced into the lungs through a flexible bronchoscope, with sensitivity rates for diagnosing the pneumonia between 70 and 90 percent. Accuracy is enhanced by using quantitative bacteriology to distinguish better colonization and invasive infection; finding more than 10³ colony-forming-units/mL is indicative of invasive infection. Bronchoalveolar lavage has increased the accuracy of bronchoscopic diagnosis.

Lower respiratory tract infections are common in intubated patients in intensive care units, occurring in as many as 20 to 25 percent of patients, with a mortality rate of 50 percent. Many of these pneumonias are attributed to low levels of aspiration. Oropharyngeal decontamination using topical, nonabsorbable antibiotics or antiseptics can reduce tracheobronchial contamination by gram-negative bacteria and pneumonia in these patients.

Vascular Catheter–Related Infection

The incidence of vascular catheter–related infections has increased greatly with the increased use of central vascular catheters that are left in place for prolonged periods in patients with compromised host defenses. It is estimated that 20 million hospitalized patients each year undergo vascular catheterization of some sort. An estimated 20,000 to 50,000 cases of hospital-acquired bacteremia per year are caused by vascular catheters. Central venous catheters have a higher infection rate than peripheral venous catheters, and polyethylene catheters have a higher infection rate than Silastic catheters. The most common source of catheter sepsis is believed to be microorganisms at the skin exit site that follow the catheter into the vein rather than microorganisms originating from a distant site that colonize the catheter via the bloodstream.

Staph. aureus and *S. epidermidis* usually originate from the skin and cause most catheter–related infections. Most yeast vascular-access infections result from hematogenous dissemination from another site. Gram-negative enteric bacteria also may infect catheters hematogenously.

The duration of catheterization, the number of catheter manipulations, inexperience of the inserter, violations of aseptic technique, and use of multilumen catheters are all associated with an increased risk of infection. Transparent plastic dressings increase the risk of infection two- to fourfold compared with traditional gauze dressings. Teams specially designated for catheter insertion and maintenance can reduce the risk of catheter infection. Skin preparation with chlorhexidine gluconate and use of topical antibiotics can also reduce the risk of infection. There are no data proving that practices such as changing catheters at intervals, changing infusion tubing every 24 to 48 h, and using in-line filters reduce the risk of infection.

Finding bacteria on the catheter tip does not establish that it was infected, since the bacteria may have come from the blood that inevitably adheres to catheter tips. Some investigators have attempted to do semiquantitative cultures of the catheter tips by rolling them on solid culture plates, flushing sterile broth through the catheter, or subjecting them to sonication to remove adherent organisms before quantitative culture as methods of diagnosing catheter–related infection. If there is unequivocal purulent discharge around the catheter insertion site, the diagnosis of vascular- catheter infection can be made without a positive culture.

Any evidence of phlebitis or cellulitis or any suspicion of septic complications caused by intravenous cannulas should lead to prompt removal of the cannulas. Because many central venous catheters are used in compromised hosts who are prone to fever, these catheters generally should not be removed because of fever alone until other potential sources of fever have been eliminated. Most surgeons remove central venous catheters that are suspected of being infected and replace them after 24 to 48 h. Frequently, central venous catheters are required, and the surgeon does not have the luxury of being able to wait before inserting a new catheter. When an infected catheter is removed and another central venous catheter is immediately inserted at the same site, infection of the new catheter usually does not occur.

Catheter infections caused by *S. epidermidis* can occasionally be treated with antibiotics alone or by removal of the catheter. If antibiotics are used, a short course (3 to 7 days) is recommended. Vascular-access infections caused by *Staph. aureus* always require antibiotic therapy. The controversy in the treatment of catheter infections caused by *Staph. aureus* is in the duration of antibiotic therapy, with most experts recommending a 2- to 3-week course. Vascular-catheter infection caused by yeasts should always be treated by catheter removal and administration of an antifungal agent if cultures remain positive following removal or if there is infection elsewhere.

SURGICAL MICROBIOLOGY

Surgical infections are usually caused by bacteria, but fungal and viral infection can also occur, especially as postoperative infections in immunocompromised hosts. Most bacterial infections are caused by organisms that are part of the patient's endogenous flora—bacteria that are normal residents of the skin or gastrointestinal tract.

Bacteria

Bacteria can be classified according to staining characteristics with Gram stain (positive or negative), shape (cocci, rods, spirals), and ability to grow without oxygen (aerobic, facultative, anaerobic), or according to a combination of these characteristics. Gram-positive cocci, gram-negative aerobic and facultative rods, and anaerobic bacteria are three groups into which most bacteria causing surgical infections can be placed.

Gram-Positive Cocci

Staphylococcus and *Streptococcus* species are the gram-positive cocci (often referred to as pyogenic) of interest to surgeons because of their ability to cause primary surgical infections and postoperative infections.

The genus *Staphylococcus* is composed of facultatively anaerobic gram- positive cocci that are found on moist areas of the body, the anterior nares, and mucous membranes. In addition, these bacteria can be found on the body surface of many species of mammals and birds, in the air and dust of occupied buildings, and in milk, food, and sewage.

Staph. aureus is the most common pathogen isolated from wound infections. A major factor in its pathogenicity is coagulase production, although the mechanism whereby coagulase production increases virulence is not known. In addition to coagulase production, a variety of other cell surface components and extracellular products are related to pathogenicity.

Cell wall peptidoglycan inhibits edema production and migration of leukocytes, allowing bacteria to proliferate in tissues. Capsules inhibit opsonization and thus phagocytosis. Some strains produce a surface- associated exopolysaccharide or glycocalyx (slime), which is associated with virulence, probably by permitting the bacteria to resist phagocytosis and adhere to prosthetic materials.

Other extracellular products also contribute to the pathogenicity of *Staph. aureus*. An enterotoxin is responsible for food poisoning. Epidermolytic toxin can cause a variety of skin lesions; the most characteristic are the diffuse exfoliative bullae seen in children with the scalded skin syndrome. Another exotoxin, TSS toxin-1, is responsible for toxic shock syndrome. Other extracellular products make *Staph. aureus* resistant to H₂O₂- mediated intracellular killing (catalase) and cause cell death (leukocidin, alpha toxin, beta toxin).

Staph. epidermidis, a member of the flora of the skin and mucous membranes, was long thought to be a commensal. Although not as pathogenic as *Staph. aureus*, *S. epidermidis* causes infection in the presence of foreign bodies such as plastic catheters, ventricular shunts, and prosthetic joints and heart valves. *Staph. aureus* and *S. epidermidis* are important surgical pathogens. *Staph. aureus* is a major cause of wound infection. It can cause infection of skin and soft tissues and abscesses of these and other structures. Bacteremia can lead to infection of heart valves and other deep structures such as bone, kidney, and brain.

Surgically important members of the genus *Streptococcus* include *S. pyogenes*, *S. pneumoniae*, and the viridans group, which includes *S. mutans*, *S. mitior*, *S. salivarius*, *S. sanguis*, and *S. milleri*. Streptococci are classified according to Lancefield classification, which is based on cell surface antigens, and according to their ability to cause hemolysis on blood agar: in alpha hemolysis, a zone of green discoloration is seen around colonies containing intact red blood cells; in beta hemolysis, there is complete clearing of the area around colonies and destruction of red blood cells; and in gamma hemolysis, there is no hemolysis. *Strep. pyogenes* is Lancefield group A and beta-hemolytic. Group A streptococci have cell surface components and extracellular products that inhibit host defenses or promote spread of the bacterium. The cell surface M protein and the capsule help streptococci resist phagocytosis. Hyaluronidase and streptokinase promote the spread of infection. Streptolysin O and streptolysin S are hemolysins. Streptococcal proteinase may be responsible for tissue invasion. Pyogenic exotoxins share many properties with endotoxins from gram-negative bacteria.

Group A streptococci can cause infection of almost any organ, although the skin, subcutaneous tissues, and pharynx are by far the most frequent sites. *Strep. pyogenes* can cause pharyngitis and result in scarlet fever or rheumatic fever. Erysipelas is streptococcal cellulitis and lymphangitis and is a spreading infection with sharp, irregular, red borders. Erythrogenic toxin produced by streptococci is responsible for the intense cutaneous erythema but is not found in all streptococcal infections.

Streptococci are important pathogens because of their ability to cause postoperative infections, including cellulitis, wound infection, endocarditis, urinary tract infection, and bacteremia. These bacteria can also cause primary necrotizing soft tissue infections and abscesses. In some hospitals streptococci other than group A streptococci are the principal

streptococcal pathogens. *Strep. pyogenes* is currently an uncommon cause of necrotizing soft tissue infections. In the nineteenth century and the early part of the twentieth century streptococci were believed to be the most common cause of necrotizing soft tissue infection.

Enterococcus faecalis, *E. faecium*, and *E. durans* were formerly classified as members of the genus *Streptococcus*, but a separate genus is now recognized. They are part of the normal flora of the gastrointestinal tract and vagina. They are commonly found in patients with peritoneal and pelvic infections as part of the mixed flora typical of these infections. Enterococcal bacteremia has a poor prognosis when associated with intraabdominal or pelvic infection and is found most often in patients who have been hospitalized for a long time. These bacteria are an important problem in intensive care units because of their resistance to antibiotics.

Aerobic and Facultatively Anaerobic Gram-Negative Bacilli

There are numerous gram-negative rods that can cause human disease, but relatively few are of surgical significance. Their cell walls have common chemical constituents, most prominent of which is lipopolysaccharide or endotoxin, which is responsible for most of the biologic effects of these bacteria. Some genera also have capsules. Most are members of the family Enterobacteriaceae that are inhabitants of the gastrointestinal tract. The genera *Escherichia*, *Klebsiella*, *Proteus*, *Enterobacter*, *Serratia*, and *Providencia* frequently can be cultured from patients with intraabdominal and pelvic peritonitis and abscess, postoperative wound infection, pneumonia, and urinary tract infection.

The family Vibrionaceae includes *Vibrio* among its genera. Some *Vibrio* species are found in marine water and can cause bacteremia and necrotizing soft tissue infections in susceptible hosts, usually those with hepatic disease. They can be found in seafood and can cause bacteremia and death if the uncooked seafood is ingested.

The family Pseudomonadaceae is composed of obligate aerobes that lack the ability to ferment sugars, unlike members of the Enterobacteriaceae. *Pseudomonas aeruginosa* is the species in this family responsible for most surgical infections. They cause infections similar to those of gram-negative enteric bacteria in association with gastrointestinal disease, pneumonia, urinary tract infection, and burns. They are frequently found in immunologically compromised patients, especially if they have been hospitalized for some time. They cause necrotizing infections, especially pneumonia and vasculitis. Ecthyma gangrenosum is the cutaneous manifestation of necrotizing vasculitis due to *Pseudomonas* bacteremia and is characterized by small, round, necrotic skin lesions. Because of its resistance to antibiotic therapy, *Pseudomonas* infections are frequently treated with a combination of two antibiotics.

Anaerobic Bacteria

Anaerobic bacteria require reduced oxygen tension for growth. They are found predominantly in the mouth, vagina, and gastrointestinal tract, where they greatly outnumber the aerobic bacteria. Anaerobic bacteria, which are pathogenic, can tolerate an initial exposure of up to 3% oxygen. Virtually all anaerobic infections arise endogenously. A low oxidation- reduction potential is common to all anaerobic infection. Vascular disease, cold, shock, edema, trauma, devitalized tissue, operation, foreign bodies, malignant disease, and growth of aerobic microorganisms can lower the oxidation-reduction potential and predispose to infection with these organisms.

In most infections with anaerobic bacteria, facultative or aerobic bacteria are also present. Aerobic or facultative bacteria make conditions favorable for anaerobic bacteria by lowering the oxidation-reduction potential. The aerobic bacteria may also supply a growth factor necessary for another organism or may interfere with local or systemic host resistance.

Anaerobes such as the *Bacteroides fragilis* group have an endotoxin, but it differs chemically from the endotoxin of the enteric facultative or aerobic gram-negative bacilli, and it exhibits poor biologic activity. The cell wall of anaerobic bacteria is important in abscess formation.

The genus *Clostridium* is the most virulent of all anaerobes. *Clostridium*, which can be found in soil and stool, can cause necrotizing soft tissue infection. Clostridia produce exotoxins that have biologic effects in cell culture. Their precise role in clinical disease is unclear, but the exotoxins produced by these bacteria are believed to be responsible for most of the local and systemic manifestations. *C. perfringens*, *C. septicum*, and *C. novyi*, which can cause necrotizing infections, produce toxins that can destroy cell membranes and lyse red blood cells, collagenase, hyaluronidase, and other enzyme toxins that enhance spread of the infection through the tissues.

C. perfringens and *C. difficile* both produce an enterotoxin. *C. difficile* causes pseudomembranous colitis and occurs in patients treated with broad-spectrum antibiotics. It produces a cytotoxin that is cytopathic for almost all tissue culture cell lines. *C. tetani* and *C. botulinum* produce neurotoxins that cause muscle spasms and paralysis, respectively.

In the colon the ratio of anaerobic bacteria to aerobic bacteria is between 300:1 and 1000:1. The most common pathogens in the colon are members of the genera *Bacteroides*, *Fusobacterium*, and *Peptostreptococcus*. Of these, *Bacteroides* is the most commonly cultured genus in patients with intraabdominal infections. The *Bacteroides fragilis* group, composed of *B. fragilis*, *B. thetaiotaomicron*, *B. distasonis*, *B. ovatus*, and *B. vulgatus*, accounts for most infections with this genus. Colonic anaerobes almost never cause infections by themselves but only as part of a mixed flora, often with facultative enteric gram-negative bacilli.

Fungi

Together with algae and protozoa, fungi are classified as protists, the most primitive eukaryotic organisms. They grow as single-celled organisms, as yeasts, or as long, branching filaments known as hyphae. Their cell walls show little similarity to those of bacteria, but they have much in common with mammalian cells. Because of this and other structural and biochemical similarities to mammalian cells (both are eukaryotic cells), they are not sensitive to antibacterial agents and many antifungal agents are toxic to human cells.

Fungi can be grouped as primary pathogens, which can cause disease in individuals with intact host defenses, and opportunists, which cause disease in patients with compromised host defenses. Among the primary pathogens are *Histoplasma*, *Coccidioides*, and *Blastomyces*. *Candida*, *Cryptococcus*, *Aspergillus*, and the phycomycetes (*Mucor*, *Absidia*, and *Rhizopus*) cause most of the opportunistic infections.

In surgical patients opportunists cause most infections. *Candida albicans* and other *Candida* species are by far the most common. They cause infections in patients being treated with broad-spectrum antibiotics and in those receiving steroids and other immunosuppressive agents, in malnourished patients, in patients with malignant neoplasms, and in other compromised hosts. In these patients they can cause vascular catheter-related infections, bacteremia, intraabdominal infection, pneumonia, and urinary tract infection. These infections can be treated by stopping antibiotic administration, correcting host defenses, and therapy with amphotericin B or one of the azole antifungal agents.

Viruses

Viruses are distinguished by their small size, by their being obligate intracellular parasites, and by their having either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) but not both. Members of the herpesvirus family, especially *cytomegalovirus* (CMV), herpes simplex virus, varicella- zoster virus, and Epstein-Barr virus, can cause infections in immunosuppressed patients such as organ transplant recipients.

CMV causes most viral infections in organ transplant recipients. In these patients CMV can cause ulcerative lesions of the gastrointestinal tract leading to bleeding or perforation for which operations might be required. Epstein-Barr virus is implicated as the cause of a polyclonal B-cell lymphoma in transplant recipients.

Hepatitis B virus, hepatitis C virus, and human immunodeficiency virus (HIV) are of importance to surgeons because of the possibility that they can become infected from patient exposure and that patients can potentially be infected by physicians who harbor these viruses. Hepatitis B prophylaxis is available should a health care worker sustain a percutaneous or permucosal exposure (Table 5-8).

Human Immunodeficiency Virus

An apparently new disease was first reported in December 1981 with the description of opportunist infections and Kaposi's sarcoma occurring in homosexual men. These men also had a profound depletion of T lymphocytes. Human immunodeficiency virus (HIV), the cause of acquired immunodeficiency syndrome (AIDS), was isolated in 1983. Since 1983, more than 600,000 individuals in the United States have developed AIDS.

HIV is a retrovirus of the lentivirus family. It is an RNA virus with a cylindrical core containing RNA, the RNA-dependent DNA polymerase (reverse transcriptase), and core proteins. The core is surrounded by a viral envelope derived from the nuclear membrane of the host cells. A glycoprotein (GP-120) on the envelope has an affinity for the CD4+ receptor on T lymphocytes. After binding of the GP-120 to the CD4+ receptor on helper/inducer T cells, the virus is internalized and uncoated. The reverse transcriptase synthesizes DNA complementary to viral RNA. This DNA is incorporated into the host genome, leading to a lifelong infection.

Infected CD4+ cells are not able to carry out their normal immune functions, which leads to opportunist infections and the development of Kaposi's sarcoma. The development of opportunist infections and tumors (Kaposi's sarcoma and lymphomas) is accompanied by a decrease in the number of T cells to less than 200/mm³. The most recent definition of AIDS includes all patients infected with HIV who have a CD4+ count less than 200 cells/mm³.

Epidemiology

The CDC does not collect data on the number of individuals infected with HIV but estimates that for every person with AIDS there are approximately eight persons with HIV infection who have not yet developed clinical AIDS. There are approximately 5 million people infected with HIV in the United States. Approximately 30.6 million people are infected with HIV worldwide.

HIV has been isolated from blood, semen, saliva, tears, vaginal secretions, alveolar fluid, cerebrospinal fluid, breast milk, synovial fluid, and amniotic fluid. Only blood and blood products, semen, vaginal secretions, and breast milk have been linked to transmission.

The groups at highest risk for HIV infection are (1) homosexual and bisexual men, (2) intravenous drug abusers, (3) persons with hemophilia and other coagulation disorders, (4) heterosexual contacts of the individuals in the three previous categories, and (5) children born to HIV- positive mothers. Recipients of transfusions of blood and blood products from HIV-positive donors have approximately a 95 percent chance of developing HIV infection. The CDC has estimated that the number of cases of transfusion-acquired AIDS could eventually reach 12,000. Since testing blood donors for evidence of HIV became mandatory in 1985, transfusion- acquired HIV infection has been virtually eliminated. The current risk of transmission of HIV by screened blood in the United States is estimated to be between 1 in 450,000 and 1 in 660,000.

HIV seroprevalence varies greatly, depending on the specific population studied, the location of the population, sex, race, and ethnic origin, and year of study. The lowest HIV seropositive rates are found among blood donors (0.0041 percent for repeat female donors, 0.0189 percent for repeat male donors) and are highest among hemophiliacs (50 to 100

percent positive), intravenous drug users (20 to 60 percent positive) homosexual and bisexual men (30 to 60 percent positive).

Relatively few seroprevalence studies have been done among hospital or emergency room patients. Seroprevalence rates vary widely with the type of hospital studied, ranging from 0.24 percent among 26,275 patients in the CDC sentinel hospital study to 9.1 percent at St. Paul–Ramsey Medical Center. One can expect that hospital type, location, and specific hospital population will greatly affect HIV-seropositivity rates. According to a later CDC sentinel hospital study of 89,547 blood specimens from hospitalized patients, the overall seropositivity was 1.3 percent, but the HIV- seropositive rate of individual hospitals ranged from 0.1 to 7.8 percent.

Serologic Events

Patients infected with HIV develop viremia accompanied by a generalized lymphadenopathy, fever, and malaise. Approximately 6 to 12 weeks after infection, antibody to HIV develops. During this time the viral titer in blood decreases markedly from 10⁴/mL to 10 to 100/mL. A low virus titer persists until the patient develops AIDS approximately 7 to 9 years after infection. When AIDS develops, the virus titer rapidly increases to a level of 10⁴/mL. Serologic testing examines antibody to HIV, and seroconversion usually occurs within 12 weeks of infection but has been known to take as long as 6 months. During this period (“the window”) it is possible for patients to have circulating virus and to be potentially infectious to those around them and yet test negative for HIV. Other techniques, measuring viral antigen or using the polymerase chain reaction to look for viral nucleic acid, can detect HIV earlier.

Surgery in HIV-Infected Patients

While surgeons are not the primary caregivers for individuals with HIV infection and AIDS, patients with this infection may require operation for unrelated reasons, for diagnosis of an infection, or for treatment of surgical complications of AIDS. Patients with HIV infection and AIDS generally do not require any extra preoperative preparation. Malnutrition associated with HIV infection may require correction if time permits. Perioperative antimicrobial therapy is given for the same indications as for patients without HIV infection. These patients generally do not have difficulty with wound healing and do not have a higher rate of wound infections or other postoperative hospital-acquired infections. Drains and open wounds require precaution to avoid contamination with HIV-infected blood and other body fluids.

Patients with HIV infection may require surgery for the same reasons that anyone else might need an operation, or for problems related to their viral illness. These problems include peritonitis caused by bowel perforation, which occurs as a result of *CMV* infection; gastrointestinal obstruction as a result of Kaposi's sarcoma or lymphoma of the gastrointestinal tract; gastrointestinal hemorrhage due to *CMV*, lymphoma, or Kaposi's sarcoma; and (4) intraabdominal or retroperitoneal infection by mycobacterial and other opportunistic organisms.

HIV and AIDS in Health Care Workers

Over 7250 *health care workers (HCWs)* with AIDS have been reported to the CDC. There may be more than 50,000 health care workers infected with HIV, based on an estimate of eight persons infected with HIV for every person with AIDS. As of June, 1996 there have been 51 documented cases of HCWs who have acquired AIDS occupationally. There are 108 cases of possible occupationally acquired AIDS.

HCWs with AIDS comprise approximately 5 percent of the total number of patients with AIDS in the United States. HCWs as a whole comprise approximately 5.7 percent of the labor force in the United States. Most HCWs with AIDS are homosexual or bisexual men (Table [5-9](#)). Significantly fewer, however, are intravenous drug abusers. A relatively large “undetermined” category accounts for 5.9 percent of *HCWs* with AIDS; one-half of these patients can be classified under one of the other risk factors, one-fourth are dead or refused to be interviewed, and the remaining one-fourth are still being investigated.

As of June 30, 1997, 166 *HCWs* had developed HIV infection as a result of occupational exposure, most as a result of exposure to blood from HIV- infected patients. Of these, 52 had the occupationally acquired infection confirmed by being seronegative at the time of exposure and becoming seropositive after exposure. The remaining 114 *HCWs* did not have a blood specimen obtained at the time of exposure, but other risk factors were excluded. Most of the infected HCWs are nurses or technicians, and six are surgeons. There are probably other *HCWs* with occupationally acquired HIV infection who are not included in the CDC data set.

Risk of HIV Seroconversion in Health Care Workers

Many prospective studies have examined the actual risk of *HCWs* becoming infected with HIV after sustaining a percutaneous exposure to blood or blood-containing body fluids from patients with HIV infection. Of 1,948 *HCWs* in 12 reports who sustained a total of 1,051 mucous membrane exposures to blood or blood-containing body fluids from HIV-infected patients, six (0.29 percent per exposure) seroconverted. Risk of HIV infection is associated with deep injury, visible blood on the device, procedure involving a needle placed directly in a vein or artery, terminal illness in the source patient, and no postexposure use of zidovudine (AZT).

Surgeons are frequently exposed to patient's blood and other body fluids. Most exposures are to the skin, and their numbers can be minimized by wearing two pairs of gloves and face shields. Survey studies show that percutaneous injuries occur in 5.6 percent of operations, and 86 percent of surgeons report at least one percutaneous injury per year.

Prevention of Blood-Borne Infections in Health Care Workers

Beginning in 1983 the CDC began issuing guidelines designed to minimize the risk of transmission of HIV in the health care setting. In 1987 the CDC issued new guidelines, which have come to be called “universal precautions” (Table [5-10](#)). These guidelines have been updated and extended but not substantially altered. They are applicable to clinical and laboratory staffs, emergency service personnel, and health care workers performing invasive procedures as well as those who are not included in direct patient care (e.g., housekeeping personnel, kitchen staff, and laundry workers). Although universal precautions were issued to reduce the transmission of HIV in health care settings, they are also appropriate for reducing the transmission of other blood-borne viruses, including *hepatitis B virus (HBV)*, *hepatitis C virus (HCV)*, and the recently described *hepatitis G virus (HGV)*.

The intent of the CDC guidelines is that all patients should be regarded as potentially harboring blood-borne pathogens, because the medical history, physical examination, and laboratory testing cannot identify all patients infected with HIV or other blood-borne pathogens and because in emergencies there may be no time for testing patients. Since all patients should be treated alike—i.e., as if they potentially have a blood-borne infection—there is no need for testing patients; testing would not alter health care worker behavior.

Compliance with universal precautions has been examined in the emergency room and hospital environment. Compliance in a large inner-city hospital emergency room was found to be only 18 percent, and it fell to 5 percent if the patient was bleeding from an external injury. The rates of noncompliance with universal precautions are reported to be 74 percent in the surgical intensive care unit and 34 percent on the surgical wards. The noncompliance rate fell to 43 percent in the intensive care unit after an educational program about universal precautions, but it did not change on the surgical wards. Wong and associates found, however, that frequency of use of barrier precautions increased from 54 to 73 percent and blood exposures decreased after universal precautions were put into effect. Although the CDC can only suggest guidelines and has no regulatory authority, the Occupational Safety and Health Administration has made the CDC guidelines mandatory. Failure to adhere to universal precautions is not acceptable and is subject to sanctions.

Testing Patients for Blood-Borne Pathogens

The CDC does not recommend routine HIV testing of all patients. HIV testing of patients is recommended for management of health care workers who sustain parenteral or mucous membrane exposure to blood or other body fluids from a patient, for patient diagnosis and management, and for counseling associated with efforts to prevent and control HIV transmission in the community.

If hospitals, physicians, or health care agencies choose to perform HIV testing, the CDC advocates certain principles: (1) obtain patient consent for testing; (2) inform patients of results and provide counseling for seropositive patients; (3) ensure confidentiality; (4) ensure that seropositive patients will not receive compromised care; and (5) prospectively evaluate the efficacy of the program in reducing the incidence of exposure of health care workers to blood or other body fluids of patients who are infected with HIV. Most states have laws regulating testing of patients for HIV. Many of these laws require written informed consent of the patient with pretest and posttest counseling for both HIV-positive and HIV-negative patients.

Management of Health Care Workers Exposed to Patients' Blood and Other Body Fluids

The CDC and others have issued recommendations for the management of health care workers exposed to patient blood and other body fluids. Hospitals, physicians' offices, and other employers of health care workers should establish a systematic approach for managing adverse exposures that is consistent with CDC and Department of Labor guidelines and state laws. The Department of Labor and the CDC have published detailed employer responsibilities in protecting workers from acquisition of blood- borne diseases in the workplace. Employers should develop standard operating procedures for all activities having the potential for exposure and should provide an initial and periodic workers' education programs.

The CDC recommends that if an exposure occurs, a blood sample should be drawn after consent is obtained from the individual from whom the exposure occurred and tested for hepatitis B surface antigen (HB sAg) and antibody to HIV. Now that a serologic test is available for *HCV* the patient should also be tested for that virus (and probably also for *HGV* when testing becomes available). Local laws regarding consent for testing source individuals should be followed. Policies should be available for testing source individuals when consent cannot be obtained (e.g., an unconscious patient). Pretest counseling, posttest counseling, and referral for treatment, if appropriate, of the source individual should be provided.

HIV Postexposure Management

If a health care worker is exposed percutaneously or by a splash to the eye or mucous membrane from a patient who has HIV infection or AIDS or who refuses to be tested, the worker should be counseled regarding the risk of infection and be evaluated clinically and serologically for evidence of HIV infection as soon as possible after the exposure. The worker should be advised to report and seek medical evaluation for any acute febrile illness that occurs within 12 weeks after exposure. Following the initial test at the time of exposure, seronegative workers should be retested 6weeks, 12 weeks and 6 months after exposure to determine whether transmission has occurred. During this period the worker should refrain from blood or semen donation and should use appropriate protection during sexual intercourse. If the source individual is found to be seronegative, baseline testing of the exposed worker with follow-up 12 weeks later may be performed if desired or recommended by the health care provider.

AZT is used to treat patients with HIV infection and has been proposed as chemoprophylaxis to prevent occupational infection in health care workers. Postexposure AZT use by *HCWs* is associated with a lower risk of HIV transmission. The CDC now recommends that HCWs exposed to blood from HIV-infected individuals be treated with AZT and lamivudine (3TC). If the exposure is high-risk (a large volume of blood containing a high titer of HIV), the protease inhibitor indinavir should also be given. Prophylaxis should be given within 1 to 2 h of exposure. If the HIV status of the source patient is unknown, the use of postexposure prophylaxis should be decided on a case-by-case basis. A dilemma arises when the source individual refuses to be tested; some states permit testing blood specimens obtained for another purpose if a health care worker has been exposed to a patient's blood or other body fluid and the patient refuses testing.

AZT prophylaxis protocols generally advise administering 200 mg AZT every 4 h for 28 to 42 days. Some protocols skip the 4:00 a.m. dose. Since an exposure can occur anytime, AZT should be available 24 hours a day.

Transmission of Blood-Borne Pathogens from Health Care Workers to Patients

Blood-borne pathogens can also be transmitted from *HCWs* to patients. HIV, *HBV*, and *HCV* can potentially be transmitted to a patient during invasive procedures when a surgeon sustains a percutaneous injury with a needle or sharp instrument which then recontacts the patient. Only *HBV* and *HCV* have been demonstrated to be transmitted from physicians to patients. One dentist has transmitted HIV to six patients; the mechanism of transmission is unclear, however. There are no other reports of transmission of HIV from a health care worker to a patient. There are four reports following patients of surgeons with HIV infection. None of 767 serologically tested patients developed HIV infection as a result of being cared for by HIV-infected surgeons. None of more than 1000 patients of another surgeon who died of AIDS developed HIV infection. More than 9000 patients cared for by more than 75 health care workers with AIDS have been followed, and no cases of transmission by health care worker to patient have been reported. Approximately 60 of the 9000 patients were HIV-positive, but they were HIV-positive before being cared for by the health care worker, or they had other risk factors, or transmission from the health care worker was excluded.

Management of the HIV-, HBV-, or HCV-Infected Health Care Worker

The report of a dentist's having passed HIV to his patients sparked considerable discussion in the scientific and popular press about the HIV- positive health care worker, especially surgeons and dentists, since they are most likely to participate in invasive procedures. The CDC first issued guidelines for the management of HIV-infected personnel in 1985. It subsequently issued guidelines for management of HIV-infected *HCWs* who participate in invasive procedures. In its early guidelines the CDC recommended that health care personnel who are otherwise fit for duty and who do not participate in invasive procedures be allowed to perform their regular duties. In its subsequent guidelines the CDC recommended that HIV-infected personnel who do participate in invasive procedures be evaluated on a case-by-case basis. These recommendations were consolidated in 1987 with the suggestion that whether *HCWs* could perform their regular duties be decided on an individual basis.

After a dentist had been reported to have transmitted HIV to patients, the CDC issued another set of recommendations, suggesting that testing of *HCWs* for HIV or *HBV* not be required but that they should be tested voluntarily. (*HCV* testing was not yet commercially available and so was not mentioned in this set of recommendations.) *HCWs* who are infected with HIV or *HBV* should not perform “exposure-prone” procedures unless they have sought counsel from an expert panel; in addition, their patients should be informed of their seropositivity. The recommendations specify the composition of the panel but provide no guidelines on which the panel can base their decisions—and, in fact, defining “exposure-prone” procedures itself has met with resistance from the medical community.

ANTIMICROBIAL THERAPY

The use of antimicrobial agents in treating surgical infection does not differ fundamentally from antimicrobial usage in general medicine. The same basic considerations apply in treating all infections. One difference, however, is that antimicrobial therapy is only an adjunct in treating surgical infection; operative treatment (or percutaneous radiologically guided drainage of infected material) is more important. The goal of antimicrobial therapy is to prevent or treat infection by reducing or eliminating organisms until the host's own defenses can get rid of the last pathogens.

The basic considerations in antimicrobial therapy are efficacy, toxicity, and cost. Efficacy is the most important consideration in choosing an antimicrobial agent. Effective antimicrobial agents must be active against the pathogens causing the infection and must be able to reach the site of infection in adequate concentrations.

All antibiotics have potential toxicity. Toxic effects may be idiosyncratic, such as allergy or the rare instances of bone marrow aplasia caused by chloramphenicol. They can also cause damage to tissues and organs, such as in the renal toxicity or ototoxicity seen with the aminoglycosides or amphotericin B. Antimicrobial agents also exert selective pressures on the microbial ecology of the hospital that lead to resistant microbes, a problem that is especially important in intensive care units.

Cost is the final consideration in the selection of antimicrobial agents. Determining the costs of antimicrobial therapy includes more than just the cost of the drug. Drug administration charges, nursing time, intravenous fluid and lines, and monitoring costs must also be considered. In addition, any increased hospital time that occurs when an inexpensive agent that is less effective or that causes more toxicity is used ultimately makes that agent a more expensive antimicrobial.

Distribution of Antimicrobial Agents

Successful treatment of localized infections with systemic antimicrobial agents requires that an adequate concentration of drug be delivered to the site of infection. Ideally the tissue concentration of antibiotics should exceed the minimum inhibitory concentration. Tissue penetration depends in part on protein binding of antibiotics. Only the unbound form of antibiotics will pass through the capillary wall or act to inhibit bacterial growth. Therapeutic outcome, on the other hand, does not appear to be correlated with protein affinity, presumably because protein binding is easily reversible. Lipid solubility of antibiotics is also an important factor in tissue penetration. It determines the ability of antibiotics to pass through membranes by non- ionic diffusion or into wounds, bone, cerebrospinal fluid, the eye, endolymph of the ear, vegetations of bacterial endocarditis, and abscesses.

Blood

Rapidity of excretion and protein binding are two main determinants of blood concentration of antimicrobial agents. Protein binding affects the rapidity of excretion. Antibiotics that are highly protein bound are not excreted as rapidly as those with a low binding affinity and thus have longer half-lives. Therefore, highly protein-bound antibiotics generally do not have to be given as frequently as those with low protein binding. The efficacy of penicillins, cephalosporins, and other antibiotics that affect bacterial cell wall synthesis depends on the amount of time during which serum levels are above the minimum inhibitory concentrations rather than their peak serum concentration. The efficacy of aminoglycosides, on the other hand, is related to achieving peak serum concentrations that are four to eight times the minimum inhibitory concentration. Monitoring of serum aminoglycoside concentrations is usually necessary to ensure that these concentrations have been achieved; patients more commonly have subtherapeutic levels rather than toxic levels. In contrast, some antimicrobial agents such as nitrofurantoin and norfloxacin are excreted so rapidly in the urine that they never achieve blood (or tissue) levels sufficient to reach effective antibacterial concentrations. They do, however, reach high urinary concentrations and are effective agents for treating urinary tract infections.

Urine

Most commonly used antibiotics (sulfonamides, penicillins, cephalosporins, aminoglycosides, tetracyclines, quinolones, azoles) are excreted principally in the urine and achieve high urinary concentrations—up to 50 to 200 times their serum concentrations. Notable exceptions are erythromycin and chloramphenicol. Since concentrating ability is severely compromised in patients with renal disease, infections of the urinary tract are more difficult to treat in these patients. The pH of urine can be changed to facilitate antibiotic activity. For instance, aminoglycosides are more active in an alkaline medium, whereas other urinary antibacterial agents (tetracyclines, nitrofurantoin, methenamine mandelate) are more active in an acidic environment. Fortunately, the antimicrobials most commonly used to treat urinary tract infections have antimicrobial activity across a broad pH range.

Bile

Besides urine, only bile regularly has antibiotic concentrations higher than serum levels. The biliary concentrations of many of the penicillins (especially nafcillin, piperacillin, mezlocillin, and azlocillin), cephalosporins (especially cefazolin, cefamandole, ceforanide, cefoxitin, cefoperazone, and cefadroxil), tetracyclines, and clindamycin frequently are several times their serum concentrations. Nafcillin and rifampin achieve biliary concentrations 20 to 100 times those of serum. Aminoglycoside antibiotics enter bile less well, especially in the presence of liver disease, and their biliary concentrations are usually lower than serum levels.

Interstitial Fluid and Tissue

High, prolonged serum concentration and low protein binding favor diffusion of antibiotics from serum into extravascular tissue. Absolute tissue levels may not accurately reflect the therapeutic potential of the antibiotic, however, because the agent may be tightly bound to tissue and thus be unavailable for binding to bacteria.

Abscesses

There are few data of clinical relevance concerning the distribution of antibiotics into abscesses. The generalization that no antibiotics penetrate abscesses is not true. While the penicillins, cephalosporins, and some other antibiotics penetrate mature abscesses poorly, others such as metronidazole, chloramphenicol, and clindamycin can achieve inhibitory concentrations in abscesses.

A separate problem is whether, after penetration, an antibiotic can retain its antimicrobial efficacy under the conditions that exist in an abscess. The acidic pH, the low oxidation-reduction potential, and the large numbers of microbial and tissue products that can bind antibiotics all serve to reduce antimicrobial efficacy. Multiple types of bacteria within an abscess make it more likely that one type will inactivate an agent effective against it or another bacterium. The lack of efficacy of penicillins and cephalosporins in treating most abscesses may be a result of the high concentrations of beta-lactamases that accumulate there. Metronidazole and clindamycin can enter abscesses and retain antibacterial activity, but they are not effective against the aerobic gram-negative bacteria that are usually present together with the anaerobic bacteria against which they are effective—so the abscess usually persists.

An additional reason that antibiotics alone are seldom effective in treating abscesses is that antibiotics are most effective against actively metabolizing, rapidly dividing bacteria. Conditions in abscesses are usually unfavorable for bacterial growth, so the antibiotic is not able to enter and be active against the bacteria.

For all these reasons antibiotics alone should not be relied on for the treatment of most abscesses. Despite occasional reports of success with such treatment, drainage remains the mainstay of treating abscess.

Use of Antibiotics in Surgery

Prophylactic Antibiotics

Antibiotics are frequently administered prophylactically to patients undergoing operation to prevent wound infection when the likelihood of infection is high (e.g., when the tissues have been exposed to bacteria such as occurs during colon surgery) or when the consequences of infection are great even though the risk of infection is low (e.g., when a prosthetic device is implanted). The use of prophylactic antibiotics to prevent wound infection was discussed earlier in the section Wound Infection. Antibiotic prophylaxis should be administered to patients with previously placed prosthetic devices such as cardiac valves or artificial joints who are having any operation or dental procedure.

Therapeutic Use of Antibiotics

Many infections can be successfully treated with oral antibiotics on an outpatient basis. Severe surgical infections should be treated with intravenous antibiotics. Initial antibiotic therapy is usually empiric, since it should not be postponed until microbiotic studies are complete (Table 5-11). Antibiotic therapy should generally be initiated before cultures are obtained in patients with peritonitis, abscesses, and necrotizing soft tissue infections. Since cultures are usually obtained promptly during operative procedures or when percutaneous drainage has been performed, it is unlikely that prior antibiotic therapy will affect culture results.

Empiric Therapy

Rational empiric antibiotic therapy requires familiarity with the microbes most likely to cause infection at the involved site and antibiotic susceptibility patterns in the hospital or unit (e.g., intensive care unit). Intraabdominal surgical infections are nearly always caused by mixed gram- negative and gram-positive aerobic and anaerobic bacteria. Initial antibiotic therapy should provide broad-spectrum activity against these bacteria.

Most necrotizing soft tissue infections, especially those originating after an intraabdominal operation or occurring below the waist, are also due to a mixed bacterial flora, and broad-spectrum empiric therapy should be initiated. Because clostridia or streptococci can also cause these infections, penicillin G should generally be included. Once Gram stain and culture results are available, antibiotic therapy can be modified.

Prosthetic device infections usually progress much more slowly than intraabdominal or necrotizing soft tissue infections. Gram-positive cocci, especially *Staph. aureus* and *S. epidermidis*, play a prominent role in these infections, but they can also be caused by gram-negative bacteria.

Numerous single and combination antimicrobials are available for initial and empiric therapy. The Surgical Infection Society (SIS) has made recommendations for use of antimicrobial agents for empiric therapy of intraabdominal infections. The SIS recommends against using drugs such as cefazolin and other first-generation cephalosporins, penicillin, cloxacillin and other antistaphylococcal penicillins, ampicillin, erythromycin, and vancomycin because these drugs do not provide adequate coverage for both aerobic and anaerobic organisms.

Metronidazole and clindamycin should not be used as single agents for mixed infection, because they lack activity against aerobic enteric organisms. Other antibiotics, such as aminoglycosides, aztreonam, cefuroxime, cefonicid, cefamandole, ceforanide, cefotetan, cefotaxime, ceftizoxime, cefoperazone, ceftriaxone, ceftazidime, and polymyxin should not be used alone because of the inadequate coverage of anaerobic gram- negative bacilli. Because of inadequate clinical data documenting efficacy and concerns about resistance, the SIS also recommends against using as single agents for empiric therapy antibiotics such as piperacillin, mezlocillin, azlocillin, ticarcillin, and carbenicillin despite their relative safety and broad in vitro antibacterial activity. Chloramphenicol has an appropriate in vitro spectrum of activity but is not acceptable because it can produce serious side effects.

Acceptable agents for community-acquired intraabdominal infections include cefoxitin, cefotetan, cefmetazole, and ticarcillin/clavulanic acid. However, these antibiotics should not be used for patients whose abdominal infection develops in the hospital after previous antibiotic therapy. For these infections and serious intraabdominal infections antibiotics such as imipenem-cilastatin (Primaxin) should be used. Combination therapy such as metronidazole or clindamycin plus an aminoglycoside or an antianaerobic antibacterial agent plus a third-generation cephalosporin or clindamycin plus a monobactam is acceptable. Cost and toxicity considerations can make one of these recommendations preferable to the others. The combination of an antianaerobic antibiotic plus an aminoglycoside plus penicillin or ampicillin is recommended only if enterococcal infection is suspected on the basis of a Gram stain or thought to be clinically relevant (e.g., associated with enterococcus bacteremia). Community-acquired intraabdominal infections are seldom associated with serious enterococcus infection.

Definitive Therapy

Antimicrobial therapy may have to be altered when the results of Gram stain, culture, and sensitivity data are available (Table 5-12). Sensitivity data may determine that one of the antibiotics currently being used is not active against one of the bacteria isolated. In addition, change to a less toxic or less costly antimicrobial agent may be possible once laboratory results are available.

Infections originating in the intensive care unit are frequently caused by antibiotic-resistant bacteria. This is especially true for hospital-acquired *Staph. aureus*, which is frequently resistant to methicillin. For hospital- acquired staphylococcal infections vancomycin should generally be initiated if methicillin-resistant *Staph. aureus* is a problem in the hospital until definitive sensitivity data are available. If the *Staph. aureus* is sensitive to penicillin G or methicillin, these agents should be used because they are more effective and less costly than vancomycin. Two drugs are generally used to treat *P. aeruginosa* infections, an antipseudomonal beta-lactam drug such as mezlocillin or ceftazidime in combination with an aminoglycoside, in an attempt to prevent development of resistance and to take advantage of possible synergism.

Drug Administration

Route

For seriously ill surgical patients the antimicrobial agent should be administered intravenously to ensure adequate serum levels. Absorption by other routes is inconsistent in seriously ill patients whose gastrointestinal tract is not functioning properly and who have problems maintaining blood pressure. If patients need prolonged antimicrobial therapy, other routes can be used once they have begun to recover, or long-term intravenous antimicrobial therapy can be given on an outpatient basis.

Recommendations provided by the manufacturer should be used as guidelines for appropriate doses of antimicrobial agents. In general there is a wide margin between therapeutic and toxic concentrations with drugs such as the penicillins and cephalosporins. Other agents, such as the aminoglycosides, have a much narrower margin between therapeutic and toxic levels. For these antibiotics the calculated dose in adults is based on lean body weight. For children, antibody dosing is frequently based on surface area.

Duration

There are few data defining the appropriate duration of antibiotic treatment. Most surgical infections can be treated effectively in 5 to 7 days of antibiotic therapy. It is generally safe to stop antibiotics as long as the patient is making clinical progress and has a normal temperature and white blood cell count, and gastrointestinal function has returned in patients with peritonitis. If clinical improvement is not evident within 4 to 5 days after operation and fever or leukocytosis persists after more than 5 days of therapy, a reason for the apparent treatment failure should be sought.

Treatment Failure

Although failure of a bacterial infection to respond to a particular antibiotic is commonly regarded as evidence that the wrong antibiotic was selected, usually other factors are responsible. Patients with intraabdominal infections who remain febrile or have persistent leukocytosis usually have recurrent (tertiary) peritonitis or an intraabdominal abscess that requires drainage. Patients with necrotizing soft tissue infections may have persistent infections. Other causes of fever such as pneumonia, urinary tract infection, vascular catheter-related infections, drug fever, and thrombophlebitis should be investigated.

Finally, the antibiotic may be inappropriate. It may be the wrong antibiotic, or it may have been given in an inadequate dose or by an inappropriate route. The bacteria may not be susceptible to the antibiotic at the concentration achievable at the site of infection, or the site may have become superinfected by another bacterium not sensitive to the antibiotic.

Drug Toxicity

Normally antibiotics are excreted primarily by the kidneys and accumulate in the serum of patients with impaired renal function. Therefore, with many antibiotics it is necessary to reduce the dose or to increase the interval between doses in patients with renal failure (there are many schedules that detail how to estimate dosages of antibiotics strongly excreted by the kidneys, but none of them is perfect). Toxic drugs such as the aminoglycosides should either not be used in patients with renal failure or impaired renal function, or, if used, their serum or plasma concentrations must be obtained frequently to verify that toxic levels are not being reached.

The general approach to antibiotic usage in patients with renal failure is to give a first dose of 80 to 100 percent of the usual amount and then to estimate the timing and the amount of the second dose according to various schedules based on the normal half-life of the antibiotic.

Immunotherapy and Biologic Therapy of Infection

Antibodies to bacterial products and to mediators of sepsis are new (and extremely costly) therapeutic modalities that are currently being evaluated. Results thus far have been disappointing. There are no currently approved immunotherapeutic agents for treating infections. A previously approved anti-endotoxin antibody (HA-1A) has been taken off the market.

Other molecules or antagonists of molecules of the inflammatory or septic response are being investigated in the laboratory or undergoing clinical trials. Although none is currently available for clinical use, they may prove to be efficacious in the future.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 6 - Trauma

Jon M. Burch
Reginald J. Franciose
Ernest E. Moore

Trauma or injury has been defined as damage to the body caused by an exchange with environmental energy that is beyond the body's resilience. Trauma remains the most common cause of death for individuals between the ages of 1 and 44 years, and the third most common cause of death for all ages. The U.S. government classifies accidental death under the following categories: accidents and adverse effects; suicide, homicide, and legal intervention; and all other external causes. Accidents and adverse effects account for approximately 100,000 deaths per year, of which motor vehicle accidents account for nearly 50 percent. Homicides, suicides, and other causes are responsible for another 50,000 deaths each year. Death rates are a poor indicator of the magnitude of the problem, however, because most injured patients survive. For example, in 1985 there were approximately 140,000 trauma-related deaths, but 57 million reported injuries and 23 million hospitalizations. For the same year the aggregate lifetime costs for all injured patients was estimated to be \$158 billion. Trauma is a major public health issue.

INITIAL EVALUATION AND RESUSCITATION OF THE INJURED PATIENT

Treatment of trauma patients often begins in the field by *emergency medical services (EMS)* personnel and completed by rehabilitation specialists. Although the Advanced Trauma Life Support (ATLS) course of the American College of Surgeons Committee on Trauma is directed at primary care physicians in rural communities, its format and basic tenets are sound for all physicians. The initial treatment of seriously injured patients consists of a primary survey, resuscitation, secondary survey, diagnostic evaluation, and definitive care. The concepts are presented in a sequential fashion, but in reality they often proceed simultaneously. The process begins with the identification and treatment of conditions that constitute an immediate threat to life. The ATLS course refers to this as the primary survey or the ABCs- Airway, with cervical spine protection, Breathing, and Circulation. Any life-threatening problem identified in the initial survey must be treated before advancing.

Airway Management

Ensuring an adequate airway is the first priority in the primary survey. Efforts to restore cardiovascular integrity will be futile if the oxygen content of the blood is inadequate. Simultaneously, all blunt-trauma patients require cervical spine immobilization until injury is ruled out. This can be accomplished with a hard (Philadelphia) collar or sandbags on both sides of the head taped to the backboard. Soft collars do not immobilize the cervical spine.

Patients who are conscious and have a normal voice do not require further evaluation or early attention to their airway. Exceptions to this principle include patients with penetrating injuries to the neck and an expanding hematoma, evidence of chemical or thermal injury to the mouth, nares, or hypopharynx, extensive subcutaneous air in the neck, complex maxillofacial trauma, or airway bleeding. These patients initially may have a satisfactory airway, but it may become obstructed if soft tissue swelling or edema progresses. In these cases, elective intubation should be performed before evidence of airway compromise is apparent.

Patients who have an abnormal voice or altered mental status require further airway evaluation. Direct laryngoscopic inspection often reveals blood, vomit, the tongue, foreign objects, or soft tissue swelling as sources of airway obstruction. Suctioning can offer immediate relief in many patients. Altered mental status is the most common indication for intubation because of the patient's inability to protect the airway. Options for airway access include nasotracheal, orotracheal, or operative intervention. Nasotracheal intubation can be accomplished only in patients who are breathing spontaneously and is contraindicated in the apneic patient. Although nasotracheal intubation frequently is used by paramedics in the field, the primary use for this technique in the emergency room is becoming limited to those few patients requiring emergent airway support who are prohibitive candidates for paralyzation.

Orotracheal intubation also can be performed in patients with potential cervical spine injuries provided that manual in-line cervical immobilization is maintained. The advantages of orotracheal intubation are the direct visualization of the vocal cords, the ability to use larger-diameter endotracheal tubes, applicability to apneic patients, and its familiarity to most physicians. The disadvantage of orotracheal intubation is that conscious patients usually require neuromuscular blockade or deep sedation. To a large extent, rapid-sequence induction of anesthesia with orotracheal intubation has become the standard in experienced trauma centers with the availability of pulse oximetry. The major advantage is rapid, definitive airway control. The disadvantages include the inability to intubate, aspiration, and complications of the required medications. Those who attempt rapid-sequence induction must be thoroughly familiar with the details and contraindications of the procedure.

Patients in whom attempts at intubation have failed or are precluded because of extensive facial injuries require a surgical airway. Cricothyroidotomy (Fig. 6-1) and percutaneous transtracheal ventilation are preferred over tracheostomy in most emergency situations because of their simplicity and safety. One disadvantage of cricothyroidotomy is the inability to place a tube greater than 6 mm in diameter because of the limited aperture of the cricothyroid space. Cricothyroidotomy also is contraindicated in patients under the age of twelve because of the risk of damage to the cricoid cartilage and the subsequent risk of subglottic stenosis.

Percutaneous transtracheal ventilation is accomplished by inserting a large-bore intravenous catheter through the cricothyroid membrane into the trachea and attaching it with tubing to an oxygen source capable of delivering 50 psi or more. A hole cut in the tubing allows for intermittent ventilation by occluding and releasing the hole. Adequate oxygenation can be maintained for more than 30 min. Because exhalation occurs passively, ventilation is limited and carbon dioxide retention can occur. Emergent tracheostomy has fallen into disfavor because of its technical difficulties; it may be necessary in cases of laryngotracheal separation or laryngeal fractures when cricothyroidotomy might cause further damage or result in the complete loss of the airway.

Breathing

Once a secure airway is obtained, adequate oxygenation and ventilation must be assured. All injured patients should receive supplemental oxygen therapy and be monitored by pulse oximetry. The following conditions may constitute an immediate threat to life because of inadequate ventilation: (1) tension pneumothorax, (2) open pneumothorax, or (3) flail chest/pulmonary contusion. These diagnoses can be made with a combination of physical examination and chest x-ray.

The diagnosis of tension pneumothorax is implied by the finding of respiratory distress in combination with any of the following physical signs: tracheal deviation away from the affected side; lack of or decreased breath sounds on the affected side; distended neck veins or systemic hypotension; or subcutaneous emphysema on the affected side. Immediate tube thoracostomy is indicated without awaiting chest x-ray confirmation (Fig. 6-2). In tension pneumothorax the collapsed lung acts as a one-way valve so that each inhalation allows additional air to accumulate in the pleural space. The normal negative intrapleural pressure becomes positive, depressing the ipsilateral hemidiaphragm and forcing the mediastinal structures into the contralateral chest. The contralateral lung is then compressed, and the heart is rotated about the superior and inferior venae cavae, decreasing venous return and cardiac output while distending the neck veins. An unrecognized simple pneumothorax can be converted to a tension pneumothorax if the patient is placed on a positive-pressure mechanical ventilator. A tension pneumothorax also can develop in a patient who is breathing spontaneously.

An open pneumothorax or sucking chest wound occurs with full-thickness loss of the chest wall, permitting a free communication between the pleural space and the atmosphere. This compromises ventilation by two mechanisms. In addition to collapse of the lung on the injured side, if the diameter of the injury is greater than the narrowest portion of the upper airway, air preferentially moves through the injury site rather than the trachea and impairs ventilation on the contralateral side. Occlusion of the injury may result in converting an open pneumothorax into a tension pneumothorax. Proper treatment in the field involves placing an occlusive dressing over the wound, which is taped on three sides. The occlusive dressing permits effective ventilation on inspiration while the untaped side allows accumulated air to escape from the pleural space, preventing a tension pneumothorax. Definitive treatment requires wound closure and tube thoracostomy.

Flail chest occurs when four or more ribs are fractured in at least two locations. Paradoxical movement of this free-floating segment of chest wall may be sufficient to compromise ventilation. It is of greater physiologic importance that patients with flail chest frequently have an underlying pulmonary contusion. Pulmonary contusion with or without rib fractures may compromise oxygenation or ventilation to the extent that intubation and mechanical ventilation is required. Respiratory failure in these patients may not be immediate, and frequent reevaluation is warranted. The initial chest x-ray usually underestimates the degree of pulmonary contusion, and the lesion tends to evolve with time and fluid resuscitation.

Circulation

With a secure airway and adequate ventilation established, circulatory status is determined. A rough first approximation of the patient's cardiovascular status is obtained by palpating peripheral pulses. A systolic blood pressure of 60 mmHg is required for the carotid pulse to be palpable, 70 mmHg for the femoral pulse and 80 mmHg for the radial pulse. At this point in the patient's treatment, hypotension is assumed to be caused by hemorrhage. Blood pressure and pulse should be measured at least every 15 min.

External control of hemorrhage should be obtained before restoring circulating volume. Manual compression and splints frequently control extremity hemorrhage as effectively as tourniquets and with less tissue damage. Blind clamping should be avoided because of the risk to adjacent structures, particularly nerves. The importance of digital control of hemorrhage for penetrating injuries of the head, neck, thoracic outlet, groin, and extremities cannot be overemphasized. This should be done with a gloved finger placed through the wound directly on the bleeding vessel applying only enough pressure to control active bleeding. The surgeon performing this maneuver must then walk along beside the patient on the way to the operating room for definitive treatment. Scalp lacerations through the galea aponeurotica tend to bleed profusely; these can be temporarily controlled with Rainey clips or a full-thickness large nylon continuous stitch.

Intravenous access for fluid resuscitation is begun with two peripheral catheters, 16-gauge or larger in an adult. Blood should be drawn simultaneously and sent for typing and hematocrit measurement. Because the flow of liquid through a tube is proportional to diameter and inversely proportional to length, venous lines for volume resuscitation should be short with a large diameter. For patients requiring vigorous fluid resuscitation, saphenous vein cutdowns at the ankle (Fig. 6-3) or percutaneous femoral vein catheter introducers are preferred. The saphenous vein is reliably found 1 cm anterior and 1 cm superior to the medial malleolus. Short 10-gauge catheters can be quickly placed even in an exsanguinating patient with collapsed veins. Venous access in the lower extremities provides effective volume resuscitation in cases of abdominal venous injury, including the inferior vena cava. Jugular and subclavian central venous introducers are less desirable for initial access in trauma patients because placement can interfere with the work of staff members performing other lifesaving procedures. Secondary central venous introducers should be placed in the operating room in the event that vena caval cross-clamping is performed.

In hypovolemic pediatric patients less than 6 years of age, percutaneous femoral vein cannulation is contraindicated because of the risk of venous thrombosis. If two attempts at percutaneous peripheral access are unsuccessful, interosseous cannulation should be performed in the proximal tibia, or in the distal femur if the tibia is fractured (Fig. 6-4). This is a safe emergency technique; however, once alternative access has been established, the cannula should be removed because of the risk of osteomyelitis.

Initial Fluid Resuscitation

Initial fluid resuscitation is a 1-L intravenous bolus of normal saline, lactated Ringer's solution, or other isotonic crystalloid in an adult, or 20 mL/kg of body weight lactated Ringer's solution in a child. In the United States crystalloid alone is used, whereas in other parts of the world colloid is often added. This is repeated once in an adult and twice in a child before administering red blood cells. The goal of fluid resuscitation is to reestablish tissue perfusion. Classic signs and symptoms of shock are tachycardia, hypotension, tachypnea, mental status changes, diaphoresis, and pallor. None of these signs or symptoms taken alone can predict the patient's organ perfusion status, but when viewed together they can help in evaluating the patient's response to treatment. Patients who have a good response to fluid infusion, i.e., normalization of vital signs, clearing of the sensorium, evidence of good peripheral perfusion (warm fingers and toes with normal capillary refill) are presumed to have adequate perfusion.

There are several caveats to keep in mind when making this presumption. Although tachycardia may be the earliest sign of ongoing blood loss, individuals in good physical condition, particularly trained athletes with a low resting pulse rate, may manifest only a relative tachycardia. Patients on beta-blocking medications may not be able to increase their heart rate in response to stress. In children, bradycardia or relative bradycardia can occur with severe blood loss and is an ominous sign, often heralding cardiovascular collapse. Conversely, hypoxia, pain, apprehension, and stimulant drugs (e.g., cocaine, amphetamines) produce a tachycardia unrelated to physiologic demands. Hypotension is not a reliable early sign of hypovolemia. In healthy patients blood volume must decrease by 30 to 40 percent before hypotension occurs (Table 6-1). Younger patients with good sympathetic tone can maintain systemic blood pressure with severe intravascular deficits until they are on the verge of cardiac arrest. In contrast, pregnancy increases circulating blood volume, and a relatively larger volume of blood loss must occur before signs and symptoms become apparent.

Acute changes in mental status can be caused by hypoxia, hypercarbia, or hypovolemia, or they may be an early sign of increasing intracranial pressure (ICP). An abnormal mental status should prompt an immediate reevaluation of the ABCs and consideration of an evolving central nervous system injury. A deterioration in mental status may be subtle and may not progress in a predictable fashion; for example, a previously calm and cooperative patient may become anxious and combative as hypoxia develops, or a patient who is agitated and combative from drugs or alcohol may become somnolent if hypovolemic shock develops. Urine output is a quantitative and relatively reliable indicator of organ perfusion. Adequate urine output is 0.5 mL/kg/h in an adult, 1 mL/kg/h in a child, and 2 mL/kg/h in an infant less than 1 year of age.

On the basis of the initial response to fluid resuscitation, hypovolemic injured patients may be placed into three broad categories: responders, transient responders, and nonresponders. Individuals who are stable or have a good response to the initial fluid therapy as evidenced by normalization of vital signs, mental status, and urine output are unlikely to have significant continuing hemorrhage, and further diagnostic evaluation for occult injuries can proceed. At the other end of the spectrum are nonresponders with persistent hypotension. This group requires immediate diagnosis and treatment to prevent a fatal outcome. Patients who respond transiently and then deteriorate present the most complex decision-making challenge. They usually are underresuscitated or have ongoing hemorrhage. In patients with penetrating trauma, the need for operative intervention for the control of hemorrhage usually is evident. Blunt trauma patients with multisystem injury, however, require careful planning. It is in this group that the greatest number of preventable deaths is likely to occur.

Persistent Hypotension Nonresponders

The spectrum of disease in this category ranges from nonsurvivable multisystem injury to problems as simple as a tension pneumothorax. Persistent hypotension in these patients usually is cardiogenic or a result of uncontrolled hemorrhage. An evaluation of the patient's neck veins and *central venous pressure (CVP)* usually distinguishes between these two categories. CVP determines right ventricular preload; in otherwise healthy trauma patients, its measurement yields objective information regarding the patient's overall volume status. Central venous catheters are inappropriate for administering large volumes of fluid, but they are valuable for measuring CVP. A hypotensive patient with flat neck veins and a CVP less than 5 cmH₂O is hypovolemic and is likely to have ongoing hemorrhage. A hypotensive patient with distended neck veins or a CVP more than 15 cmH₂O is likely to be in cardiogenic shock. The CVP may be falsely elevated if the patient is agitated and straining or fluid administration is overzealous; isolated readings must be interpreted with caution.

In trauma patients the differential diagnosis of cardiogenic shock is indicated by: (1) tension pneumothorax, (2) pericardial tamponade, (3) myocardial contusion or infarction, and (4) air embolism. Tension pneumothorax is the most frequent cause of cardiac failure. Traumatic pericardial tamponade is most often associated with penetrating injury to the heart. As blood leaks out of the injured heart, it accumulates in the pericardial sac. Because the pericardium is not acutely distensible, the pressure in the pericardial sac rises to match that of the injured chamber. This pressure usually is greater than that of the right atrium; right atrial filling is impaired, and right ventricular preload is reduced. This leads to decreased right ventricular output and increased CVP. Increased intrapericardial pressure also impedes myocardial blood flow, which leads to subendocardial ischemia and a further reduction in cardiac output. This cycle may progress insidiously with injury of the venae cavae or atria, or precipitously with injury of either ventricle. With acute tamponade, as little as 100 mL of blood within the pericardial sac can produce life-threatening hemodynamic compromise. The usual presentation is a patient with a penetrating injury in proximity to the heart who is hypotensive and has distended neck veins or an elevated CVP. The classic findings of Beck's triad (hypotension, distended neck veins, and muffled heart sounds) and pulsus paradoxus are not reliable indicators of acute tamponade. Ultrasound imaging in the emergency room using a subxiphoid or parasternal view is extremely helpful if the findings are clearly positive (Fig. 6-5), but equivocal findings are common. Early in the course of tamponade, blood pressure and cardiac output transiently improve with fluid administration, which may lead the surgeon to question the diagnosis—or lull the surgeon into a false sense of security.

Once the diagnosis of cardiac tamponade is established, pericardiocentesis should be performed (Fig. 6-6). Evacuation of as little as 15 to 25 mL of blood can dramatically improve the patient's hemodynamic profile. Pericardiocentesis should be done even if the patient stabilizes with volume loading because subclinical myocardial ischemia can lead to sudden lethal arrhythmias, and patients with tamponade can decompensate unpredictably. While pericardiocentesis is being performed, preparation should be made for emergent transport to the operating room. Emergent pericardiocentesis is successful in decompressing the tamponade in approximately 80 percent of cases; most failures are a result of clotted blood within the pericardium. If pericardiocentesis is unsuccessful and the patient remains severely hypotensive (systolic blood pressure <70 mmHg) or shows other signs of hemodynamic instability, emergency room thoracotomy should be performed (Fig. 6-7). This is best accomplished using a left anterolateral thoracotomy and a longitudinal pericardiotomy anterior to the phrenic nerve, followed by evacuation of the pericardial sac and temporary control of the cardiac injury. The patient is then transported to the operating room for definitive repair (Fig. 6-8).

Myocardial contusion from direct myocardial impact occurs in approximately one-third of patients sustaining significant blunt chest trauma. The diagnostic criteria for myocardial contusion include specific electrocardiogram abnormalities, i.e., ventricular dysrhythmias, atrial fibrillation, sinus bradycardia, and bundle branch block. Transient sinus tachycardia is not indicative of contusion. Serial cardiac enzyme determinations (CPK-MB fraction) lack sensitivity and are not predictive of complications under these conditions, and they are not recommended. While the diagnosis is common, acute life-threatening complications of ventricular arrhythmias and cardiac pump failure occur in less than 5 percent and less than 1 percent, respectively, of patients sustaining major blunt chest trauma. Arrhythmias are treated by pharmacologic suppression. The management of cardiogenic shock from cardiac pump failure includes early placement of a Swan-Ganz pulmonary artery catheter to optimize fluid administration, inotropic support, and urgent echocardiography to rule out septal or free wall rupture, valvular disruption, or pericardial tamponade. Patients with refractory cardiogenic shock might require placement of an intraaortic balloon pump to decrease myocardial work and enhance coronary perfusion.

Acute myocardial infarction is itself frequently the cause of motor vehicle accidents or other trauma in older patients. While the ideal initial management is to provide optimal treatment for the evolving infarction, decisions regarding lytic therapy and emergent angioplasty must be individualized according to the patient's other injuries.

Air embolism is a frequently overlooked lethal complication of pulmonary injury. It occurs when air from an injured bronchus enters an adjacent injured pulmonary vein and returns to the left heart. Air accumulation in the left ventricle impedes diastolic filling, and during systole it is pumped into the coronary arteries, disrupting coronary perfusion. The typical scenario is a patient with a penetrating chest injury who appears hemodynamically stable but suddenly goes into cardiac arrest after being intubated and placed on positive-pressure ventilation. Air emboli also have been described in conjunction with blunt thoracic trauma and can occur at any time when a pulmonary venous injury is being manipulated. The patient should be placed in the Trendelenburg position to trap the air in the apex of the left ventricle. Emergency thoracotomy is followed by cross-clamping the pulmonary hilum on the side of the injury to prevent further introduction of air. Air is aspirated from the apex of the left ventricle with an 18-gauge needle and 50-mL syringe. Vigorous open cardiac massage is used to force the air bubbles through the coronary arteries. The highest point of the aortic root also is aspirated to prevent air from entering the coronary arteries or embolizing to the brain. The patient should be kept in the Trendelenburg position and the hilum clamped until the pulmonary venous injury is controlled.

A state of persistent hypotension and flat neck veins resulting from uncontrolled hemorrhage is associated with a high mortality. A rapid search for the source or sources of hemorrhage, including visual inspection with knowledge of the injury mechanism, abdominal ultrasound imaging, anteroposterior chest and pelvic x-rays, usually indicates the regions of the body responsible for the blood loss. Type O red blood cells (O-negative for women of childbearing age) or type-specific red blood cells should be administered and the patient taken directly to the operating room for exploration. For patients with a sustained systolic blood pressure of less than 70 mmHg, in spite of crystalloid and blood administration, emergency room thoracotomy should be considered. The clearest indication for this procedure is penetrating chest trauma, and survival is reported as high as 30 percent. A small number of patients with penetrating abdominal trauma survive, but the role of emergency room thoracotomy in blunt abdominal trauma is controversial. The goal of emergency room thoracotomy for thoracic injuries is control of hemorrhage; for abdominal injuries the goal is to sustain central circulation and limit abdominal blood loss by clamping the descending thoracic aorta. Every effort should be made to replace the aortic clamp to below the renal arteries within 30 min. Longer clamping times proximal to the abdominal viscera are seldom associated with survival. The decision to perform an emergency room thoracotomy can be assisted by use of the algorithm in Fig. 6-9.

Transient Responders

Hypotensive patients who transiently respond to fluid administration usually have some degree of active hemorrhage. Those with penetrating injuries should be taken to the operating room for exploration. Those with multiple blunt injuries constitute a diagnostic and therapeutic dilemma. These patients often require sophisticated evaluation such as computed tomography (CT) and angiography. It is during these diagnostic evaluations and the necessary transportation that the greatest hazard exists, because monitoring is compromised and the environment is suboptimal for dealing with acute problems. The surgeon must accompany the patient and be prepared to abort the examination if hypotension recurs. If it does, the patient should be given type-specific red blood cells and transported immediately to the operating room to localize the hemorrhage. An operating room should be immediately available when these patients arrive in the emergency room.

The traditional volume resuscitation (described above) of patients sustaining penetrating torso trauma has been questioned. It has been assumed that any hypotension is dangerous and must be treated, preferably with blood or crystalloid, but some have argued that hemostatic mechanisms frequently control hemorrhage initially, and increased venous and subsequent arterial pressure from fluid resuscitation can disrupt tenuous hemostasis. Furthermore, active bleeding increases as venous and arterial pressure increases. Laboratory studies support these concepts. In a prospective randomized study of hypotensive patients who sustained penetrating torso trauma and required operative treatment, half the patients received volume resuscitation and fluid was withheld in the others until the operation was begun, but there was no survival advantage for those resuscitated in the traditional fashion. Subgroup analysis suggested a survival disadvantage for pericardial tamponade. Patients with profound hypotension (systolic blood pressure <70 mmHg) are at risk for sudden death. Controlled hypotension is the optimal middle ground.

Secondary Survey

When the conditions that constitute an immediate threat to life have been attended to or excluded, the patient is examined in a systematic fashion to identify occult injuries. Special

attention should be given to the patient's back, axillae, and perineum because injuries in these areas are easily overlooked. Patients should undergo digital rectal examination to evaluate sphincter tone and to look for blood, perforation, or a high-riding prostate. A Foley catheter should be inserted to decompress the bladder, obtain a urine specimen, and monitor urine output. Stable patients at risk for urethral injury should undergo urethrography before catheterization. Signs of urethral injury include blood at the meatus, perineal or scrotal hematomas, or a high-riding prostate. In the case of persistent hypovolemic shock, an initial attempt at a Foley catheterization should be made; if this is unsuccessful, a percutaneous suprapubic cystostomy should be placed. A nasogastric tube should be inserted to decrease the risk of gastric aspiration and allow inspection of the contents for blood suggestive of occult gastroduodenal injury.

Selective radiographs are obtained early in the emergency room evaluation. For patients with severe blunt trauma, anteroposterior chest and pelvic radiographs should be obtained as soon as possible. For patients with truncal gunshot wounds, posteroanterior and lateral radiographs of the chest and abdomen are warranted. It is helpful to mark the entrance and exit sites of penetrating wounds with metallic clips or staples so that the trajectory of the missile or blade can be estimated.

Many trauma patients cannot provide specific information about the nature of their injury mechanism. Emergency medical personnel and police are trained to evaluate an injury scene and should be questioned. For automobile accidents the speed of the accident, the angle of impact (if any), the use of restraints, airbag deployment, condition of the steering wheel and windshield, the amount of intrusion, whether the patient was ejected from the vehicle, and whether anyone was dead at the scene should all be ascertained. The patient's physiologic condition in the field also is important. Vital signs and mental status in the emergency room can be compared with those at the scene; improvement or deterioration provide critical prognostic information.

Mechanisms and Patterns of Injury

Evaluation and decision making are far more difficult in blunt trauma than in penetrating trauma. More energy is transferred over a wider area during blunt trauma than from a gunshot wound or a stab wound. As a result, blunt trauma is associated with multiple, widely distributed injuries, whereas in penetrating injuries the damage is localized to the path of the bullet or knife in penetrating wounds. Patients who have sustained blunt trauma are separated into categories according to their risk for multiple injuries, high energy transfer and low energy transfer. Injuries involving high energy transfer include auto-pedestrian accidents, motor vehicle accidents in which the car's change of speed exceeds 20 mph or in which the patient has been ejected, motorcycle accidents, and falls from heights greater than 20 feet. The greatest risk factors reflecting magnitude of injury from the field associated with life-threatening injuries are death of another occupant in the vehicle and an extrication time greater than 20 min.

Patients who have sustained high-energy-transfer trauma have certain patterns of injury related to the mechanism, e.g., when unrestrained drivers suffer frontal impacts, their heads strike the windshield, their chests and upper abdomens hit the steering column, and their legs or knees contact the dashboard. The resultant injuries frequently include facial fractures, cervical spine fractures, laceration of the thoracic aorta, myocardial contusion, injury to the spleen and liver, and fractures of the pelvis and lower extremities. The discovery of one of these injuries should prompt a search for others.

Low-energy trauma, such as being struck with a club or falling from a bicycle, usually does not result in widely distributed injuries, but potentially lethal lacerations of internal organs can still occur because the net energy transfer to that location may be substantial.

Penetrating injuries are classified according to the wounding agent, i.e., stab wounds, gunshot wounds, or shotgun wounds. Gunshot wounds are subdivided further into high- and low-velocity injuries because the speed of the bullet is much more important than its weight in determining kinetic energy. Experience in urban trauma centers indicates that high-velocity gunshot wounds (bullet speed greater than 2000 ft/s) are rare in the civilian setting. Shotgun injuries are divided into close-range (< 7 meters) and long-range wounds. Close-range shotgun wounds are comparable to high-velocity wounds because the entire energy of the load is delivered to a small area, often with devastating results. Long-range shotgun wounds result in a diffuse pellet pattern in which many pellets miss the victim, and those that do strike are dispersed and of comparatively low energy.

Regional Assessment and Special Diagnostic Tests

Additional diagnostic studies are often indicated on the basis of mechanism of injury, location of injuries, screening x-rays, and the patient's overall condition. The patient is in constant jeopardy when undergoing special diagnostic testing. The surgeon should be in attendance and be prepared to alter plans as circumstances demand. Hemodynamic, respiratory, and mental statuses determine the most appropriate course of action.

Head

A score based on the *Glasgow Coma Scale (GCS)* should be determined for all injured patients (Table 6-2). It is calculated by adding the scores of the best motor response, best verbal response, and eye opening. Scores range from 3 (the lowest) to 15 (normal). Scores of 13 to 15 indicate mild head injury, 9 to 12, moderate injury, and less than 9, a severe injury. The GCS is useful for triage and prognosis.

Examination of the head should focus on potentially treatable neurologic injuries. The presence of lateralizing findings are important, e.g., a unilateral dilated pupil unreactive to light, asymmetric movement of the extremities either spontaneously or in response to noxious stimuli, or a unilateral Babinski's reflex suggest a treatable intracranial mass lesion or major structural damage. Stroke syndromes should prompt a search for carotid dissection or thrombosis using duplex scanning or angiography. Otorrhea, rhinorrhea, "raccoon eyes," and Battle's sign (ecchymosis behind the ear) can be seen with basilar skull fractures. While not necessarily requiring treatment, these fractures carry an increased risk of meningitis in the postinjury period. The head and face should be systematically palpated for fractures. Patients with a significant closed head injury (GCS less than 14) should have a CT scan performed. For penetrating injuries plain skull films should be obtained as well, as they can provide information that CT does not.

Cerebral pathologic lesions from blunt trauma include hematomas, contusions, hemorrhage into ventricular and subarachnoid spaces, and diffuse axonal injury (DAI). Hematomas are further classified according to location. Epidural hematomas occur when blood accumulates between the skull and the dura and are caused by disruption of the middle meningeal artery or other small arteries in that potential space from a skull fracture (Fig. 6-10). Subdural hematomas occur between the dura and the cerebral cortex and are caused by venous disruption or laceration of the parenchyma of the brain (Fig. 6-11). Because of the underlying brain injury, prognosis is much worse with subdural hematomas. Intraparenchymal hematomas and contusions can occur anywhere within the brain. Hemorrhage may occur into the ventricles, and though usually not massive, this blood may cause postinjury hydrocephalus. Diffuse hemorrhage into the subarachnoid space may cause vasospasm and reduce cerebral blood flow. DAI results from high- speed deceleration injury and represents direct axonal damage. On CT a blurring of the gray/white matter interface may be seen, with multiple, small, punctate hemorrhages. While prognosis is difficult to predict, early evidence of DAI on CT scan is associated with a poor outcome. Magnetic resonance imaging (MRI) can often identify DAI with greater precision than CT.

Significant penetrating injuries usually are produced by bullets from handguns, but an array of other weapons or instruments can injure the cerebrum via the orbit or through the thinner temporal region of the skull. While the diagnosis usually is obvious, in some instances wounds in the auditory canal, mouth, and nose can be elusive. Prognosis is variable, but most supratentorial wounds that injure both hemispheres are fatal.

Neck

In evaluating the neck of a blunt trauma victim, attention should be focused on signs and symptoms of an occult cervical spine injury. Because of the devastating consequences of quadriplegia, all patients should be assumed to have cervical spine injuries until proven otherwise. The presence of posterior midline pain or tenderness should provoke a thorough radiologic evaluation. There is no perfect test to detect all injuries. A cervical spine series including lateral view with visualization of C7–T1, anteroposterior view, and transoral odontoid view is sufficient to detect most significant fractures and subluxations. If pain or tenderness persists in spite of normal appearance on plain x-ray films, a CT scan should be done. CT identifies most fractures but can miss some subluxations. A combination of plain film and CT imaging can identify virtually all injuries. An exception to this is a purely ligamentous injury. These rare and dangerous injuries may not be visible with standard imaging techniques. Flexion and extension views can be performed and may reveal opening of the intervertebral space. This should only be done in the presence of an experienced surgeon: patients with injuries have become permanently quadriplegic when flexed and extended by inexperienced individuals (Fig. 6-12). A safer method may be to instruct the patient to carefully move his or her head without assistance from the surgeon; patients will not pith themselves.

Spinal cord injuries can be complete or partial. Complete injuries cause permanent quadriplegia or paraplegia, depending on the level of the injury. These patients have a complete loss of motor function and sensation two or more levels below the bony injury. Patients with high spinal cord disruption are at risk for spinal shock from physiologic disruption of sympathetic fibers. Significant neurologic recovery is rare. There are several partial or incomplete spinal cord injury syndromes. Central cord syndrome usually occurs in older persons who suffer hyperextension injuries. Motor function, pain, and temperature sensation are preserved in the lower extremities but diminished in the upper extremities. Some functional recovery usually occurs, but it is seldom a return to normal. Anterior cord syndrome is characterized by diminished motor function and pain and temperature sensation below the level of the injury. Position, vibratory, and crude touch sensation are maintained. Prognosis for recovery is poor. Brown-Séquard's syndrome usually is the result of a penetrating injury in which the right or left half of the spinal cord is transected. This rare lesion is characterized by the ipsilateral loss of motor function, proprioception, and vibratory sensation; pain and temperature sensation are lost on the contralateral side.

Penetrating injuries of the anterior neck that violate the platysma are considered significant because of the density of critical structures in this region. Mandatory exploration may be appropriate in some circumstances, but patients are now managed selectively in most centers (Fig. 6-13). Selective management is based on the neck's division into three zones (Fig. 6-14). Zone I is between the clavicles and the cricoid cartilage, and is also referred to as the thoracic outlet. Zone II is between the cricoid cartilage and the angle of the mandible, and Zone III is above the angle of mandible. The evaluation and management of visceral and vascular injuries in Zone I (the thoracic outlet) are complicated by the overlying ribs, sternum, and clavicles. Because the operative incision to be made may depend on the injured structures, a precise preoperative diagnosis is desirable. Patients with Zone I injuries should undergo angiography of the great vessels, soluble-contrast esophagram followed by barium esophagram, esophagoscopy, and bronchoscopy. Hemodynamically unstable patients should not undergo this extensive evaluation but should be taken directly to the operating room.

Patients with Zone II injuries are the easiest to evaluate. Unstable patients or those with evidence of airway compromise, an expanding hematoma, or significant external hemorrhage (including hemorrhage into the mouth) should be explored promptly. Stable patients without these findings can be evaluated selectively. Penetrating neck wounds in stable patients should be locally explored to determine the depth of penetration. Wounds that do not penetrate the platysma are insignificant and should be closed; these patients can be discharged. The vast majority of the remaining Zone II penetrating wounds are observed for 12 h. Patients with right-to-left transcervical gunshot wounds may require diagnostic studies. Carotid and vertebral angiography, direct laryngoscopy, tracheoscopy, esophagoscopy, and esophagram might be necessary, depending on the bullet's trajectory.

Patients with Zone III penetrating injuries require carotid and vertebral angiography if there is evidence of arterial bleeding. This is important for three reasons: (1) exposure of the distal internal carotid and vertebral arteries is difficult; (2) the internal carotid artery may have to be ligated, a maneuver associated with a high risk of stroke; and (3) active hemorrhage from the external carotid and vertebral arteries can be controlled by selective embolization. Associated injuries of the pharynx are inconsequential and require no special evaluation.

Chest

Blunt trauma to the chest may involve the chest wall, thoracic spine, heart, lungs, thoracic aorta and great vessels, and the esophagus. Most of these injuries are assessable by physical examination and chest x-ray. Patients with large air leaks after tube thoracostomy and those who are difficult to ventilate should undergo fiberoptic bronchoscopy to search for bronchial tears or foreign bodies.

The most threatening occult injury in trauma surgery is a tear of the descending thoracic aorta. Widening of the mediastinum on anteroposterior chest x-ray strongly suggests this injury. The widening is caused by the formation of a hematoma around the injured aorta that is temporarily contained by the mediastinal pleura. Posterior rib fractures and laceration of small vessels also can produce similar hematomas. Should the hematoma rupture into the chest with an aortic injury, the patient exsanguinates in seconds. Other findings suggestive of an aortic tear are noted in Table 6-3. This injury may be present with an entirely normal chest x-ray, although the incidence is approximately 2 percent. Because of the dire consequences of missing the diagnosis, CT and angiography are frequently performed after certain types of injury. Aortic tears occur when shearing forces are created in the chest. This is most often seen in high- energy-transfer deceleration motor vehicle accidents with frontal or lateral impact. It may occur after an ejection injury or fall. The tear usually occurs just distal to the left subclavian artery, where the aorta is tethered by the ligamentum arteriosum (Fig. 6-15). In 2 to 5 percent of cases the tear occurs in the ascending aorta, in the transverse arch, or at the diaphragm. Dynamic, spiral CT is an excellent screening test. Positive findings are a hematoma around the aorta or injury of the aorta. This test is highly sensitive, but its specificity is unknown. A clearly widened mediastinum on chest x-ray or abnormalities on CT are an absolute indication for emergent aortography.

Penetrating thoracic trauma is considerably easier to evaluate. Physical examination, plain posteroanterior and lateral chest x-rays with metallic markings of entrance and exit wounds, and CVP measurement disclose the vast majority of injuries. Injuries of the esophagus and trachea are exceptions. Depending on the estimated trajectory of the missile or blade, bronchoscopy should be performed to evaluate the trachea. Esophagoscopy can be performed to evaluate the esophagus, but injuries have been missed with the use of this technique alone. Patients at risk also should undergo a soluble contrast esophagram. If no extravasation of contrast medium is seen, a barium esophagram should be performed for greater detail. Failure to identify extant esophageal injuries leads to fulminant mediastinitis that is often fatal. As in the neck, right-to-left transmediastinal gunshot wounds frequently cause visceral or vascular injuries. Stable patients should be carefully evaluated for tracheal and esophageal injuries. Angiography occasionally is indicated.

Abdomen

With few exceptions, it is not necessary to determine which intraabdominal organs are injured, only whether an exploratory laparotomy is necessary. Physical examination of the abdomen is unreliable in making this determination, but most authorities agree that the presence of abdominal rigidity or gross abdominal distention in a patient with truncal trauma is an indication for prompt surgical exploration. For the majority of patients suffering blunt abdominal trauma, however, it is not clear whether exploration is needed. Serial examinations by the same surgeon can detect early peritoneal inflammation and the need for laparotomy before serious infections and hemorrhagic complications occur. Drugs, alcohol, or injuries of the head or spinal cord complicate physical examination. Laparotomy also may be impractical in patients who require general anesthesia for the treatment of other injuries. These patients require additional diagnostic testing.

The diagnostic approaches to penetrating and to blunt abdominal trauma differ substantially. Little preoperative evaluation is required for firearm injuries in which the peritoneal cavity is penetrated because the chance of internal injury is over 90 percent and laparotomy is mandatory. Anterior truncal gunshot wounds between the fourth intercostal space and the pubic symphysis whose trajectory as determined by x-ray or entrance/exit wound suggests peritoneal penetration should be operated on. Gunshot wounds to the back or flank are more difficult to evaluate because of the greater thickness of tissue between the skin and the abdominal organs. If in doubt, it is always safer to explore the abdomen than to equivocate when the depth of penetration is uncertain.

In contrast to gunshot wounds, stab wounds that penetrate the peritoneal cavity are less likely to injure intraabdominal organs. Anterior and lateral stab wounds to the trunk should be explored under local anesthesia in the emergency room to determine whether the peritoneum has been violated. Injuries that do not penetrate the peritoneal cavity do not require further evaluation. Stab wounds to the flank and back are more difficult to evaluate. Some authorities have recommended a triple contrast CT scan to detect occult retroperitoneal injuries of the colon, duodenum, and urinary tract. Because CT does not always identify enteric injuries, the authors have used soluble contrast radiographs of the colon and duodenum followed by barium if necessary. The larger final images may improve sensitivity. *Diagnostic peritoneal lavage (DPL)* remains the most sensitive test available for determining the presence of intraabdominal injury (Fig. 6-16). For stab wounds to the abdomen, its sensitivity for detecting intraabdominal injury exceeds 95 percent. The results of DPL are considered to be grossly positive if more than 10 mL of free blood can be aspirated after insertion of the catheter. If less than 10 mL is withdrawn, 1 L of normal saline solution is instilled and the patient is gently rocked from side to side and up and down. The effluent is withdrawn and sent to the laboratory for red blood cell count and determination of amylase and alkaline phosphatase levels. A red blood cell count greater than 100,000/mm³ is considered positive. The detection of bile, vegetable or fecal material, or the observation of effluent draining through a chest tube, a nasogastric tube, or a Foley catheter also constitutes a positive result. In equivocal cases, measurement of amylase and alkaline phosphatase levels can be helpful in identifying perforation of hollow viscera. The white blood cell count of the lavage effluent is not considered a valid indicator of intraperitoneal injury.

Stab wounds to the lower chest present a diagnostic opportunity. After the administration of adequate local anesthesia and extension of the wound as necessary, a finger is placed into the thoracic cavity to palpate the diaphragm. Confirmation of diaphragm penetration is an indication for laparotomy. When a hole is not palpable but risk of a diaphragmatic injury exists, a *DPL* should be performed. A red cell count in the effluent of more than 10,000/mm³ is considered positive when evaluating for a diaphragmatic injury. For red cell counts between 1,000 and 10,000/mm³, thoracoscopy should be considered.

Blunt abdominal trauma is evaluated by ultrasound imaging in most major trauma centers and, in selected cases, with CT scanning to refine the diagnosis. Ultrasonography performed by a surgeon or an emergency physician in the emergency room has largely replaced *DPL*. Evaluation of the entire abdomen is not the goal, but ultrasound is used in specific anatomic regions (e.g., Morison's pouch, the left upper quadrant, the pelvis) to identify free intraperitoneal fluid (Fig. 6-17). This method is exquisitely sensitive for detecting intraperitoneal fluid collections larger than 250 mL, but it is relatively poor for staging solid organ injuries. DPL is appropriate for patients whose condition cannot be explained with ultrasound imaging.

The use of CT scanning for the diagnosis of blunt abdominal trauma gained considerable popularity in the early 1980s. It was reported that injuries of the liver, spleen, and kidneys could be diagnosed with great precision using this method. Much of the initial enthusiasm has been tempered by the recognition of several limitations: (1) the need for high-quality radiographs, (2) the need for radiologists skilled in the interpretation of postinjury CT images, (3) the need for proper patient preparation, (4) poor sensitivity for intestinal injuries and acute pancreatic injuries, and (5) relatively poor correlation between splenic and hepatic CT images and the subsequent risk of bleeding requiring an operation. Despite these limitations, CT is an important diagnostic tool because of its specificity for hepatic, splenic, and renal injuries (Fig. 6-18). CT is indicated primarily for hemodynamically stable patients who are candidates for nonoperative therapy. CT also is indicated for hemodynamically stable patients who have unreliable physical examinations or other conditions (i.e., intracranial injury) requiring CT evaluation. The algorithm the authors use to evaluate blunt abdominal trauma is outlined in Fig. 6-19.

Because of the excellent view provided of the liver and anterior diaphragm, laparoscopy seems to be an ideal diagnostic tool for stable patients who have possible anterior upper abdominal injuries. One theoretic concern is carbon dioxide gas embolism through injuries of the hepatic veins, but this potential complication can be eliminated with gasless laparoscopy. The role of laparoscopy remains to be clarified, but it may expand with the availability of a smaller laparoscope that can be inserted under local anesthesia.

Pelvis

Blunt injury to the pelvis frequently produces complex fractures (Fig. 6-20). Plain x-rays reveal gross abnormalities, but CT scanning may be necessary to assess the pelvis for stability. Sharp spicules of bone can lacerate the rectum or vagina. The finding of gross blood on digital examination strongly suggests injury to these organs. Proctoscopy or speculum examination may reveal the injury. In questionable cases, soluble-contrast x-rays are diagnostic. The bladder can be lacerated by sharp fracture fragments, or, if the bladder is full, a direct blow to the hypogastrium can generate sufficient intravesicular pressure to cause rupture. Gross blood on urinalysis may not always occur, and a cystogram should be performed if more than a few red cells per high-power field are seen on urinalysis. Urethral injuries are suspected by the findings of blood at the meatus, scrotal or perineal hematomas, and a high-riding prostate on rectal examination. Urethrograms should be done in stable patients before placing the Foley catheter to avoid false passage and subsequent stricture.

Major vascular injuries are uncommon in blunt pelvic trauma; however, thrombosis or disruption of the arteries or veins in the iliofemoral system may occur. Angiography is indicated if thrombosis of the arterial system is suspected. Evaluation of penetrating injuries of the pelvis is similar to that for blunt injuries in stable patients. Visceral and vascular injuries are much more common, and laparotomy is often required. Life-threatening hemorrhage can be associated with pelvic fractures. The source may be the lower lumbar arteries and veins or branches of the internal iliac arteries and veins. These injuries are frequently not amenable to surgical repair and usually occur with disruption of the posterior elements of the pelvis.

Extremities

Injury of the extremities from any cause requires plain x-ray films to evaluate fractures. Ligamentous injuries, particularly those of the knee and shoulder when related to sports activities, can be imaged with MRI. The assessment of vascular injuries is somewhat controversial. Vascular diagnosis usually is limited to the arterial system unless there is uncontrolled external venous hemorrhage or venous injuries are uncovered during operative exploration. It is uncommon to see a patient with a venous complication related to trauma that was not identified and treated while an arterial injury was being evaluated.

Physical examination serves to identify and localize arterial injuries in many instances. Physical findings are classified as hard signs or soft signs (Table 6-4). Hard signs constitute indications for operative exploration, whereas soft signs are indications for observation or additional testing. Arteriography may be helpful in localizing the injury in some patients with penetrating injuries and hard signs. For example, a bullet that enters the lateral hip and exits below the knee medially, when there is also a femoral shaft fracture and no popliteal pulse, could have injured the femoral or popliteal artery anywhere. Arteriography would be useful to localize the injury and limit the dissection.

The controversy in vascular trauma is in the management of patients with soft signs of injury, particularly injuries that are in proximity to major vessels. It is known that some of these patients will be found to have arterial injuries that require repair. One approach has been to measure systolic blood pressures using Doppler ultrasonography and compare the injured side to the uninjured side. If the pressures are within 10 percent of each other, a significant injury is excluded and no further evaluation is performed. If the difference is greater than 10 percent, an arteriogram is indicated. Some argue that there are occult injuries, such as pseudoaneurysms or injuries of the profunda femoris or peroneal arteries, which may not be sampled with this technique. If hemorrhage occurs from these injuries, compartment syndrome and limb loss may occur. While trauma centers debate this issue, the surgeon who is treating the injured patient can perform angiography in selected patients who have soft signs.

TREATMENT

General Considerations

Seriously injured patients are more fragile than their age and physical condition imply. Procedures that are well tolerated in older patients, such as hepatic lobectomy for hepatoma, usually are lethal in the multiply injured patient. There has been a remarkable change in operative approach over the past 20 years. Faster techniques are used, and shorter serial operations have become common. For example, at the authors' institution virtually all suture lines are created with a running single layer. There is no evidence that this method is less secure than interrupted multilayer techniques, and it is clearly faster. Drains, once considered mandatory for many parenchymal injuries and some anastomoses, have virtually disappeared. Fluid collections that accumulate in a delayed fashion are now effectively managed by interventional radiologists. Injuries once thought to mandate resection, such as splenic injuries, are now managed with suture repair or even nonoperatively. The treatment of colonic injuries by primary repair is another example. These conceptual changes have significantly improved survival in trauma patients, and all have been developed through the extensive experience of major urban trauma centers and the forums for the free exchange of ideas provided by the American College of Surgeons Committee on Trauma, the American Association for the Surgery of Trauma, the International Association of Trauma and Surgical Intensive Care, the Pan-American Trauma Congress, and other surgical organizations.

The management of patients with multiple injuries requires the early establishment of therapeutic priorities. While the concept of life over limb and limb over cosmesis seems obvious, decision making can be subtle. Many combinations of injuries and physiologic states can be anticipated that have an impact on decision making. It is the trauma surgeon's responsibility to assess the probable outcomes of different therapeutic strategies and choose which is in the patient's best interest.

Transfusion

Because fresh whole blood, the optimal replacement material for shed blood, is no longer available, whole blood must be recreated from its component parts: *packed red blood cells (pRBC)*, *fresh-frozen plasma (FFP)*, and platelet packs. Not all trauma patients requiring transfusions receive all three components. Most trauma patients receive between 1 and 5 units of pRBC and no other components, but major trauma centers have the capability of transfusing tremendous quantities of blood components. It is not unusual for 100 component units to be transfused during one procedure. Red cell transfusion rates of 20 to 40 units of pRBC per hour are common in severely injured patients.

Transfusion practices in trauma require the surgeon to identify the insidious signs of coagulopathy, such as excessive bleeding from the cut edges of skin, fascia, and peritoneum that were previously controlled. While the local volume of coagulopathic hemorrhage in one visual field seems low compared to that of a hole in the aorta or venae cavae, blood loss from the entire area of dissection can lead to exsanguination. The usual measurements of coagulation capability, i.e., prothrombin time (PT), partial thromboplastin time (PTT), and platelet count have a turnaround time of more than 30 min in most institutions. These tests are of limited value in patients who have lost two or three blood volumes while waiting for test results. Under such conditions, transfusion must be empiric and based on the surgeon's observations. At the first sign of coagulopathic hemorrhage, the previously lost plasma proteins and platelets must be restored with *FFP* and platelet packs. Additional transfusions should be administered with equal ratios of *pRBC*, FFP, and platelets.

The causal relationship of core hypothermia metabolic, acidosis, and postinjury coagulopathy has been observed in a number of studies. The pathophysiology is multifactorial and includes inhibition of temperature-dependent enzyme-activated coagulation cascades, platelet dysfunction, endothelial abnormalities, and a poorly understood fibrinolytic activity. The role of metabolic acidosis in the pathogenesis of a coagulopathy is unclear. Experiments have demonstrated impaired hemostasis at a pH of 7.20; others have suggested that pH directly affects platelet function. Other series implicated acidosis in the propagation of disseminated intravascular coagulation with the secondary consumption of clotting factors. Hypothermia and metabolic acidosis have adverse effects on myocardial performance and tissue perfusion.

Primary hemostasis relies on platelet adherence and aggregation to injured endothelium, resulting in the formation of the platelet plug. A platelet count of 50,000/mm³ is considered adequate for tissue hemostasis if the platelets are normal. However, platelet dysfunction is a well-documented complication of massive transfusion that is aggravated by associated hypothermia. Consequently, the recommended target of more than 100,000/mm³ for platelet transfusion in other high-risk patients should be extended to the severely injured.

Blood typing and, to a lesser extent, crossmatching is essential to avoid life-threatening intravascular hemolytic transfusion reactions. A complete type and crossmatch requires 20 to 45 min to complete and reduces the risk of an intravascular hemolysis to approximately 0.004 percent. If 20 units of *pRBC* are needed within an hour, an army of technicians would be required to perform this service. Twenty to 45 min is too long for a patient with an exsanguinating hemorrhage to wait. Therefore, trauma patients requiring emergency transfusions are given type O, type-specific, or biologically compatible red blood cells. As a cross-check for ABO compatibility, a saline crossmatch is often performed.

The administrative and laboratory time required is approximately 5 min, and the risk of intravascular hemolysis is about 0.05 percent. The risk increases to 1.0 percent with a history of previous transfusions or pregnancy, and up to 3.0 percent with both. This increased risk of transfusion reaction is a result of the presence of irregular antibodies (e.g., Kell, Duffy, Kidd, etc.) in the patient's plasma that occur in about 1/1000 patients. Intravascular hemolysis can occur with ABO-compatible *pRBC* if the patient has an irregular antibody. It usually is not as severe as ABO incompatible hemolytic reactions, and the time required to detect the antibodies biochemically or by crossmatch makes the increased risk of hemolytic reaction a reasonable trade for rapid availability. Preformed antibodies are rapidly depleted by hemorrhage and are produced slowly, diminishing the severity of intravascular hemolysis if it occurs. An alternative strategy for those patients who are consistently stable and do not have serious injuries is to perform a type and screen as a cost-saving measure. If blood is subsequently needed urgently, low-titer, type-specific red cells can be administered with the same risk of intravascular hemolysis as with fully typed and crossmatched blood, provided the screen for irregular antibodies is negative. Unstable patients should receive O-negative, O-positive, or type-specific red cells, depending on the patient's age and sex and the availability of blood cell types. Other components should be type specific or biologically compatible.

Prophylaxis

All injured patients undergoing an operation should receive preemptive antibiotic therapy. The authors use second-generation cephalosporins for laparotomies and first-generation cephalosporins for all other operations. Additional doses should be administered during the procedure on the basis of blood loss and the half-life of the antibiotic. The role of postoperative antibiotic therapy in trauma patients remains to be defined, but the trend has been to reduce the duration. Tetanus prophylaxis is administered to all patients according to the American College of Surgeons guidelines (see Chap. 5).

Deep venous thrombosis and other venous complications occur more often in injured patients than is generally believed. This is particularly true for patients with major fractures of the pelvis and lower extremities, those with spinal cord injury or in a coma, and those with injury of the large veins in the abdomen and lower extremities. The authors use pulsatile compression stockings in all injured patients and selectively place inferior vena caval filters for those at very high risk. The role of inferior vena caval filters may expand in the future when removable devices become commercially available. Low-molecular-weight heparins have been demonstrated to be safe and effective in patients with orthopaedic injuries. Their use in patients with other injuries remains to be elucidated.

Another prophylactic measure is thermal protection. Hemorrhagic shock impairs perfusion and metabolic activity throughout the body. With declining metabolism, heat production and body temperature decrease. The injured patient receives a second thermal insult with the removal of insulating clothing. As a result, trauma patients can become seriously hypothermic, with temperatures as low as 34°C by the time they reach the operating room. Hypothermia impairs coagulation and myocardial contractility and increases myocardial irritability. Intentional hypothermia has protective features for patients with massive head injuries, but most authorities agree that the deleterious effects outweigh the potential benefits. Injured patients whose intraoperative core temperature drops below 32°C are at risk for fatal arrhythmias and defective coagulation. Thermal prophylaxis should begin in the emergency room by maintaining the ambient temperature comfortable for an exposed patient. Fluids should be stored at body temperature and blood products should be administered through rapid-warming devices. When examination is completed, the patient should be kept scrupulously covered with warm blankets or other devices until body temperature returns to normal.

Vascular Repair

The initial control of vascular injuries should be accomplished digitally by applying enough pressure directly on the bleeding site to stop the hemorrhage. Some bleeding vessels may need to be gently pinched between the thumb and index finger. These maneuvers, along with suction, usually create a dry enough field to safely permit the dissection necessary to define the injury. Sharp dissection with fine scissors is preferable to blunt dissection because the latter can aggravate the injury. When a sufficient length of vessel is available, a vascular thumb forceps is used to grasp the vessel. If the vessel is not transected, forceps can be placed directly across the injury. This minimizes or eliminates bleeding while the dissection necessary for clamping is completed. If the vessel is transected (or nearly so) digital control is maintained on one side while the other is occluded with a thumb forceps. The vessel is then sharply mobilized to allow an appropriate vascular clamp to be applied. When definitive control of all injuries is achieved, heparinized saline is injected into the proximal and distal ends of the injured vessel to prevent thrombosis. The exposed intima and media at the site of the injury are highly thrombogenic, and small clots often form. These clots should be carefully removed to prevent thrombosis or embolism when the clamps are removed. Because of the frequency that embolism occurs, routine balloon catheter exploration of the distal vessel has been recommended. Ragged edges of the injury site should be judiciously debrided using sharp dissection.

Injuries of the large veins such as the venae cavae, or the innominate and iliac veins pose a special problem for hemostasis. Numerous large tributaries make adequate hemostasis difficult to achieve, and their thin walls render them susceptible to additional iatrogenic injury. When a large-vessel injury is encountered, tamponade with a folded laparotomy pad held directly over the bleeding site usually establishes hemostasis sufficient to prevent exsanguination. If hemostasis is not adequate to expose the vessel proximal and distal to the injury, sponge sticks can be placed strategically on either side of the injury and carefully adjusted to improve hemostasis. This maneuver requires skill and discipline to maintain a dry field. The operative field is sometimes sufficient to delineate and repair the injury. It is often difficult for the assistant to maintain complete control of hemorrhage with sponge sticks. In this situation, the vessel can be exposed on either side of the sponge stick and a vascular clamp applied. The clamp can then be sequentially advanced toward the injury until hemostasis is complete.

Options for the treatment of vascular injuries are listed in Table 6-5. Some arteries and most veins can be ligated without significant sequelae. Arteries for which repair should always be attempted include the aorta and the carotid, innominate, brachial, superior mesenteric, proper hepatic, renal, iliac, femoral, and popliteal arteries. In the forearm and lower leg at least one of the two palpable vessels should be salvaged. The list of veins for which repair should be attempted is short: the superior vena cava, the inferior vena cava proximal to the renal veins, and the portal vein. There are notable vessels for which repair is not necessary, e.g., the subclavian artery and the superior mesenteric vein. The portal vein can be ligated successfully provided adequate fluid is administered to compensate for the dramatic but transient edema that occurs in the bowel. Ligation of some vessels, such as the popliteal vein and the left or right branch of the portal vein, can result in morbidity for the patient that is not life threatening. The authors attempt to repair all arteries larger than 3 mm and all veins larger than 10 mm in diameter, depending on the patient's physiologic condition.

Some arterial injuries have been treated by observation without subsequent complications. These include small pseudoaneurysms, intimal dissections, small intimal flaps and arteriovenous fistulas in the extremities, and occlusions of small (<2 mm) arteries. Follow-up angiography is obtained within 2 to 4 weeks to ensure that healing has occurred.

Lateral suture is appropriate for small arterial injuries with little or no loss of tissue. End-to-end anastomosis is used if the vessel is transected or nearly so. The severed ends of the vessel are mobilized, and small branches are ligated and divided as necessary to obtain the desired length. Arterial defects of 1 to 2 cm usually can be bridged. The surgeon should not be reluctant to divide small branches to obtain additional length because most injured patients have normal vasculature and the preservation of potential collateral flow is not as important as in atherosclerotic surgery. To avoid postoperative stenosis, particularly in smaller arteries, some technique such as beveling or spatulation should be used so that the

completed anastomosis is slightly larger in diameter than the native artery (Fig. 6-21).

Interposition grafts are used when end-to-end anastomosis cannot be accomplished without tension despite mobilization. For vessels less than 6 mm in diameter, autogenous saphenous vein from the groin should be used because *polytetrafluoroethylene (PTFE)* grafts less than 6 mm in diameter have a prohibitive rate of thrombosis. Injuries of the brachial, popliteal, and internal carotid arteries require the saphenous vein for interposition grafting. When the saphenous vein is harvested for treating an arterial injury in the lower extremity, it should be taken from the contralateral extremity. Because the status of the ipsilateral venous system is unknown, the saphenous vein on that side may become an important tributary. Larger arteries must be bridged by artificial grafts. Some authorities advocate the use of free internal iliac artery grafts because of the greater thickness and strength of its wall compared to the saphenous vein. The authors believe that this vessel is overly tedious to remove and has no advantage over the saphenous vein.

Transposition procedures can be used when an artery has a bifurcation of which one vessel can safely be ligated. Injuries of the proximal internal carotid can be treated by mobilizing the adjacent external carotid, dividing it distal to the internal injury and performing an end-to-end anastomosis between it and the distal internal carotid (Fig. 6-22). The proximal stump of the internal carotid is oversewn in such a way as to avoid a blind pocket where clot may form. Injuries of the ipsilateral external and contralateral common iliac arteries can be handled in a similar fashion provided flow is maintained in at least one internal iliac artery (Fig. 6-23).

Arterial injuries are often grossly contaminated from enteric or external sources, in which case many surgeons are reluctant to place artificial grafts in situ. This situation arises most often in injuries to the aortic or iliac artery when the colon also is injured. For the aorta there are few options. Ligation of the aorta with unilateral or bilateral axillofemoral bypass can be performed. These are lengthy procedures that are prone to thrombosis and infection. Most patients who require an aortic graft cannot tolerate surgery for the amount of time required to perform an axillofemoral bypass. Therefore, even in the presence of fecal contamination, it is common practice to use *PTFE* or Dacron in situ for aortic injuries. Every effort is made to remove and control contamination after the control of hemorrhage but before the graft is brought into the operative field. This includes copious irrigation of the abdominal cavity and changing of drapes, gowns, gloves, and instruments. After placement of the graft, it is covered with peritoneum or omentum before definitive treatment of the enteric injuries. Graft infection is rare in these instances. A similar approach can be used for injuries to the iliac artery, but in most cases this can be avoided by the innovative use of transposition procedures.

Suture selection for arterial injuries is based on the diameter of the vessel being repaired (Table 6-6). The use of progressively finer suture for smaller- diameter vessels encourages the inclusion of less tissue with more closely placed sutures, which is necessary for successful repair. When performing anastomoses in which the vessels are tethered, e.g., the thoracic artery and the abdominal aorta, the authors use the parachute technique to ensure precise placement of the posterior suture line (Fig. 6-24). If this technique is used, traction on both ends of the sutures must be maintained or leakage from the posterior aspect of the suture line is probable. A single temporary suture 180 degrees from the posterior row is used to maintain alignment.

Venous injuries are more difficult to repair successfully because of their propensity to thrombose. Small injuries without loss of tissue can be treated with lateral suture. More complex repairs often fail. Thrombosis does not occur acutely but rather gradually over 1 to 2 weeks. Adequate collateral circulation, sufficient to avoid acute venous hypertensive complications, usually develops within several days. Therefore it is reasonable to use *PTFE* for venous interposition grafting and accept a gradual but eventual thrombosis while waiting for collateral circulation to develop. Conversely, chronic venous hypertensive complications in the lower extremities often can be avoided with any level of ligation by (1) elastic bandages carefully applied in the operating room at the end of the procedure, and (2) continuous elevation of the lower extremities to 30 degrees. These measures should be maintained for 1 week, after which the patient is ambulated. If no edema occurs with the bandages removed, elevation is no longer necessary. It is a reasonable precaution to have the patient wear compressive stockings up to the knee for a few months afterward.

There are several circumstances in which a more aggressive approach should be considered. Ligation of the superior vena cava has been associated with sudden blindness resulting from compression of the optic nerve from venous hypertension. Ligation of the suprarenal inferior vena cava is believed to be associated with acute renal failure from venous hypertension. Chronic venous insufficiency of the lower extremities may be caused by ligation of the infrarenal vena cava or any in-line vein below that level, particularly the popliteal vein. Interposition grafting can be considered in these situations, but the choice of material is problematic. One option is to use artificial material because it is rapidly available in hemodynamically correct sizes. The drawback is that thrombosis is inevitable when such grafts are placed below the renal veins. Artificial grafts have performed satisfactorily in cases of suprarenal inferior vena caval and superior vena caval replacement. The jugular vein can be used to replace vessels of similar size, e.g., the portal or femoral vein. The saphenous vein is too small to replace any important vein. Panel grafts and spiral grafts constructed around a mandrel (chest tube) using saphenous vein have occasionally been performed, but these procedures are extremely tedious and have no apparent advantage over ligation in most instances.

The technology used by interventional radiologists is advancing rapidly. They have the ability to cannulate virtually any artery in the body and dilate it, place an intraluminal filter, stent, or graft in it, or occlude it. Their services are most valuable for treating arterial or venous injuries that are surgically inaccessible, such as stent placement in the internal carotid artery near or in the base of the skull, or controlling hemorrhage in hepatic injuries or pelvic fractures (Figs. 6-25, 6-26, and 6-27).

Staged Operations

The most common causes of death for trauma patients are head injury, exsanguination from cardiovascular injuries, and sepsis with multiple organ failure. Another cause of death has become apparent as the capability of delivering massive quantities of red blood cells and other components has developed. Surgeons are able to operate on the most severely injured patients until a constellation of metabolic derangements develops. These are characterized by the triad of an obvious coagulopathy, profound hypothermia, and metabolic acidosis. Hypothermia from evaporative and conductive heat loss and diminished heat production occurs in spite of warming blankets and blood warmers. The metabolic acidosis of shock is exacerbated by aortic clamping, vasopressors, massive transfusions, and impaired myocardial performance. Coagulopathy is caused by dilution, hypothermia, and acidosis. Each of these factors reinforces the others, resulting in a critically ill patient who is at high risk for a fatal arrhythmia. This downward spiral has been referred to as “the bloody vicious cycle” (Fig. 6-28).

Heat loss appears to be the central event because neither of the other components can be corrected until core temperature returns toward normal. Laboratory and mathematical heat exchange models have demonstrated that evaporative heat loss from an open abdomen is by far the greatest source. A concomitant open thoracic cavity greatly accelerates the rate of the patient's deterioration and can cause the syndrome by itself. This is the rationale for the immediate abdominal closure and the reason it has been successful.

Staged operations are indicated when a coagulopathy develops and core temperature drops below 34°C. A refractory acidosis is almost always present. Several unorthodox techniques can be used to expedite wound closure. Bleeding raw surfaces, often of the liver, are packed with laparotomy pads. Small enteric injuries are closed with staples, and large ones are stapled on both sides with the GIA stapler and the damaged segment removed. Clamps may be left on unrepaired vascular injuries, or the vessels may be ligated. Injuries of the pancreas and kidneys are not treated if they are not bleeding. No drains are placed, and the abdomen is closed with sharp towel clips placed 2 cm apart, which include only the skin (Fig. 6-29). Towel clips are used because they do not cause bleeding as needles do, and they can be applied very rapidly, usually in 60 to 90 seconds. The closure of just the skin allows for the abdominal or thoracic cavity to accommodate a greater volume without increased pressure. The clips are covered with a towel, and a plastic adhesive sheet is placed over the towel to prevent excessive fluid from draining onto the patient's bedding. Cold wet drapes are removed, and the patient is covered from head to toe with layers of warm blankets. Some of the unorthodox treatments used, including the creation of closed-loop bowel obstructions and unrepaired renal injuries, are not compatible with survival; however, reoperation is planned within 2 to 24 h, and the treatments are tolerated well within that time frame. The goal is to complete the procedure as soon as possible, or the patient will die. If the surgeon believes that the patient's metabolic problems can be corrected in a short time (2 h or less), the patient can remain in the operating room while additional blood products are administered and rewarming measures are instituted. Patients who are in very poor condition and require several hours for metabolic corrections should be transferred to the surgical intensive care unit. If the patient's condition improves as evidenced by normalization of coagulation studies, the correction of acid/base imbalance, and a core temperature of at least 36°C, the patient should be returned to the operating room for removal of packs and definitive treatment of injuries.

There are several complications associated with this treatment. Failure to identify noncoagulopathic hemorrhage can lead to exsanguination. Most patients with coagulopathic hemorrhage have a gradual decrease in the need for *pRBC*, *FFP*, and platelets and an improvement in coagulation studies as temperature rises. In the case of vascular hemorrhage coagulopathy does not correct itself, and these patients must be returned to the operating room for reexploration.

A second complication is referred to as the abdominal or thoracic compartment syndrome, which are caused by an acute increase in intracavitary pressure. In the abdomen the compliance of the abdominal wall and the diaphragm permit the accumulation of many liters of fluid before *intraabdominal pressure (IAP)* increases. There are primarily two sources for this fluid, blood, and edema. Blood accumulates as a result of the coagulopathy or missed vascular injury described above. The cause of edema is multifactorial. Ischemia and reperfusion cause capillary leakage, loss of oncotic pressure occurs, and in the case of the small bowel, which is often eviscerated, prolongation and narrowing of veins and lymphatics caused by traction impairs venous and lymphatic drainage.

The resulting edema may be dramatic (Fig. 6-30). Similar phenomena occur in the chest. As fluid continues to accumulate, the compliant limit of the abdominal cavity is eventually exceeded, and *IAP* increases. When *IAP* exceeds 15 mmHg, serious physiologic changes begin to occur. The lungs are compressed by the upward displacement of the diaphragm. This causes a decrease in functional residual capacity, increased airway pressure, and, ultimately, hypoxia. Cardiac output decreases as a consequence of diminished venous return to the heart and increased afterload. Blood flow to every intraabdominal organ is reduced because of increased venous resistance. As *IAP* exceeds 25 to 30 mmHg, life-threatening hypoxia and anuric renal failure occur. Cardiac output is further reduced but can be returned toward normal with volume expansion and inotropic support.

The only method for treating hypoxia and renal failure is to decompress the abdominal cavity by opening the incision. This results in an immediate diuresis and a resolution of hypoxia. Failure to decompress the abdominal cavity eventually causes lethal hypoxia or organ failure. There have been a few reports of sudden hypotension when the abdomen is opened, but volume loading to enhance cardiac output has largely eliminated this problem. *IAP* is measured using the Foley catheter. Because the bladder is a passive reservoir at low volumes (50 to 100 mL), it imparts no intrinsic pressure but can transmit *IAP*. Fifty mL saline solution is injected into the aspiration port of the urinary drainage tube with an occlusive clamp placed across the tube just distal to the port. The saline is used to create a standing column of fluid between the bladder and port that can transmit *IAP* to a recording device. The needle in the port is connected to a *CVP* manometer using a three-way stopcock. The manometer is filled with saline and opened to the drainage tube. *IAP* is read at the meniscus with manometer zeroed at the pubic symphysis. Bladder pressures measured in this fashion are reliable and consistent. Pressures less than 15 mmHg do not require decompression. Table 6-7 lists recommendations according to *IAP*.

In the chest similar phenomena occur. Edema of the heart and lungs develops, and the heart also may dilate. Blood accumulation is rarely a problem because of the use of chest tubes. The diagnosis usually is apparent in the operating room because the heart tolerates compression poorly. Attempts to close the chest in this setting are associated with profound hypotension, and an alternative method of closure is necessary. The most popular material used to accommodate the addition of volume in the chest or abdomen is a 3-liter plastic urologic irrigation bag that has been cut open and sterilized. The bag is sewn to the skin or fascia using No. 2 nylon suture with a simple running technique (Fig. 6-31). As many as four bags may need to be sewn together to cover a large defect. Closed- suction drains are placed beneath the plastic to remove blood and serous fluid that inevitably accumulate. The entire closure is covered with an iodinated plastic adhesive sheet to simplify nursing care. Patients whose renal function has not been impaired will have a

remarkable diuresis. Exogenous fluids are held to a minimum to facilitate the resolution of edema. Definitive wound closure can usually be performed in 48 to 72 h.

In the case of patients who develop sepsis and multiple organ failure (MOF), edema does not resolve until sepsis and MOF have resolved. This may require several weeks. The bags have been left in place up to 3 weeks with the patient surviving. The authors make every effort to close the skin over the viscera to decrease protein and heat loss and to inhibit infection. If these attempts are unsuccessful and the abdomen remains open with granulating tissue exposed, lateral traction forces of the abdominal wall eventually cause an enteric fistula. The risk of developing a fistula increases rapidly after 2 weeks with an open abdomen. These problems are extremely difficult to treat. Several approaches have been used to avoid this catastrophic complication, including polyglycolic acid or polypropylene mesh sewn to the fascia, split-thickness skin grafts placed directly on the bowel, musculocutaneous flaps, and traction devices. Of these options, skin grafts have the greatest success rates, although the abdominal wall hernia eventually requires reconstruction.

Nonoperative Management

Nonoperative treatment for blunt injuries of the liver, spleen, and kidneys is now the rule rather than the exception. Up to 90 percent of children and 50 percent of adults are treated in this manner. As interventional radiology continues to advance, these numbers will increase. The primary requirement for this therapy is hemodynamic stability. The extent of the patient's injuries should be delineated by CT scanning. Recurrent hemorrhage from the liver and kidneys has been infrequent, but delayed hemorrhage or rupture of the spleen is an important consideration in the decision to pursue nonoperative management. The patient should be monitored in the intensive care unit for the first 24 h. Because CT misses some enteric injuries, frequent abdominal examination should be performed. Usually the fall in hematocrit level stabilizes within 24 h. If the hematocrit level continues to fall, angiography with embolization of bleeding sites should be considered, particularly for hepatic and renal injuries. CT scanning usually is repeated at least once during the hospitalization to assess major hepatic or splenic lesions requiring transfusion. Gradually increasing activity is permitted after discharge. Patients involved in contact sports should have complete healing of the injury documented radiographically before resuming participation. This can take several months.

Complications of nonoperative treatment include continuing hemorrhage, delayed hemorrhage, necrosis of liver, spleen, or kidney from embolization, abscess, biloma, and urinoma. Hemorrhage may be treated by interventional radiology, though open operative control often is necessary. Most infectious complications can be treated by percutaneous drainage. Bilomas usually are resorbed.

Head

General principles for the management of cerebral injuries have changed in recent years. Attention is now focused on maintaining or enhancing cerebral perfusion rather than merely lowering intracranial pressure (ICP). Hyperventilation to a pCO₂ below 30 mmHg to induce cerebral vasoconstriction exacerbates cerebral ischemia in spite of decreasing ICP. These secondary iatrogenic cerebral injuries cause more harm than previously appreciated. Other treatments or conditions that must be avoided include decreased cardiac output because of the excessive use of osmotic diuretics, sedatives, or barbiturates, and hypoxia. Nevertheless, the measurement of ICP is important and is efficiently accomplished with a ventriculostomy tube. The tube also permits the withdrawal of cerebrospinal fluid, which is the safest method for lowering ICP. Although an ICP of 10 mmHg is believed to be the upper limit of normal, therapy is not usually initiated until the ICP reaches 20 mmHg. *Cerebral perfusion pressure (CPP)*, which is equal to the mean arterial pressure (MAP) minus the ICP, is an important measurement that is used to monitor therapy. The lowest acceptable CPP is 60 mmHg. This figure can be adjusted by lowering ICP or raising MAP. Induced paralysis, sedation, osmotic diuresis, and barbiturate- induced coma are used. The goal of fluid therapy is to achieve a euvolemic state, and arbitrary fluid restriction is avoided. It is unclear whether outcome is improved by boosting MAP with pressors or inotropes in patients with an elevated ICP resistant to treatment.

Indications for operative intervention for space-occupying hematomas are based on the amount of midline shift, the location of the clot, and the patient's ICP. A shift of more than 5 mm usually is considered an indication for evacuation. This is not an absolute rule, however. Smaller hematomas causing less shift in treacherous locations, such as the posterior fossa, can require drainage because of the threat of brainstem compression or herniation. Removal of small hematomas also may improve ICP and *CPP* in patients with an elevated ICP that is refractory to medical therapy.

The treatment of diffuse axonal injury includes the control of cerebral edema and general supportive care. The authors frequently use percutaneous tracheostomy for airway control and percutaneous endoscopic gastrostomy for enteral access in head-injured patients whose recovery is unlikely or prolonged. Prognosis is related to Glasgow Coma Scale score. Serious head injuries, GCS 3–8, have a poor prognosis, and an institutional existence is almost a certainty. Mild brain injuries, GCS 13–15, have a good prognosis; independent living is probable, but neuropsychiatric testing often reveals significant abnormalities.

General surgeons in small or rural communities without emergency neurosurgical coverage may be required to drill a burr hole in one life- saving circumstance: in a patient with an epidural hematoma. As blood from a torn vessel, usually the middle meningeal artery, accumulates, the temporal lobe is forced medially, which compresses the third cranial nerve and eventually the brainstem. The typical course is: (1) initial loss of consciousness; (2) awakening and a lucid interval; (3) recurrent loss of consciousness with a unilaterally fixed, dilated pupil; and (4) cardiac arrest. These patients usually do not have a serious underlying cortical injury, and complete recovery often is possible. The burr hole should be made on the same side as the dilated pupil, as shown in Fig. 6-32. The goal of the procedure is not to control the hemorrhage but to decompress the intracranial space. A craniotomy is required for the control of hemorrhage. The patient's head should be loosely wrapped with a thick layer of gauze to absorb the bleeding, and the patient should be transferred to a facility with emergency neurosurgical capability for a craniotomy.

Neck

Blunt Injury Cervical Spine

Treatment of injuries to the cervical spine is based on the level of injury, the stability of the spine, the presence of subluxation, the extent of angulation, and the extent of neurologic deficit. Cautious axial traction in line with the mastoid process is used to reduce subluxations. A halo-vest combination can accomplish this and also provide rigid external fixation for definitive treatment when left in place for 3 to 6 months. This device is the treatment of choice for many cervical spine injuries. Surgical fusion usually is reserved for those with neurologic deficit, those who demonstrate angulation greater than 11 degrees on flexion and extension x-rays, or those who are unstable after external fixation.

Spinal Cord

Injuries of the spinal cord, particularly complete injuries, are essentially untreatable. Approximately 3 percent of patients who present with flaccid quadriplegia have concussive injuries, and these patients represent the very few who seem to have miraculous recoveries. A prospective randomized study comparing methylprednisolone with placebo demonstrated a significant improvement in outcome (usually one or two spinal levels) for those who received the corticosteroid within 8 h of injury. The standard dosage is 30 mg/kg given as an intravenous bolus followed by a 5.4 mg/kg infusion administered over the next 23 h.

Larynx

The larynx may be fractured by a direct blow, which can result in airway compromise. A hoarse voice in a trauma patient is highly suggestive of laryngeal fracture. In cases of severe fracture a cricothyroidotomy or tracheostomy should be performed to protect the airway. The larynx is repaired with fine wires and sutures. If direct repair of internal laryngeal structures is necessary, the thyroid cartilage is split longitudinally in the midline and opened like a book. This is referred to as a laryngeal fissure.

Carotid and Vertebral Arteries

Blunt injury to the carotid or vertebral arteries may cause dissection, thrombosis, or pseudoaneurysm. More than half the patients with such injuries have a delayed diagnosis. Facial contact resulting in hyperextension and rotation appears to be the mechanism of injury. To reduce delayed recognition, the authors use CT angiography in patients at risk to identify these injuries before neurologic symptoms develop. The injuries frequently occur at or extend into the base of the skull and usually are not surgically accessible. Accepted treatment for thrombosis and dissection is anticoagulation therapy with heparin followed by warfarin sodium (Coumadin) for 3 months. Pseudoaneurysms also occur near the base of the skull. If they are small, they can be followed with repeat angiography. If enlargement occurs, consideration should be given to the placement of a stent across the aneurysm by an interventional radiologist. Another possibility is to approach the intracranial portion of the carotid artery by removing the overlying bone and performing a direct repair. This method has only recently been described and has been performed in a limited number of patients.

Venous Injuries

Thrombosis of the internal jugular veins caused by blunt trauma can occur unilaterally or bilaterally. These injuries usually are discovered incidentally and are generally asymptomatic. Bilateral thrombosis can aggravate cerebral edema in patients with serious head injuries. Stent placement should be considered in such patients if their ICP remains elevated. Laryngeal edema resulting in airway compromise also can occur.

Penetrating Injuries

Penetrating injuries in Zone II or III that require operative intervention are explored using an incision along the anterior border of the sternocleidomastoid muscle. If bilateral exploration is necessary, the inferior end of the incision can be extended to the opposite side. Midline wounds or significant bilateral injuries can be exposed via a large collar incision at the appropriate level. Alternatively, bilateral anterior sternocleidomastoid incisions can be used.

Carotid and Vertebral Arteries

Exposure of the distal internal carotid artery in Zone III is difficult (Fig. 6- 33). The first step is to divide the ansa cervicalis and mobilize the hypoglossal nerve. Next, the portion of the posterior belly of the digastric muscle that overlies the internal carotid artery is resected. The glossopharyngeal and vagus nerves are mobilized and retracted. If accessible, the styloid process and attached muscles are removed. At this point anterior displacement of the mandible may be helpful, and various methods for accomplishing this have been devised. Some authorities have advocated division and elevation of the vertical ramus, but two remaining structures still prevent exposure of the internal carotid to the base of the skull: the parotid gland and the facial nerve. Excessive anterior traction on the mandible or parotid may damage the facial nerve, particularly the mandibular branch. Unless the surgeon is willing to resect the parotid and divide the facial nerve, division of the ramus is seldom helpful.

Penetrating carotid artery injuries, regardless of the patient's neurologic status, usually require repair, except in comatose patients. Inaccessible carotid artery injuries near the base of the skull can be treated by interventional radiologists with a stent if the anatomy of the injury is favorable. Otherwise the artery will need to be thrombosed or ligated. If ligation is necessary, the patient should be given anticoagulation therapy with heparin followed by warfarin sodium (Coumadin) for 3 months. This treatment may prevent a stroke by inhibiting the generation of thrombi from the surface of the clot at the circle of Willis while the endothelium heals. Without anticoagulation therapy the risk of stroke with ligation has been approximately 20 to 30 percent, and most strokes occur a few days after ligation. Tangential wounds of the internal jugular vein should be repaired by lateral venorrhaphy, but extensive wounds are efficiently attended to by ligation.

Vertebral artery injuries usually result from penetrating trauma, although thrombosis and pseudoaneurysms can occur from blunt injury. The diagnosis is made by angiography or when significant hemorrhage is noted posterior to the carotid sheath during neck exploration. Exposure of the vertebral artery above the C6 vertebra where it enters its bony canal is complicated by the overlying anterior elements of the canal and the tough fascia covering the artery between the elements. The artery is approached through an anterior neck

incision by retracting the contents of the carotid sheath laterally (Fig. 6-34). The muscular attachments to the anterior elements are removed. Care must be taken to avoid injury to the cervical spinal nerves that are located directly behind and lateral to the bony canal. Some authorities have recommended using a high-speed burr to remove the anterior aspect of the canal, thereby avoiding the venous plexus between the elements. The authors have not found this to be a problem and have often excised the fascia between the elements and lifted the artery out of its canal with a tissue forceps.

The treatment for vertebral artery injuries is ligation proximal and distal to the injury. There is rarely, if ever, an indication for repair. Neurologic complications are uncommon. Exposure of the vertebral artery above C2 is extremely difficult. Rather than using a direct operative approach, the authors expose the vessel below C5, outside the bony canal, clamp the artery proximally, and insert a No. 3 balloon-tipped catheter. The catheter is advanced to the level of the injury or distal to it, and the balloon is inflated with saline solution until back bleeding stops. The tube to the catheter is crimped over on itself and secured in this position with several heavy silk sutures. The catheter is trimmed so that it can be left in the wound under the skin. The proximal end of the artery is ligated. One week later the catheter is removed under local anesthesia. Rebleeding has not occurred in our experience.

The same approach can be used for the distal internal carotid artery. An alternative approach is to have the interventional radiologist place coils to induce thrombosis proximal and distal to the injury if the lesion is diagnosed by angiography. Not all vertebral artery injuries can be treated by this method. Injuries of the proximal vertebral artery can be exposed by a median sternotomy with a neck extension.

Trachea and Esophagus

Injuries of the trachea are repaired with a running 3-0 absorbable monofilament suture. Tracheostomy is not required in most patients. Esophageal injuries are repaired in a similar fashion. If an esophageal wound is large, or if tissue is missing, a sternocleidomastoid muscle pedicle flap is warranted, and a closed-suction drain is a reasonable precaution. The drain should be near but not in contact with the esophageal or any other suture line. It can be removed in 7 to 10 days if the suture line remains secure. Care must be taken when exploring the trachea and esophagus to avoid iatrogenic injury to the recurrent laryngeal nerves.

Penetrating injuries of the neck often create wounds in adjacent hollow structures, e.g., the trachea and esophagus or the carotid artery and esophagus. If, after repair, these adjacent suture lines are in contact, the stage is set for devastating postoperative fistulous complications. To avoid these complications, viable tissue should routinely be interposed between adjacent suture lines. Viable strips of the sternocleidomastoid muscle or strap muscles are useful for this purpose.

Thoracic Outlet

Great Vessels

Most injuries of the great vessels of the thoracic outlet (Zone I) are caused by penetrating trauma, although the innominate and subclavian arteries are occasionally injured from blunt trauma. Angiography is desirable for planning the incision. If this is not possible because of hemodynamic instability, a reasonable approach can be inferred from the chest x-ray and the location of the wounds. If the patient has a left hemothorax, a left third or fourth interspace anterolateral thoracotomy should be performed because the proximal left subclavian artery may be injured. Hemorrhage can be controlled digitally until the vascular injury is delineated. Additional incisions or extensions are often required. A third or fourth interspace right anterolateral thoracotomy may be used for thoracic outlet injury presenting with hemodynamic instability and a right hemothorax. A median sternotomy with a right clavicular extension also can be used. Unstable patients with injuries near the sternal notch may have a large mediastinal hematoma or have lost blood directly to the outside. These patients should be explored via a median sternotomy.

If angiography has identified an arterial injury, a more direct approach can be used. Fig. 6-35 shows the various incisions that are used depending on the location of the arterial injury. A median sternotomy is used for exposure of the innominate, proximal right carotid and subclavian, and proximal left carotid arteries.

The proximal left subclavian artery presents a unique challenge. Because it arises from the aortic arch far posteriorly, it is not readily approached via a median sternotomy. A posterolateral thoracotomy provides excellent exposure but severely limits access to other structures and is not recommended. The best option is to create a full-thickness flap of the upper chest wall. This is accomplished with a third or fourth interspace anterolateral thoracotomy for proximal control, a supraclavicular incision with a resection of the medial third of the clavicle, and a median sternotomy, which links the two horizontal incisions. The ribs can be cut laterally for additional exposure, which allows the flap to be folded laterally with little effort. This incision has been referred to as a book or trapdoor thoracotomy (Fig. 6-36). The midportion of the subclavian artery is accessible by removing the proximal third of either clavicle, with the skin incision made directly over the clavicle. Muscular attachments are stripped away, and the clavicle is divided with a Gigli's wire saw. The medial remnant of the clavicle is forcibly elevated. The periosteum is dissected from the posterior aspect of the bone until the sternoclavicular joint is reached. The capsular attachments are cut with a heavy scissors or knife, and the bone is discarded. The periosteum and underlying fascia are very tough and must be sharply incised along the direction of the vessel. The subclavian vein is mobilized, and the artery is directly underneath. The anterior scalene muscle is divided for injuries just proximal to the thyrocervical trunk; the relatively small phrenic nerve should be identified on its anterior aspect and spared. Iatrogenic injury to cords of the brachial plexus can occur.

The great vessels are fragile and easily torn during dissection or crushed with a clamp. Some advocate oversewing proximal injuries of the artery on the side of the aortic arch and sewing a graft onto a new location on the arch. The graft is then sewn to the artery without tension. The authors have not found this necessary, provided the vessels are handled with care.

Trachea and Esophagus

The trachea and esophagus are difficult to approach at the thoracic outlet. The combination of a neck incision and a high anterolateral thoracotomy may be used. Alternatively, these structures can be approached via a median sternotomy, provided the left innominate vein and artery are divided. Temporary division of the innominate artery is tolerated well in otherwise healthy people, but the vessel should be repaired after treatment of the tracheal or esophageal injury. The vein does not need to be repaired. As in the neck, adjacent suture lines should be separated by viable tissue. A portion of the sternocleidomastoid can be rotated down for this purpose.

Chest

The most common life-threatening complications from blunt and penetrating thoracic injury are hemothorax, pneumothorax, or a combination of the two. Approximately 85 percent of these patients can be treated definitively with a chest tube. Because of the viscosity of blood at various stages of coagulation, a 36F or larger chest tube should be used. If one tube fails to completely evacuate the hemothorax (a "caked hemothorax"), a second tube should be placed (Fig. 6-37). If the second chest tube does not remove the blood, a thoracotomy should be performed because of the risk of life-threatening hemorrhage. Common sources of blood loss include intercostal vessels, internal thoracic artery, pulmonary parenchyma, and the heart. Less common sources are the great vessels, aortic arch, azygos vein, superior vena cava, and inferior vena cava. Blood may also enter the chest from an abdominal injury through a perforation or tear in the diaphragm. Indications for operative treatment of penetrating thoracic injuries are listed in Table 6-8.

The indications for thoracotomy in blunt trauma are based on specific preoperative diagnoses. These include pericardial tamponade, tear of the descending thoracic aorta, rupture of a main bronchus, and rupture of the esophagus. Thoracotomy for hemothorax in the absence of the above diagnoses is rarely indicated. A shattered chest wall that produces a hemothorax is better treated by the interventional radiologist with embolization.

Thoracic Incisions

The selection of incision is important and depends on the organs being treated. For exploratory thoracotomy for hemorrhage, the patient is supine and an anterolateral thoracotomy is performed. Depending on findings, the incision can be extended across the sternum or even farther for a bilateral anterolateral thoracotomy. The fifth interspace usually is preferred unless the surgeon has a precise knowledge of which organs are injured and knows that exposure would be enhanced by selecting a different interspace. The heart, lungs, aortic arch, great vessels, and esophagus are accessible with these incisions. Care should be taken to ligate the internal thoracic artery and veins if they are transected. This step often is overlooked, resulting in continuous blood loss that obscures the field and endangers the patient.

The heart also can be approached via a median sternotomy. Because little else can be done in the chest through this incision, it usually is reserved for stab wounds of the anterior chest in patients who present with pericardial tamponade. Posterolateral thoracotomies rarely are used since ventilation is impaired in the dependent lung, and the incision cannot be extended. There are two specific exceptions. Injuries of the posterior aspect of the trachea or main bronchi near the carina tracheae are inaccessible from the left or from the front. The only possible approach is through the right chest using a posterolateral thoracotomy. A tear of the descending thoracic aorta can be repaired only through a left posterolateral thoracotomy. Because the authors use left heart bypass for these procedures, the patient's hips and legs are rotated toward the supine position to gain access to the left groin for femoral artery cannulation. It is also helpful for optimal exposure to resect the fourth rib and enter the chest through its bed.

Heart

Most cardiac injuries are the result of penetrating trauma, and any part of the heart is susceptible. Control of hemorrhage while the heart is being repaired is crucial, and several techniques can be used. The atria can be clamped with a Satinsky vascular clamp. Digital control and suturing beneath the finger is possible anywhere in the heart, though the technique requires skill and a long, curved cardiovascular needle. However, the reality of blood-borne viral infections raises the question of whether this method should be used. If the hole is small, a "peanut" sponge clamped in the tip of a hemostat can be placed into the wound, or the blood loss may be accepted while sutures are being placed. For larger holes a 16F Foley catheter with a 30 mL balloon can be inflated with 10 mL of saline solution. Gentle traction on the catheter controls hemorrhage from any cardiac wound because wounds too large for balloon tamponade are incompatible with survival. Suture placement with the balloon inflated is problematic. Usually the ends of the wound are closed progressively toward the middle until the amount of blood loss is acceptable with the balloon removed. The use of skin staples for the temporary control of hemorrhage has become popular, particularly when emergency room thoracotomy has been performed. The use of staples has the advantages of reducing the risk of needle-stick injury to the surgeon and of not demanding the attention required by a balloon catheter. In most instances, hemostasis is neither perfect nor definitive. Inflow occlusion of the heart by clamping the superior and inferior venae cavae can be performed for short periods, and this may be essential for the treatment of extensive or multiple wounds as well as for those that are difficult to expose.

Immediate repair of valvular damage or acute septal defects rarely is necessary and requires total cardiopulmonary bypass, which has a high mortality in this situation. Most patients who survive to make it to the hospital do well with only external repair. After recovery, the heart can be thoroughly evaluated, and, if necessary, secondary repair can be performed under more controlled conditions. Coronary artery injuries also pose difficult problems. Ligation leads to acute infarction distal to the tie, but reconstruction requires bypass. The right coronary artery can be ligated anywhere, but the resultant arrhythmias may be extremely resistant to treatment. The left anterior descending and circumflex arteries cannot be ligated proximally without causing a large infarct. These injuries are rare and usually produce death in the field.

Lungs

Pulmonary injuries requiring operative intervention usually result from penetrating injury. Formerly the entrance and exit wounds were oversewn to control hemorrhage. This set the

stage for air embolism, which occasionally caused sudden death in the operating room or in the immediate postoperative period. Pulmonary tractotomy has been used to reduce this problem as well as the need for pulmonary resection. Linear stapling devices are inserted directly into the injury tract and positioned to cause the least degree of devascularization (Fig. 6-38). Two staple lines are created and the lung is divided between. This allows direct access to the bleeding vessels and leaking bronchi. No effort is made to close the defect. Lobectomy or pneumonectomy rarely is necessary. Lobectomy is indicated only for a completely devascularized or destroyed lobe. Parenchymal injuries severe enough to require pneumonectomy rarely are survivable, and major pulmonary hilar injuries necessitating pneumonectomy usually are lethal in the field.

Trachea and Esophagus

Injuries of the trachea and esophagus are managed in the same fashion as described above for lung injuries. Because exposure can be difficult, provisions should be made to deflate the lung on the operative side by using a double-lumen endotracheal tube (a double-lumen tube is seldom needed for cardiac or pulmonary injury). Repair of injuries of the main bronchi and the trachea near the carina tracheae can result in a complete loss of ventilation when the overlying pleura is opened, even if a double-lumen tube is used. Gases from the ventilator preferentially escape from the injury and neither lung will be ventilated. Digital occlusion of the injury can control air loss if the injury is small. Larger injuries are an imminent threat to life. To avoid this catastrophe, a 6 or 7 mm cuffed endotracheal tube should be on the operative field and a second ventilator available. If ventilation is inadequate, the surgeon can insert and inflate the endotracheal tube into the main bronchus on the opposite side through the injury to permit ventilation of one lung while the injury is repaired. Eventually, the tube will have to be removed to close the defect, but the remaining hole can be controlled digitally. Alternatively, it may be possible for the anesthesiologist to cannulate the opposite bronchus.

Descending Thoracic Aorta

The occurrence of paraplegia from ischemic injury of the spinal cord has been a concern in injuries to the descending thoracic aorta. Conceptually, two techniques have been advocated. The simpler technique, often referred to as “clamp and sew,” is accomplished with the application of vascular clamps proximal and distal to the injury and repair or replacement of the damaged portion of the aorta. This method results in transient hypoperfusion of the spinal cord distal to the clamps as well as all abdominal organs. Large doses of vasodilators also are required to reduce afterload and avoid acute left heart failure. If the clamping time is short, less than 30 min, paraplegia is uncommon. Longer clamping times have been associated with paraplegia in approximately 10 percent of patients. Clamping times of less than 30 min are difficult to achieve where there are many tears requiring complex repair. An alternative approach is to provide some method for maintaining a reasonable degree of perfusion for organs distal to the clamps. Two techniques have been used to accomplish this goal. The first is with the use of a shunt, a temporary extraanatomic route around the clamps. A heparin-impregnated tube, the Gott shunt, has been designed specifically for this purpose, but the volume of blood flow to the distal aorta is marginal. The second method is to use left heart bypass. With this method a volume of oxygenated blood is siphoned from the left heart and pumped into the distal aorta. Flow rates of 2 to 3 L/min appear to provide adequate protection by maintaining a distal perfusion pressure higher than 65 mmHg. This is the preferred method. The left superior pulmonary vein, rather than the left atrium, is cannulated to remove blood from the heart because the vein is tougher and less prone to tearing (Fig. 6-39). The left femoral artery is cannulated to return the blood to the distal aorta. A centrifugal pump is used because it is not as thrombogenic as a roller pump and, strictly speaking, heparinization is not required. This can be a significant benefit in patients with multiple injuries, particularly in those with intracranial hemorrhage. Occasional small cerebral infarcts have occurred, and 5,000 to 10,000 units of heparin usually is administered unless contraindicated by associated injuries.

Once bypass is initiated, the proximal vascular clamp is applied between the left common carotid and left subclavian arteries, and the distal clamp is placed distal to the injury. The left subclavian artery is clamped separately. The hematoma is entered and the injury evaluated. In most patients a short gelatin-sealed Dacron graft is placed, usually 18 to 22 mm in diameter. Primary repair without a graft is possible in some patients. For the anastomoses or suture lines, 3-0 polypropylene suture is used. Air and clot are flushed from the aorta between two clamps and the subclavian artery before tying the final suture. After completion of the repair the clamps are removed and the patient is weaned from the pump. The cannulae are removed, and the vessels are repaired. Meta-analysis comparing the clamp-and-sew and left heart bypass methods revealed a significantly lower incidence of paraplegia when the pump is used.

Injuries of the transverse aortic arch do occur from blunt trauma. The proximal clamp usually can be placed between the innominate and left carotid arteries without cerebral infarction. The proximal clamp, however, cannot be placed proximal to the innominate artery. A possible approach to injuries in which the clamps completely exclude the cerebral circulation is to use profound hypothermia and circulatory arrest.

Small intimal flaps of the thoracic aorta without hematomas can be treated nonoperatively. Intraluminal mediastinal stents also may provide a solution, but their role remains to be defined. Penetrating injuries of the thoracic aorta are rare and do not afford enough time to set up the pump. There is no choice but to use the clamp-and-sew technique with partially occluding clamps, if possible.

Abdomen

All abdominal explorations in adults are performed using a long midline incision because of its versatility. For children under the age of 6 years, a transverse incision may be advantageous. If the patient has been in shock or is currently unstable, no attempt should be made to control bleeding from the abdominal wall until major sources of hemorrhage have been identified and controlled. The incision should be made with a scalpel rather than with an electrosurgical unit, because it is faster. Liquid and clotted blood are rapidly evacuated with multiple laparotomy pads and suction. Additional pads are placed in each quadrant to localize hemorrhage, and the aorta is palpated to estimate blood pressure.

If exsanguinating hemorrhage is encountered on opening the abdomen, it usually is caused by injury to the liver, aorta, inferior vena cava, or iliac vessels. If the liver is the source, the hepatic pedicle should be immediately clamped (a Pringle maneuver) and the liver compressed posteriorly by tightly packing several laparotomy pads between the hepatic injury and the underside of the right anterior chest wall (Figs. 6-40, 6-41). This combination of maneuvers temporarily controls the hemorrhage from most survivable hepatic injuries.

If exsanguinating hemorrhage originates near the midline in the retroperitoneum, direct manual pressure is applied with a laparotomy pad, and the aorta is exposed at the diaphragmatic hiatus and clamped. The same approach is used in the pelvis except that the infrarenal aorta can be clamped, which is easier and safer because splanchnic and renal ischemia are avoided. Injuries of the iliac vessels pose a particular problem for emergency vascular control. Because there are so many large vessels in proximity, multiple vascular injuries are common. Venous injuries are not controlled with aortic clamping. A helpful maneuver in these instances is pelvic vascular isolation (Fig. 6-42). For stable patients with large midline hematomas, clamping the aorta proximal to the hematoma also is a wise precaution. Many surgeons take a few moments, once overt hemorrhage has been controlled, to identify obvious sources of enteric contamination and minimize further spillage. This can be accomplished with a running suture or with Babcock clamps.

Any organ can be injured by blunt or penetrating trauma, but certain organs are injured more often, depending on the mechanism. In blunt trauma, organs that cannot yield to impact by elastic deformation are most likely to be injured. The solid organs—liver, spleen, and kidneys—are representative of this group. For penetrating trauma, organs with the largest anterior surface area are most prone to injury, i.e., the small bowel, liver, and colon. Bullets and knives usually follow straight lines, and adjacent structures are commonly injured, e.g., the pancreas and duodenum. Penetrating trauma is not limited by the elastic properties of the tissue, and vascular injuries are far more common. While these general principles simplify the localization of injuries, unless the patient has exsanguinating hemorrhage, a methodical exploration always should be performed.

Abdominal organs are systematically examined by visualization or palpation. Missed injuries are a serious problem with often fatal results. In penetrating trauma missed injuries can occur if wound tracks are not followed their entire course. Injuries also can be missed if the surgeon fails to explore retroperitoneal structures such as the ascending and descending colon, the second and third portion of the duodenum, and the ureters. Injuries of the aorta or venae cavae may be temporarily tamponaded by overlying structures. If the retroperitoneum is opened and the injury overlooked, delayed massive hemorrhage can occur after abdominal closure. Blunt abdominal injuries usually are obvious, but injuries of the pancreas, duodenum, bladder, and even the aorta can be overlooked.

Vascular Injuries

Injury to the major arteries and veins in the abdomen are a technical challenge to the surgeon and often are fatal. All vessels are susceptible to injury in penetrating trauma. Vascular injuries in blunt trauma are far less common and usually involve the renal arteries and veins, though all other vessels, including the aorta, can be injured. Several vessels are difficult to expose: the retrohepatic vena cava, the suprarenal aorta, the celiac axis, the proximal superior mesenteric artery, the junction of the superior mesenteric, splenic, and portal veins, and the bifurcation of the vena cava. Techniques to aid in the exposure of these vessels have been described. The suprarenal aorta, the celiac axis, and the proximal superior mesenteric and left renal arteries can be exposed by left medial visceral rotation (Fig. 6-43). This is accomplished by incising the left lateral peritoneal reflection beginning at the distal descending colon and extending the incision past the splenic flexure, around the posterior aspect of the spleen, behind the gastric fundus, and ending at the esophagus. This incision permits the left colon, spleen, pancreas, and stomach to be rotated toward the midline. Division of the left crus of the diaphragm permits access to the aorta well above the celiac axis. In contrast, mobilization of the right colon and a Kocher maneuver exposes the entire inferior vena cava except the retrohepatic portion, and it is technically simple. This is referred to as a right medial visceral rotation (Fig. 6-44). The kidney can be left in situ or mobilized with the remaining viscera with right and left medial rotations.

The junction of the superior mesenteric, splenic, and portal veins can be exposed in elective surgery by dissecting the vessels from the pancreas, as required when performing a distal splenorenal shunt. In the presence of massive bleeding from a venous injury, this may be impossible. Therefore, the neck of the pancreas is divided without hesitation. This provides excellent exposure of this difficult area.

The bifurcation of the vena cava is obscured by the right common iliac artery. This vessel should be divided to expose extensive vena caval injuries of this area (Fig. 6-45). The artery must be repaired after the venous injury is treated. Amputation occurs in approximately 50 percent of patients in whom the vessels are not repaired.

Liver

The lower costal margins impair visualization and a direct approach to the liver. Exposure of the right lobe can be improved by elevating the right costal margin with a large Richardson retractor. The right lobe can be mobilized by dividing the right triangular and coronary ligaments. After division of the right triangular ligament, the dissection is continued medially, dividing the superior and inferior coronary ligaments. The right lobe then can be rotated medially into the surgical field. Mobilization of the left lobe is accomplished in the same fashion. Care must be taken when dividing any of the coronary ligaments because of their proximity to the hepatic veins and the retrohepatic vena cava. On occasion it may be necessary to extend the midline abdominal incision into the chest. This is best accomplished with a median sternotomy. The pericardium and diaphragm can be divided toward the center of the inferior vena cava. The combination of incisions provides outstanding exposure of the hepatic veins and retrohepatic vena cava while avoiding injury to the phrenic nerves.

The Pringle maneuver is one of the most useful techniques for evaluating the extent of hepatic injuries (see Fig. 6-40). In patients with extensive hepatic injuries, the Pringle maneuver differentiates between hemorrhage from the hepatic artery and portal vein, which ceases when the clamp is applied, and hemorrhage from the hepatic veins and retrohepatic vena cava, which does not. The authors prefer to manually tear the lesser omentum and place the clamp from the left side while guiding the posterior blade of the clamp through the foramen of Winslow with the aid of the left index finger. This approach has the advantage of avoiding injury to the structures within the hepatic pedicle, assuring

that the clamp will be placed properly the first time, and including any anomalous or accessory left hepatic arteries between the blades of the clamp.

Techniques for the temporary control of hemorrhage from the liver are necessary when dealing with an extensive injury to provide the anesthesiologist with sufficient time to restore circulating blood volume before proceeding, and because it is not possible to control hemorrhage from more than one location in the abdomen simultaneously. The temporary hemostatic techniques that have proved most useful are hepatic compression, the Pringle maneuver, and perihepatic packing. Manual compression of a bleeding hepatic injury may be a lifesaving maneuver (Fig. 6-46). The addition of laparotomy pads on the surface of the liver distributes digital forces and lessens the chance of aggravating the injury. If the lacerated edges of the liver are carefully opposed and the proper forces applied, hemorrhage from almost any hepatic injury can be controlled. The obvious drawback is that considerable skill is required and that little else can be done while the liver is being compressed. Manual compression is best suited for immediate attempts to prevent exsanguination and for periodic control during a complex procedure.

Perihepatic packing also is capable of controlling hemorrhage from most hepatic injuries, and it has the advantage of freeing the surgeon's hands. The laparotomy pads, two or three stacked together, should remain folded. The right costal margin is elevated, and the pads are strategically placed over and around the bleeding site (see Fig. 6-41). Additional pads should be placed between the liver, diaphragm, and anterior chest wall until the bleeding has been controlled. Ten to 15 pads may be required to control the hemorrhage from an extensive right lobar injury. The effectiveness of packing may be enhanced by downward pressure on the right costal margin by an assistant. Packing of injuries of the left lobe is not as effective because there is insufficient abdominal and thoracic wall anterior to the left lobe to provide adequate compression with the abdomen open. Hemorrhage from the left lobe usually can be controlled by mobilizing the lobe and compressing it between the surgeon's hands.

Two complications might be caused by packing hepatic injuries. Tight packing can compress the inferior vena cava and reduce cardiac filling, and the right diaphragm will be forced cephalad, increasing airway pressure and decreasing tidal volume and functional residual capacity. The surgeon must decide whether these complications outweigh the risk of additional blood loss.

Perihepatic packing will not reliably control hemorrhage from larger branches of the hepatic artery. The Pringle maneuver often is used as an adjunct to packing for the temporary control of the arterial hemorrhage. Properly applied, a Pringle maneuver eliminates all hepatopetal flow. The length of time that a Pringle maneuver can remain in place without causing irreversible ischemic damage to the liver is unknown. Several authors have documented a Pringle maneuver applied for over 1 h without appreciable hepatic damage; this is a reasonable figure. Another option for temporary control of hepatic hemorrhage is to use a tourniquet. After mobilization of the bleeding lobe, a 1-inch Penrose drain is wrapped around the liver near the anatomic division between the left and right lobes. The drain is cinched until hemorrhage ceases; tension is maintained by placing a clamp on the drain. Tourniquets are difficult to use, however, because they often slip off or even tear through the parenchyma. An alternative is to use the Lin liver clamp, though it has the same shortcomings as the tourniquet. If successful, the occluding device is removed in 24 h and nonviable tissue is resected.

Special techniques have been developed for controlling hemorrhage from juxtahepatic venous injuries. These formidable procedures include hepatic vascular isolation with clamps, the atriocaval shunt, and the Moore-Pilcher balloon. Hepatic vascular isolation with clamps is accomplished by the application of a Pringle maneuver, clamping the aorta at the diaphragm, and clamping of the suprarenal and suprahepatic vena cava. Although this technique has success in elective procedures, its use in trauma patients has had mixed results because patients in profound hemorrhagic shock do not tolerate the precipitous loss of venous return to the heart.

The atriocaval shunt was designed to achieve hepatic vascular isolation while permitting venous blood to enter the heart from below the diaphragm. Enthusiasm for the shunt has declined as mortality rates with its use range from 50 to 80 percent. The shunt must be precisely constructed and properly positioned on the first attempt because patients with juxtahepatic venous injuries do not tolerate the continuing blood loss associated with repeated unsuccessful attempts to position the shunt correctly. A variation of the original atriocaval shunt has been the substitution of a 9-mm endotracheal tube for the usual large chest tube (Fig. 6-47). While this change may seem trivial, surrounding the suprarenal vena cava for a snare tourniquet is extremely difficult because exsanguinating hemorrhage must be controlled by posterior compression of the liver, which severely restricts access to that segment of the vena cava.

An alternative to the atriocaval shunt is the Moore-Pilcher balloon. This device is inserted through the femoral vein and advanced into the retrohepatic vena cava. When the balloon is inflated, the hepatic veins and vena cava are occluded, thereby achieving vascular isolation. The catheter itself is hollow, and holes placed below the balloon permit blood to flow into the right atrium from the inferior vena cava.

Surgeons attempting hepatic vascular isolation should be aware that none of the techniques provides complete hemostasis. The residual bleeding after successful vascular isolation can be removed readily with suction. Regardless of the technique used, a Pringle maneuver should always be used. Because of the technical challenge and high mortality of hepatic vascular isolation, there has been a trend toward avoiding a direct operative approach to the injured vessels. If massive venous hemorrhage is seen from behind the liver, and, if hemostasis can be achieved with perihepatic packing, the patient can be transferred to the interventional radiology suite, where hemorrhage from arterial sources are embolized and stents are placed to bridge venous injuries (see Fig. 6-26).

Numerous methods for the definitive control of hepatic hemorrhage have been developed. Minor lacerations may be controlled with manual compression applied directly to the injury site. For similar injuries that do not respond to compression, topical hemostatic techniques have been successful. Small bleeding vessels may be controlled with electrocautery, although the power output of the machine may have to be increased. Bleeding surfaces immune to electrocautery may respond to the argon beam coagulator. Microcrystalline collagen can be used. The powder is placed on a clean 4 × 4 sponge and applied directly to the oozing surface. Pressure is maintained for 5 to 10 min. Topical thrombin also can be applied to minor bleeding injuries by saturating a gelatin foam sponge or a microcrystalline collagen pad and applying it to the bleeding site.

Fibrin glue has been used for superficial and deep lacerations and is an effective topical agent. Fibrin glue is made by mixing concentrated human fibrinogen (cryoprecipitate) with bovine thrombin and calcium. Because the coagulum forms quickly, the fibrinogen and thrombin-calcium solution are placed in separate syringes joined with a Y connector. Spray-on applicators also have been used. Enthusiasm has been tempered by reports of fatal anaphylactic reactions and idiopathic hypotension related to an antigenic response to the bovine component.

Suturing of the hepatic parenchyma is an effective hemostatic technique. This treatment has been maligned as a cause of hepatic necrosis, but hepatic sutures often are used for persistently bleeding lacerations less than 3 cm in depth. It also is an appropriate alternative for deeper lacerations if the patient will not tolerate further hemorrhage. The preferred suture is 2-0 or 0 chromic attached to a large, curved, blunt needle. The large diameter of the suture helps to prevent it from pulling through Glisson's capsule. A simple running technique is used to approximate the edges of shallow lacerations. Deeper lacerations can be managed with interrupted horizontal mattress sutures placed parallel to the edge of the laceration. When tying the suture, adequate tension exists when visible hemorrhage ceases or the liver blanches around the suture.

Most sources of venous hemorrhage within the liver can be managed with parenchymal sutures, and even injuries of the retrohepatic vena cava and hepatic veins have been successfully tamponaded by closing the hepatic parenchyma over the bleeding vessel. Venous hemorrhage resulting from penetrating wounds that traverse the central portion of the liver can be managed by suturing the entrance and exit wounds with horizontal mattress sutures. Intrahepatic hematomas might form, which can become infected, though this may be preferable to intracaval shunt or deep hepatotomy. Suturing of the hepatic parenchyma is not always successful in controlling the hemorrhage, particularly when the hemorrhage is of arterial origin.

Hepatotomy with selective ligation of bleeding vessels is an important technique usually reserved for transhepatic penetrating wounds. Hepatotomy is performed using the finger fracture technique. The dissection continues until the bleeding vessels are identified and controlled. Considerable blood loss may be incurred because the division of additional viable hepatic tissue is often required to reach the bleeding vessels. An alternative to suturing the entrance and exit wounds of a transhepatic injury or extensive hepatotomy is the use of an intrahepatic balloon. Our method is to tie a large Penrose drain to a hollow catheter and ligate the opposite end of the drain (Fig. 6-48). The balloon is then inserted into the bleeding wound and inflated with soluble contrast medium. If control of the hemorrhage is successful, a stopcock or clamp is used to occlude the catheter and maintain inflation. The catheter is left in the abdomen and removed at an operation 24 to 48 h later. Recurrent hemorrhage may occur when the balloon is deflated but usually is amenable to selective embolization.

Hepatic arterial ligation may be appropriate for patients with recalcitrant arterial hemorrhage from deep within the liver. Its utility is limited because hemorrhage from the portal and hepatic venous systems continues. Its primary role is in transhepatic injuries when application of the Pringle maneuver results in the cessation of arterial hemorrhage. Arterial ligation is a reasonable alternative to a deep hepatotomy, particularly in unstable patients. While ligation of the right or left hepatic artery is well tolerated, the fate of dearterialized lobe is unpredictable. Lobar necrosis requiring anatomic lobectomy after arterial ligation has been described. Ligation of the proper hepatic artery may not be tolerated.

An uncommon, perplexing hepatic injury is the subcapsular hematoma. This lesion occurs when the parenchyma of the liver is disrupted by blunt trauma, but Glisson's capsule remains intact. The hematoma may be recognized at the time of the surgery or preoperatively if CT is performed, and subsequent decision making is often difficult. Subcapsular hematomas discovered during an exploratory laparotomy that involve less than 50 percent of the surface of the liver and are not expanding or ruptured should be left alone or packed. Hematomas that are expanding during an operation may require exploration. These lesions often are caused by uncontrolled arterial hemorrhage, and packing alone may not be successful. An alternative strategy is to pack the liver to control venous hemorrhage, close the abdomen, and transport the patient to the angiographic suite for hepatic arteriography and embolization of the bleeding vessel. Ruptured hematomas require exploration and selective ligation, with or without packing.

Resectional debridement is indicated for the removal of peripheral portions of nonviable hepatic parenchyma. The mass of tissue removed should rarely exceed 25 percent of the liver. Because additional blood loss may occur, it should be reserved for patients who are in good metabolic condition and who will tolerate additional blood loss. Resectional debridement is performed by finger fracture. An alternative for patients with extensive unilobar injuries is anatomic hepatic resection, but the mortality rate for trauma patients exceeds 50 percent in most series. It has largely been replaced by perihepatic packing, resectional debridement, and hepatotomy with selective ligation and usually is not indicated in the acute setting. There are two circumstances, however, in which anatomic resections are appropriate. The first is when there are extensive injuries of the lateral segment of the left lobe. Because hemorrhage can be easily controlled with bimanual compression, uncontrolled blood loss is not as problematic as with the left or right anatomic lobectomies. Another indication for anatomic lobectomy occurs in patients whose hemorrhage has been controlled by perihepatic packing or arterial ligation but whose left or right lobe is nonviable. The mass of the remaining necrotic liver is large and the risk of subsequent infection high, and it should be removed as soon as the patient's condition permits.

Several centers have reported patients with devastating hepatic injuries or necrosis of the entire liver who have undergone successful hepatic transplantation. The patient must have all other injuries delineated, particularly of the central nervous system, and have an excellent chance of survival excluding the hepatic injury. Cost and donor availability limit such procedures, but it is probable that hepatic transplantation for trauma will continue to be performed in rare circumstances.

Omentum has been used to fill large defects in the liver, with the rationale that it provides an excellent source for macrophages and that it fills a potential dead space with viable tissue. The omentum also can provide some additional support for parenchymal sutures and often is strong enough to prevent them from cutting through Glisson's capsule.

Several prospective and retrospective studies have demonstrated that the use of Penrose or sump drains is associated with a greater risk of intraabdominal sepsis than the use of closed-suction drains or no drains. Drains are not necessary for minor lacerations. They should be used if bile is seen oozing from the liver and in most patients with deep central injuries.

The complications after significant hepatic trauma include hemorrhage, infection, and various fistulas. Postoperative hemorrhage can be expected in a considerable percentage of patients treated with perihepatic packing. The source may be persistent coagulopathy or missed vascular injury. In most instances in which postoperative hemorrhage is suspected, the patient is best served by returning to the operating room. Arteriography with embolization can be considered in selected patients.

Infections within and around the liver occur in about 3 percent of injured patients. Perihepatic infections develop more often in victims of penetrating trauma than blunt trauma, presumably because of the greater frequency of enteric contamination of the former. Persistent elevation of temperature and white blood cell count after the third or fourth postoperative day should prompt a search for intraabdominal infection. In the absence of pneumonia, line sepsis, or urinary tract infection, an abdominal CT scan with intravenous and upper gastrointestinal contrast should be obtained. Many perihepatic infections can be treated with CT-guided drainage. Infected hematomas and infected necrotic liver cannot be expected to respond to percutaneous drainage. Right twelfth rib resection remains an excellent approach for posterior infections and allows superior drainage.

Bilomas are loculated collections of bile that may or may not be infected. If infected, the biloma is essentially an abscess and should be treated as such. If sterile, it will eventually be reabsorbed. Biliary ascites is caused by disruption of a major bile duct. Reoperation with the establishment of appropriate drainage is the prudent course. Even if the source of bile leakage can be identified, primary repair of the injured duct is unlikely to be successful. It is best to wait until a firm fistulous communication is established with adequate drainage.

Biliary fistulas occur in approximately 3 percent of patients with hepatic injuries. They usually are of little consequence and most will close without specific treatment. Rarely, a fistulous communication with intrathoracic structures forms in patients with associated diaphragm injuries and results in a bronchobiliary or pleurobiliary fistula. As a result of the pressure differential between the biliary tract and thoracic cavity, most of these fistulas require operative closure, but the authors have treated a pleurobiliary fistula by endoscopic sphincterotomy with stent placement, which then closed spontaneously.

Hemorrhage from hepatic injuries often is treated without identifying and controlling each individual bleeding vessel; arterial pseudoaneurysms may develop. If the pseudoaneurysm enlarges, it eventually ruptures into the parenchyma of the liver, a bile duct, or an adjacent portal venous branch. Rupture into a bile duct results in hemobilia, which is characterized by intermittent episodes of right upper quadrant pain, upper gastrointestinal hemorrhage, and jaundice. If the aneurysm ruptures into a portal vein, portal venous hypertension with bleeding esophageal varices can occur. These complications are exceedingly rare and are best managed with hepatic arteriography and embolization. Biliovenous fistulas also have been reported. Serum bilirubin rises very rapidly, and extremely high values are common. Sphincterotomy of the papilla of Vater may hasten closure.

Gallbladder and Extrahepatic Bile Ducts

Injuries of the gallbladder are treated by lateral suture or cholecystectomy, whichever is easier. If lateral suture is performed, absorbable suture should be used to prevent the formation of calculi. Injuries of the extrahepatic bile ducts are a challenge. Because of the proximity of the portal vein, hepatic artery, and vena cava, associated vascular injuries are common and the patient's physiologic status is often poor. The ducts are of normal size and texture, i.e., small in diameter and thin-walled. These factors usually preclude primary repairs except for the smallest lacerations with no loss of tissue. These injuries can be treated by the insertion of a T tube through the wound or by lateral suture using 4-0 to 6-0 monofilament absorbable suture. Most transections and any injury associated with significant tissue loss require a Roux-en-Y choledochojejunostomy. The anastomosis is performed using a single-layer interrupted technique, as it is almost impossible to do a running stitch, using 4-0 or 5-0 monofilament absorbable suture. A round patch of seromuscular tissue the size of the common duct is removed from the jejunum at the site of the anastomosis to inhibit wound contraction. The mucosa and submucosa are punctured but not resected. Full-thickness bites of the duct and jejunum are taken. Because of the small size of the duct, only six to eight stitches can be used. T tubes are not placed. The jejunum is then sutured to the areolar tissue of the hepatic pedicle or porta hepatis to relieve tension from the anastomosis.

Injuries of the hepatic ducts are almost impossible to repair satisfactorily under emergency circumstances. One approach is to intubate the duct for external drainage and attempt a repair when the patient recovers. Alternatively, the duct can be ligated if the opposite lobe is normal and uninjured. For patients who are critically ill, the common duct also can be treated by intubation with external drainage.

Spleen

Splenic injuries are treated by splenic repair (splenorrhaphy), partial splenectomy, resection, or nonoperatively, depending on the extent of the injury and the condition of the patient. Enthusiasm for splenic salvage has been driven by the evolving trend toward nonoperative management of solid organ injuries and the rare, but often fatal, complication of overwhelming postsplenectomy infection (OPSI). These infections are caused by encapsulated bacteria, e.g., *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*, and are very resistant to treatment. OPSI occurs most often in young children and immunocompromised adults. It is uncommon in otherwise healthy adults.

To repair or remove the spleen safely, it should be mobilized to the extent that it can be brought to the surface of the abdominal wall without tension. This requires division of the attachments between the spleen and splenic flexure of the colon. An incision is made in the peritoneum and endoabdominal fascia beginning at the inferior pole, 1 or 2 cm from the spleen, and continuing posteriorly and superiorly until the esophagus is encountered; this is similar to a left medial visceral rotation (Fig. 6-49). Care must be taken not to pull on the posterior aspect of the spleen or it will tear at the peritoneal reflection, causing significant hemorrhage. The spleen should be rotated counterclockwise with posterior pressure applied to expose the peritoneal reflection. It is often helpful to rotate the operating table 20 degrees to the patient's right so that the weight of abdominal viscera aids in their own retraction. A plane can then be established between the spleen and pancreas and the renal fascia (Gerota's fascia) that can be extended to the aorta. This completes mobilization and permits repair or removal of the spleen without struggle for exposure.

Hilar injuries or a pulverized splenic parenchyma usually are treated by splenectomy. The authors have selectively reimplanted six pieces of the spleen (40 × 40 × 3 mm) within the leafs of the omentum. Technetium scans have confirmed their viability. IgM levels have normalized. The patient's response to an antigenic challenge has not been evaluated. Splenectomy also is indicated for lesser splenic injuries in patients with multiple abdominal injuries who have developed a coagulopathy; it usually is necessary in patients with failed splenic salvage attempts. Partial splenectomy can be used in patients in whom only a portion of the spleen has been destroyed, usually the superior or inferior half. After removal of the damaged portion, the same methods used to control hemorrhage from hepatic parenchyma can be used for the spleen. When placing horizontal mattress sutures across a raw edge, gentle compression of the parenchyma by an assistant facilitates hemostasis (Fig. 6-50). After ligation of the sutures and releasing compression, the spleen expands slightly and further tightens the sutures. Drains are never used after completion of the repair or resection. If splenectomy is performed, vaccines against the encapsulated bacteria are administered. The pneumococcal vaccine is routinely given, and those effective against *Haemophilus influenzae* and *Neisseria meningitidis* should be used.

Diaphragm

In blunt trauma the diaphragm is injured on the left in 75 percent of cases, presumably because the liver diffuses some of the energy on the right side. For blunt and penetrating trauma the diagnosis is suggested by an abnormality of the diaphragmatic shadow on chest x-ray. Many of these are subtle, particularly with penetrating injuries, and additional diagnostic evaluation may be warranted. The typical diaphragmatic injury from blunt trauma is a tear in the central tendon that may be large. Acute injuries are repaired through an abdominal incision. The laceration is closed with No. 1 monofilament permanent suture using a simple running technique. Occasionally, large avulsions or shotgun wounds with extensive tissue loss require polypropylene mesh to bridge the defect.

Duodenum

Duodenal hematomas are caused by a direct blow to the abdomen, and they occur more often in children than adults. Blood accumulates between the seromuscular and submucosal layers, eventually causing obstruction. The diagnosis is suspected by the onset of vomiting after blunt abdominal trauma; barium x-ray examination of the duodenum reveals the coiled- spring sign of obstruction. Most duodenal hematomas in children can be managed nonoperatively with nasogastric suction and parenteral nutrition. Resolution of the obstruction occurs in the majority of patients if this therapy is continued for 7 to 14 days. If surgical intervention is necessary, evacuation of the hematoma is associated with equal success and fewer complications than bypass procedures. Despite a lack of existing data on adults, there is no reason to believe that their hematomas should be treated differently from those of children. A new approach is laparoscopic evacuation if the obstruction persists more than 7 days.

Duodenal perforations can be caused by blunt and penetrating trauma (Fig. 6-51). Blunt injuries are difficult to diagnose because the contents of the duodenum have a neutral pH and few bacteria and are often contained by the retroperitoneum. Mortality can exceed 30 percent if the lesion is not identified and treated within 24 h. The perforations are not reliably identified by initial oral contrast CT examinations; therefore, the authors often obtain contrast x-rays with soluble contrast medium followed with barium if necessary. Most perforations of the duodenum can be treated by primary repair. The authors prefer to use a running, single-layer suture of 3-0 monofilament. The wound should be closed in a direction that results in the largest residual lumen. Occasionally, penetrating injuries damage only the pancreatic aspect of the second or third portion. Because the duodenum cannot be adequately mobilized to repair the injury directly, the wound should be extended laterally or the duodenum divided so that the pancreatic aspect can be sutured from the inside. Duodenal repairs or anastomoses do not benefit from adjunctive external drainage.

Challenges arise when there is a substantial loss of duodenal tissue. Extensive injuries of the first portion of the duodenum can be repaired by debridement and anastomosis because of the mobility and rich blood supply of the distal gastric antrum and pylorus. In contrast, the second portion is tethered to the head of the pancreas by its blood supply and the pancreatic and accessory pancreatic ducts (ducts of Wirsung and Santorini) so that the length of duodenum that can be mobilized from the pancreas is limited to approximately 1 cm. Unlike the jejunum, ileum, or colon, this mobilization yields little additional tissue to alleviate tension on suture line. As a result, suture repair of the second portion when tissue is lost often results in an unacceptably narrow lumen, and an end-to-end anastomosis is almost impossible, requiring more sophisticated repairs. For extensive injuries proximal to the accessory papilla, debridement and end-to-end anastomosis are appropriate. For lesions between the accessory papilla and the papilla of Vater, a vascularized jejunal graft, either a patch or a tubular interposition graft, may be required. Experience with these procedures is limited. Duodenal injuries with tissue loss distal to the papilla of Vater and proximal to the superior mesenteric vessels are best treated by Roux-en-Y duodenojejunostomy (Fig. 6-52). The distal portion of the duodenum is oversewn; the jejunum is sutured end-to-end to the proximal duodenum, and the defunctionalized distal duodenum and proximal jejunum are drained into the jejunum. Alternatively, the short defunctionalized duodenum can be resected. This is a rather tedious dissection behind the superior mesenteric vessels that may not be tolerated by a patient who has been in protracted shock.

Injuries to the third and fourth portions of the duodenum with tissue loss pose other problems. Because of the short mesentery of the third and fourth portions of the duodenum, the risk of ischemia limits mobilization. While end-to-end duodenojejunal anastomoses are possible in these regions, the technique used must resemble that of a hand-sewn, low anterior rectal anastomosis with a posterior row of interrupted sutures placed while the ends of the bowel are far apart. The jejunum is then parachuted down to the duodenum, and the anterior row is completed. Duodenal fistulas are common when this method is used. Resection of the third and fourth portions and a duodenojejunostomy on the right side of the superior mesenteric vessels is recommended.

An important adjunct for high-risk or complex duodenal repairs is the pyloric exclusion technique (Fig. 6-53). By occluding the pylorus and performing a gastrojejunostomy, the gastrointestinal stream can be diverted away from the duodenal repair. If a fistula does develop, it is functionally an end fistula, which is easier to manage and more likely to close than a lateral fistula, and the patient can take food by mouth to maintain nutritional status. To perform a pyloric exclusion, a gastrostomy is made on the greater curvature as close to the pylorus as possible. The pylorus is then grasped with a Babcock clamp via the gastrostomy and oversewn with a 0 polypropylene suture. A gastrojejunostomy restores gastrointestinal continuity. Vagotomy is not necessary because marginal ulceration occurs at the same frequency (approximately 3 percent) as duodenal ulceration occurs in the same patient population. Absorbable sutures do not last long enough to be effective, and even heavy polypropylene will give way in 3 to 4 weeks in most patients. A linear staple line across the outside of the pylorus provides the most enduring pyloric closure.

Pancreas

Blunt pancreatic transection at the neck of the pancreas can occur with a direct blow to the abdomen. As an isolated injury, it is more difficult to detect than blunt duodenal rupture, but a missed pancreatic injury is more benign. Because the main pancreatic duct is transected, the patient develops a pseudocyst or pancreatic ascites; there is minimal inflammation because the pancreatic enzymes remain inactivated. The diagnosis occasionally can be made with CT using fine slices through the pancreas. CT will not identify a significant number of transections if performed within 6 h of injury.

Optimal management of pancreatic trauma is determined by the location of the injury and whether or not the main pancreatic duct is injured. Pancreatic injuries in which the pancreatic duct is not injured may be treated by drainage or left alone. In contrast, pancreatic injuries associated with a ductal injury always require treatment to prevent pancreatic ascites or a major external fistula. Direct exploration of perforations or lacerations confirms the diagnosis of a ductal injury in most instances. This leaves a small but significant percentage of patients whom the diagnosis is in doubt and in whom more invasive investigations may be required. One recommendation has been to perform operative pancreatography. This procedure requires direct access to the duct by way of a duodenotomy or after resection of the tail of the pancreas. The duct is cannulated with a 5F pediatric feeding tube, and 2 to 4 mL of full-strength contrast medium is slowly injected; injuries are identified by obstruction or extravasation. Care must be taken to avoid overdistention of the duct with contrast medium, which can produce pancreatitis. The obvious shortcoming of this approach is the creation of a duodenal wound that must heal in a less than optimal environment. While those who advocate transduodenal pancreatography have had few duodenal fistulas, some have occurred. The problems associated with lateral duodenal fistulas are sufficient to dampen enthusiasm for this approach. If the patient already has a duodenal wound in the second portion, the above objections to pancreatography are mitigated.

An expeditious alternative to pancreatography is to pass a 1.5 to 2.0 mm coronary artery dilator into the main duct via the papilla and observe the pancreatic wound. If the dilator is seen in the wound, a ductal injury is confirmed. When inserted through the papilla of Vater, care must be taken to ensure that the dilator enters the pancreatic duct and not the bile duct. This can be determined by palpation of the hepatic pedicle. The limitations of this approach are the same as those for pancreatography.

A third method for identifying pancreatic ductal injuries is the use of *endoscopic retrograde pancreatography (ERP)*. This technique may be difficult to perform in an anesthetized patient in the operating room, but the surgeon can assist by manipulating the duodenum or occluding the distal portion to facilitate air insufflation. ERP is very helpful in the delayed diagnosis of a ductal injury or in those patients who are too sick to explore adequately during the initial operation.

No ideal method exists for identifying pancreatic ductal injuries that cannot be ruled out by direct exploration. This dilemma tends to encourage aggressive local exploration, which may create a ductal injury where none existed. For injuries involving the neck, body, or tail of the pancreas, this is of minor consequence because a simple resection distal to the injury cures the lesion. This is not the case for injuries to the head of the pancreas, which cannot be treated with a simple resection. Rather than accepting the risks of pancreatography or aggressive local exploration, a final option for identifying ductal injuries in the head of the pancreas is to do nothing other than drain the pancreas (Fig. 6-54). If a pancreatic fistula or pseudocyst develops, the diagnosis is confirmed. The majority of pancreatic fistulas close spontaneously with only supportive care. The authors prefer this approach to operative pancreatography when the diagnosis of ductal injury in the head of the pancreas is not apparent and *ERP* is not promptly available.

Several options are available for treating injuries of the neck, body, and tail of the pancreas when the main duct is transected. Distal pancreatectomy with splenectomy has been the preferred approach, but increasing interest in splenic preservation has stimulated the use of the splenic-preserving distal pancreatectomy. This procedure is performed by dissecting the pancreas from the splenic vein. Another method for splenic preservation is to bury the distal transected end of the pancreas in a Roux-en-Y limb. This technique also conserves the distal pancreas but is seldom performed because of the added complexity of the Roux-en-Y and the risks of pancreatojejunostomy.

For injuries to the head of the pancreas that involve the main pancreatic duct but not the intrapancreatic bile duct, there are few options. Distal pancreatectomy is rarely indicated because the risk of pancreatic insufficiency is significant if more than 85 to 90 percent of the gland is resected. A more limited resection from the site of the injury to the neck of the pancreas, with preservation of the pancreaticoduodenal vessels and common duct, allows closure of the injured proximal pancreatic duct. Pancreatic function can then be preserved by a Roux-en-Y pancreatojejunostomy with the distal pancreas (Fig. 6-55).

In contrast to injuries of the pancreatic duct, diagnosis of injuries to the intrapancreatic common bile duct is simple. The first method is to squeeze the gallbladder and observe the pancreatic wound. If bile is seen leaking from the pancreatic wound, the presence of an injury is established. Operative cholangiography is diagnostic in questionable cases. If a patient with an intrapancreatic bile duct injury is critically ill from hemorrhage, external drainage can be used until the patient is fit for definitive treatment. Small tangential perforations of the intrapancreatic bile duct may heal with this treatment, but it is seldom recommended. Most authorities advocate division of the common bile duct superior to the first portion of the duodenum, ligation of the distal common duct, and reconstruction with a Roux-en-Y choledochojejunostomy.

The use of drains has played an important role in the management of pancreatic injuries. Many authorities advocate routine drainage of all pancreatic injuries, but the authors do not drain contusions, lacerations in which the probability of a major ductal injury is small, or pancreatic anastomoses. Pancreatic injuries should be drained when there is a possible unidentified major ductal injury. If a drain is desirable, prospective studies have demonstrated that closed-suction devices are associated with fewer infectious complications than sump or Penrose drains. Almost all pancreatic fistulas close spontaneously. Nutritional support is important, and electrolyte replacement may be necessary.

Pancreaticoduodenal Injuries

Because the pancreas and duodenum are in physical contact, combined pancreaticoduodenal injuries are not uncommon, particularly in penetrating trauma. These lesions are dangerous because of the risk of duodenal suture line dehiscence and the development of a lateral duodenal fistula. The simplest treatment is to repair the duodenal injury and drain the pancreatic injury. This method is appropriate for combined injuries without major duodenal tissue loss and without pancreatic or biliary ductal injuries. With more extensive injuries, consideration should be given to providing additional protection for the duodenal suture line. The authors prefer pyloric exclusion to other alternatives.

While most pancreatic and duodenal injuries can be treated with relatively simple procedures, a few require extensive operations, such as pancreatoduodenectomy (Fig. 6-56). Examples of such injuries include transection of the intrapancreatic bile duct and the main pancreatic duct in the head of the pancreas, avulsion of the papilla of Vater from the duodenum, and destruction of the entire second portion of the duodenum. Most injuries of that nature are caused by higher-energy gunshot wounds. In patients with a pancreaticoduodenal injury who also have an intrapancreatic bile duct injury, it is possible to use the combination of a pyloric exclusion and Roux-en-Y choledochojejunostomy to avoid a pancreatoduodenectomy. The complexity and unpredictable physiology of the combined procedures makes the pancreatoduodenectomy more attractive.

Colon

There are three conceptually different methods for treating colonic injuries: primary repair, colostomy, and exteriorized repair. Primary repairs include lateral suture of perforations and resection of the damaged colon with reconstruction by ileocolostomy or colocolostomy. The advantage of primary repairs is that definitive treatment is carried out at the initial operation. The disadvantage is that suture lines are created in suboptimal conditions, and leakage may occur. Several different styles of colostomies are used to manage colonic injuries. In some instances, the injured colon can be exteriorized like a loop colostomy. The injured area can be resected and an end colostomy or ileostomy performed; the distal colon can be brought to the abdominal wall as a mucous fistula or oversewn and left in the abdominal cavity. A loop colostomy can be created proximal to a suture line that is left in the abdominal cavity.

The advantage of colostomy is that it avoids an unprotected suture line in the abdomen. The disadvantage is that a second operation is required to close the colostomy. Often overlooked disadvantages are the complications associated with the creation of a colostomy, some of which may be fatal. Exteriorized repairs are created by the suspending of a repaired perforation or anastomosis on the abdominal wall with an appliance after the fashion of a loop colostomy. If after 10 days the suture line does not leak, it can be returned to the abdominal cavity under local anesthesia without subsequent risk of leakage. If the repair breaks down before 10 days, it is treated as a loop colostomy. Healing is successful in 50 to 60 percent of cases. The advantage is avoidance of an intraperitoneal suture line when it is at risk of leakage, and the disadvantage is that 40 to 50 percent of patients require colostomy closure. Stomal complications similar to those of colostomies also can occur with the exteriorization.

Numerous large retrospective and several prospective studies have demonstrated that primary repair is safe and effective in the majority of patients with penetrating injuries. Colostomy is appropriate in a few patients but the current dilemma is how to select them. Exteriorized repair is no longer indicated because most patients who were once candidates for this treatment are successfully managed by primary repair. Two methods have been advocated that result in 75 to 90 percent of penetrating colonic injuries being safely treated by primary repair. The first is to repair all perforations that do not require resection. If resection is required because of the local extent of the injury and it is proximal to the middle colic artery, the proximal portion of the right colon up to and including the injury is resected and an ileocolostomy performed. If resection is required distal to the middle colic artery, an end colostomy is created and the distal colon oversewn and left within the abdomen. Ileocolostomy heals more reliably than colocolostomy because, in the trauma patient who has suffered shock and may be hypovolemic, assessing the adequacy of the blood supply of the colon is much less reliable than in elective procedures. The blood supply of the terminal ileum is never a problem. Another approach is to repair all injuries regardless of the extent and location (including colocolostomy) and reserve colostomy for patients with protracted shock and extensive contamination. Systemic factors are more important than local factors in determining whether a suture line heals. Both of these approaches are reasonable and result in the majority of patients being treated by primary repairs. When a colostomy is required, performing a loop colostomy proximal to a distal repair should be avoided because a proximal colostomy does not protect a distal suture line. All suture lines and anastomoses are performed with the running single-layer technique described in Fig. 6-57.

Complications related to the colonic injury and its treatment may include intraabdominal abscess, fecal fistula, wound infection, and stomal complications. Intraabdominal abscess occurs in approximately 10 percent of patients, and most are managed with percutaneous drainage. Fistulas occur in 1 to 3 percent of patients and usually present as an abscess or wound infection, which, after drainage, is followed by continuous fecal output. Most colonic fistulas heal spontaneously. Wound infection can be effectively avoided by leaving the skin and subcutaneous tissue open and relying on healing by secondary intention. The skin can be closed primarily in approximately 60 percent of patients without developing an infection. This treatment should be reserved for injuries with little contamination and in patients with minimal blood loss and little subcutaneous fat.

Stomal complications include necrosis, stenosis, obstruction, and prolapse. Taken together they occur in approximately 5 percent of patients, and most require reoperation. Necrosis is a serious complication that must be recognized and treated promptly. Failure to do so can result in life-threatening septic complications, including necrotizing fasciitis.

Rectum

Rectal injuries are similar to colonic injuries with respect to the ecology of the luminal contents, the structures and blood supply of the wall, and the nature and frequency of complications. They differ in two important ways: mechanisms of injury and accessibility. The rectum is often injured by gunshot wound, rarely by stab wound, and frequently by acts of autoeroticism and sexual misadventure. The rectum also is subject to high-pressure injuries that can be caused by air guns or water under high pressure, as used in golf course irrigation systems. The second difference is limited access to the rectum because of the surrounding bony pelvis.

The diagnosis is suggested by the course of projectiles, the presence of blood on digital examination of the rectum, and history. Patients in whom a rectal injury is suspected should undergo proctoscopy. Hematomas, contusions, lacerations, and gross blood may be seen. If the diagnosis is in question, x-ray examinations with soluble contrast medium enemas are indicated. It may be difficult to determine whether an injury is present. These patients should be treated as though an injury were present.

The portion of the rectum proximal to the peritoneal reflection is referred to as the intraperitoneal segment, and that portion distal to the reflection as the extraperitoneal segment. This distinction is blurred somewhat because the broad posterior aspect of the intraperitoneal portion could be considered as either. Injuries of the intraperitoneal portion (including its posterior aspect) are treated as previously outlined in the section on colonic injuries. Access to extraperitoneal injuries is so restricted, especially in the narrow male pelvis, that indirect treatment usually is required. While colostomies proximal to a suture line are avoided in patients with colonic injuries, there is often no option in patients with extraperitoneal injuries; sigmoid colostomies are appropriate for most patients. Properly constructed loop colostomies are preferred because they are quick and easy to fashion, and they provide total fecal diversion. Essential elements include: (1) adequate mobilization of the sigmoid colon so that the loop rests on the abdominal wall without tension; (2) maintenance of the spur of the colostomy (the common wall of the proximal and distal limbs after maturation) above the level of the skin with a 1 cm nylon rod or similar device; (3) longitudinal incision in the taeniae coli; and (4) immediate maturation in the operating room using 3-0 braided absorbable suture (Fig. 6-58). A staple line can be applied across the distal limit to ensure complete diversion, but it is not necessary, and it complicates closure of the colostomy. A mucous fistula is never required and should be avoided because of the risk of necrosis if the inferior mesenteric or superior rectal arteries were injured or otherwise ligated.

If a perforation is inadvertently uncovered during dissection, it should be repaired as described previously. Otherwise, it is not necessary to explore the extraperitoneal rectum to repair perforation. It may be extremely difficult or impossible to accomplish exploration. If the injury is so extensive that it must be repaired, the rectum is divided at the level of injury, the distal rectum is oversewn or stapled, and an end colostomy is created (Hartmann's procedure). In rare instances in which the anal sphincters have been destroyed, an abdominoperineal resection may be necessary.

Extraperitoneal injuries of the rectum should be drained via a retroanal incision (see Fig. 6-58). Waldeyer's fascia is particularly tough at this level and may need to be sharply incised. The drains, Penrose or closed-suction, should be placed close to the perforation or suture line and left until they fall out spontaneously or drainage diminishes, which usually occurs within 7 to 10 days. Irrigation of the distal rectum with various solutions is advocated by some authorities; this has not been determined to be helpful or harmful in retrospective studies, but it may be of benefit in a patient whose rectum is full of feces. If it is done, the irrigation solution should be isotonic and the anus should be mechanically dilated to avoid building up pressure that might force feces out of an unrepaired perforation. If the patient has a concomitant bladder injury and adjacent suture lines are created, a flap of viable omentum should be placed between them to reduce the risk of a rectovesical fistula.

There have been reports of treating small extraperitoneal rectal injuries by suture or drainage alone. The outcomes have been acceptable, and colostomies have been avoided. There is insufficient experience to recommend this approach because pelvic sepsis associated with rectal injury is highly lethal.

Complications of rectal injuries are similar in nature and frequency to those of colonic injuries. Pelvic osteomyelitis also may occur. Bone biopsy and bacteriology should be performed to secure the diagnosis. Culture-specific intravenous antibiotics should be administered for 2 to 3 months. Debridement might be necessary.

Stomach and Small Intestine

Injuries of the stomach and small bowel pose no special problems or controversies. Gastric injuries can be missed occasionally if a wound is located within the mesentery of the lesser curvature or high in the posterior fundus. The stomach should be clamped at the pylorus and inflated with air or methylene-blue-colored saline solution if there is any question. Patients with injuries that damage the nerves of Latarjet or both vagus nerves should have a drainage procedure. If the distal antrum or pylorus is severely damaged it can be reconstructed with a Billroth I or II procedure. A running two-layer suture line is preferred for the stomach because of its rich blood supply and because postoperative hemorrhage has occurred when the single-layer technique has been used in the stomach.

There are no special issues in treating injuries of the small bowel. Wounds of the mesenteric border can be missed if the exploration is not comprehensive. Most injuries are treated with a lateral single-layer running suture. Multiple penetrating injuries often occur close together. Rather than performing many lateral repairs, judicious resections with end-to-end anastomosis can save considerable time.

Kidneys

Three imaging techniques—CT, *intravenous pyelography (IVP)*, and arteriography—can be used to evaluate accurately the extent of renal injury. The contrast medium required for each is nephrotoxic and limits the number of studies that can be performed. The fact that there are two identical organs makes the sacrifice of one a viable therapeutic option. Nearly 95 percent of all blunt renal injuries are treated nonoperatively. The diagnosis is suspected by the finding of microscopic or gross hematuria and confirmed by CT or IVP. Most cases of urinary extravasation and hematuria resolve in a few days with bed rest. Persistent gross hematuria can be treated by embolization. Persistent urinomas can be drained percutaneously. Operative treatment is necessary occasionally for similar lesions that do not respond to these less invasive measures.

If a perinephric hematoma is encountered during laparotomy from blunt trauma, exploration is indicated if it is expanding or pulsatile. Very large hematomas should be explored because of the risk of a major vascular injury. Vascular control at the junction of the renal vessels with the aorta and vena cava is not always necessary before entering the hematoma. If emergent vascular control is needed, a large curved vascular clamp can be placed easily across the hilum from below with the clamp parallel to the vena cava and aorta.

Hemostatic and reconstructive techniques used to treat blunt renal injuries are similar to those used to treat the liver and spleen. The collecting system should be closed separately and the renal capsule preserved to close over the repair of the collecting system (Fig. 6-59). Permanent sutures should be avoided because of the risk of calculus formation. The authors prefer absorbable monofilament sutures because of their lack of abrasiveness. If nephrectomy is being considered and the status of the opposite kidney is unknown, the latter should be palpated. The presence of a palpably normal opposite kidney is assurance that the patient will not be rendered anephric by a unilateral nephrectomy. Unilateral renal agenesis occurs in 1/1000 patients.

The renal arteries and veins are uniquely susceptible to traction injury caused by blunt trauma. As the artery is stretched, the inelastic intima and media may rupture. This causes thrombus formation, resulting in high-grade stenosis or thrombosis. The injury can be detected by CT, *IVP*, or duplex scanning. If the patient does not have more urgent injuries and treatment and repair can be accomplished within 3 h of admission, it should be attempted. Successful renal artery repair in a patient who presents with complete thrombosis is rare. If repair is not possible within this time frame, leaving the kidney in situ to resorb does not necessarily lead to hypertension or abscess formation. Isolated renal vein injuries can occur from blunt trauma. The vein may be torn or avulsed from the vena cava, and a large hematoma develops that often leads to an operation and nephrectomy.

All penetrating wounds to kidneys are explored. Bleeding perforations and lacerations are treated using the same hemostatic techniques described above. Renal vascular injuries are common after penetrating trauma, and they may be deceptively tamponaded and result in delayed hemorrhage. Injuries involving the collecting system should be closed separately if they are large. Small perforations that penetrate the collecting system can be controlled by suture of the capsule and parenchyma. Perforations of the renal pelvis should be meticulously repaired with fine sutures.

Ureters

Injuries of the ureters from external trauma are rare. They occur in a few patients with pelvic fractures and are uncommon in penetrating trauma because the silhouette they present is so small. The diagnosis in blunt trauma may be made by CT, *IVP*, or retrograde ureterography. The injury often is not identified until a complication, e.g., a urinoma, is apparent. In penetrating trauma, ureteral injuries are discovered during the exploration of the retroperitoneum, although missed injuries also are not unusual. If an injury is suspected but not identified, methylene blue or indigo carmine is administered intravenously. Staining of the tissue adjacent to the injury can facilitate identification of the injury site. Most injuries can be repaired primarily using the same technique as that described earlier for small arteries, using 5-0 absorbable monofilament suture. When the ureter is mobilized, the dissection should be at least 1 cm lateral and medial to the ureter to avoid injury to its delicate vascular plexus. The kidney also can be mobilized to increase ureteral mobility. Injuries of the distal ureter can be treated by reimplantation. The psoas hitch and Boari flap may be helpful in selected distal ureteral injuries (Fig. 6-60). If the patient is critically ill and being considered for a staged laparotomy or if the surgeon is uncomfortable with ureteral repair, the ureter can be ligated on both sides of the injury and a nephrostomy performed (Fig. 6-61).

Bladder

Bladder injuries are diagnosed by cystography, CT, or during laparotomy. A postvoid view enhances the accuracy of cystography. Blunt ruptures of the intraperitoneal portion are closed with a running single-layer closure using 3-0 absorbable monofilament suture. Blunt extraperitoneal rupture is treated with a Foley catheter; direct operative repair is not necessary. Cystograms can be used to determine when the catheter can be removed, which is usually in 10 to 14 days. Penetrating bladder injuries are treated in the same fashion, although injuries near the trigone should be repaired through an incision in the dome so that iatrogenic injury to the intravesicular ureter is avoided by direct visualization.

Urethra

Blunt disruption of the posterior urethra is managed by bridging the defect with a Foley catheter. This usually requires passing catheters through the urethral meatus and through an incision in the bladder. Once the catheter bridges the defect, healing occurs as the intervening hematoma resorbs. Strictures are not uncommon but can be managed electively. Penetrating injuries are treated by direct repair.

Gynecologic Injuries

Gynecologic injuries are rare. Occasionally, the vagina is lacerated by a sharp bone fragment from a pelvic fracture. Penetrating injuries to the vagina, uterus, fallopian tubes, and ovaries also are uncommon. The usual hemostatic techniques are used to control bleeding, and suture repair is used to close defects that communicate with a lumen. Repair of a transected fallopian tube can be attempted but probably is unjustified; a suboptimal repair increases the risk of tubal pregnancy. Transection at the injury site with proximal ligation

and distal salpingectomy is a more prudent approach.

Trauma in pregnancy also is rare. Blunt trauma can cause uterine rupture, which almost always results in fetal demise. The outcome of penetrating uterine injuries is more variable and is dependent on penetration of the uterine cavity, damage to the placenta, and fetal injury. Spontaneous abortion is a frequent outcome. On occasion, a mother presents with life-threatening injuries, including severe head injury or cardiac arrest from hemorrhagic shock. If the fetus is viable by dates or examination, an emergency cesarean section should be considered whether or not the mother's life can be saved. This occurs more often with severe head injury than cardiac arrest from hemorrhagic shock.

Completion of the Laparotomy and Postoperative Considerations

After repair of all injuries, the abdomen is irrigated with saline warmed to body temperature. This will not eliminate all bacteria, but an effort should be made to remove blood clots, food particles, and gross enteric and fecal contamination.

Patients with moderate to severe injuries are at risk for multiple organ failure and nosocomial infection. The integrity of the gut has a pivotal role in the severity and outcome of these complications. Needle-catheter jejunostomies are placed in all such patients before abdominal closure (Fig. [6-62](#)). Enteric feedings are initiated as soon as the patient arrives in the intensive care unit and are advanced to full strength within 72 h. Total parenteral nutrition might be necessary in some patients, but it causes mucosal atrophy, which may impair the barrier function of the mucosa.

The abdominal incision is closed with a running No. 2 nylon suture that includes at least 1.5 cm of fascia. Tension of the suture should be just enough to approximate the fascia but no more. Subcutaneous sutures are never used. The skin is closed selectively, depending on the amount of contamination and subcutaneous tissue. The greater the degree of contamination and subcutaneous fat, the more frequently the skin and subcutaneous tissue should be left open. Patients initially treated with staged laparotomy should have the skin and subcutaneous tissue left open.

Pelvis

Pelvic fractures can cause exsanguinating retroperitoneal hemorrhage without associated major vascular injury; branches of the internal iliac vessels and the lower lumbar arteries are often responsible. Hemorrhage also comes from small veins and from the cancellous portion of the fractured bones. A direct surgical approach is rarely effective because many of the sources of hemorrhage are outside of the surgical field. Most pelvic fractures that cause life-threatening hemorrhage involve disruption of the posterior elements, i.e., the sacroiliac joints and associated ligaments, and are often biomechanically unstable.

A hemodynamically unstable patient with an unstable pelvic fracture may be bleeding from sources other than the pelvis, such as the spleen. Large retroperitoneal hematomas also can cause a hemoperitoneum, particularly if overlying peritoneum ruptures. Determining the source of hemorrhage poses a therapeutic problem because it is desirable not to operate for a retroperitoneal hematoma, whereas a laparotomy may be essential to deal with hemorrhage from the spleen or liver. Ultrasonography and *DPL* have been used to aid in this decision. If 10 mL or more of free blood can be aspirated from the peritoneal cavity, or if the ultrasound scan is unequivocally positive, then the blood is assumed to be coming from an injury unrelated to the pelvic fracture, and a laparotomy is performed. If the DPL is positive by laboratory analysis or if it is negative, attention is directed toward treating the pelvic fracture. The decision to operate or not may prove to have been the wrong course, and the plan may need to be altered accordingly.

Several methods have been used to control hemorrhage associated with pelvic fractures. These include immediate external fixation, medical antishock trousers (MAST), angiography with embolization, and pelvic packing. No single technique is effective for treating all fractures, and there is little agreement among specialists as to which should be used. Anterior external fixation is not intended to provide definitive fracture stabilization in most instances. Its advocates intend for the device to decrease pelvic volume, to tamponade bleeding, and to prevent secondary hemorrhage that may occur if the fractured bones shift. Many orthopaedic surgeons are unconvinced of the efficacy of external fixation for grossly unstable posterior fractures. Antishock trousers can provide some stability for the fracture and probably tamponade venous hemorrhage. The disadvantages are the loss of access to the abdomen and the risk of lower-extremity compartment syndrome. Angiography with embolization is very effective for controlling arterial hemorrhage, but arterial hemorrhage occurs in only 10 to 20 percent of patients with active hemorrhage from pelvic fractures. Pelvic packing may control venous hemorrhage. The only reason to consider its use is when a pelvic hematoma is inadvertently entered or if it has ruptured. An algorithm for managing patients is shown in Fig. [6-63](#).

Another challenge is the open pelvic fracture. In many instances the wounds are located in the perineum, and the risk of pelvic sepsis and osteomyelitis is high. To reduce the risk of infection, a sigmoid colostomy is recommended. The pelvic wound is manually debrided and irrigated daily with a high-pressure pulsatile irrigation system until granulation tissue covers the wound. The wound is then left to heal by secondary intention. This approach has been highly successful.

Extremities

Vascular Injuries with Fractures

Vascular injuries associated with fractures are rare, occurring in only 0.5 to 3 percent of all patients with extremity fractures. They also are more severe than isolated vascular injuries or fractures, and amputation rates of more than 50 percent have been noted. These injuries can be caused by blunt and penetrating trauma. Particular fractures and dislocations are more likely to be associated with vascular injury than others. In the upper extremity, a fracture of the clavicle or the first rib may lacerate the distal subclavian artery. The axillary artery may be injured in patients with dislocations of the shoulder or proximal humeral fractures. Supracondylar fractures of the distal humerus and dislocations of the elbow are known for their association with brachial artery injuries. In all these fractures and dislocations, vascular injuries are uncommon and occur in only a small fraction of patients.

In the lower extremity, the orthopaedic injury most commonly associated with vascular injury is dislocation of the knee, in which the popliteal artery or vein may be injured in as many as 30 percent of patients. The popliteal vessels also may be injured in patients with supracondylar fractures of the femur or tibial plateau fractures. Vascular injury can occur in patients with combined fractures of the tibia and fibula.

The importance of a careful neurologic examination in these patients is critical. Three distinctly different mechanisms can produce paralysis and numbness in an injured extremity: ischemia, nerve injury, and compartment syndrome. As a result, failure to accurately perform and document the neuromuscular function of the injured extremity can lead to missed injuries, improper treatment, and unrealistic expectations on the part of the patient.

Perhaps the greatest controversy in the treatment of patients with combined orthopaedic and vascular injuries is the order in which the procedures are to be performed. Advocates of initial fracture treatment argue that it is difficult to judge the length of a vascular graft (or whether one is required) when the ends of the fractured bone are overriding or if angulation is present. Also, extensive orthopaedic manipulation can easily disrupt delicate vascular repairs. Opponents of this approach argue that the length of time required to stabilize the fracture may cause further ischemic damage of the limb. The use of temporary intravascular shunts has been recommended as a compromise to avoid ischemia during fracture treatment.

A rational approach is to consider all the above options in light of the condition of the patient's injured extremity. If the extremity is clearly viable and there is no hemorrhage from the vascular injury, the fracture should be treated first. If the limb is at risk from ischemia, prompt revascularization is required. When little or no fracture manipulation is anticipated, definitive vascular repair is performed first. If extensive manipulation is required in an ischemic extremity, temporary shunts can be placed and vascular repair performed after the fracture has been treated (Fig. [6-64](#)).

The extent of injury causing combined orthopaedic and vascular injuries frequently results in open fractures, and the use of external fixation devices for these injuries has become common. These devices may significantly hinder vascular repair because of their location and bulk. Preoperative planning between the vascular and orthopaedic surgeon should serve to avoid this technical problem.

Because of the severity of combined orthopaedic and vascular injuries, the need for immediate amputation might arise. Primary amputation should be strongly considered when the primary nerve is transected in addition to a fracture and arterial injury—such as when the popliteal nerve is transected along with injuries to the popliteal artery and the distal femur. This difficult decision is best reached through a collaborative effort involving the trauma surgeon, the orthopaedic surgeon, and, in certain cases, the neurosurgeon. Prolonged rehabilitation resulting in a paralyzed, anesthetic extremity prone to ulceration is hardly better than the prompt fitting of a good prosthesis.

Compartment Syndrome

A compartment syndrome can occur anywhere in the extremities, including the thighs, buttocks, arms, and hands. The pathophysiology is an acute increase in pressure in a closed space that impairs blood flow to the structures within. The causes of extremity compartment syndrome include arterial hemorrhage into a compartment, venous ligation or thrombosis, crush injuries, infections, crotalid envenomation, and ischemia/reperfusion. In conscious patients, pain is the prominent symptom. Active or passive motion of involved muscles increases the pain. Progression to paralysis can occur. The most frequent site is in the anterior compartment of the leg; a well-described early sign is paresthesia or numbness between the first and second toes caused by pressure on the deep peroneal nerve.

In comatose or obtunded patients, the diagnosis is more difficult to secure. A compatible history, firmness of the compartment to palpation, and diminished mobility of the joint are suggestive. The presence or absence of a pulse distal to the affected compartment is notoriously unreliable in the diagnosis of a compartment syndrome. A frozen joint and myoglobinuria are late signs and suggest a poor prognosis. Compartment pressure can be measured. The small, hand-held Stryker manometer is a convenient tool for this purpose. Pressures greater than 45 mmHg usually require operative intervention. Pressures between 30 and 45 mmHg should be carefully evaluated and watched closely.

Treatment consists of measures to reduce compartment pressure, including elevation of the extremity, evacuation of hematomas, and fasciotomy. As long as neurologic and muscular function are intact, elevation and observation are sufficient. The evacuation of hematomas as a consequence of arterial injury almost always results in a fasciotomy, because the compartment must be opened to treat the vascular injury. Because the lower extremity is most frequently involved, the two-incision, four-compartment fasciotomy is shown in Fig. [6-65](#). Note that the soleus muscle must be detached from the tibia to decompress the deep flexor compartment.

Prognosis is related to the severity, duration, and cause of the compartment syndrome. The best results are obtained in patients with arterial hemorrhage and venous ligation or thrombosis who undergo early fasciotomy. Those who develop compartment syndrome from crush injuries, crotalid envenomation, and particularly ischemia/reperfusion have a poor prognosis because of the preexisting muscle and nerve damage caused by the original insult. Fasciotomy should be attempted, although infection and amputation are a frequent outcome.

Prognosis and Outcome Evaluation

Prognosis and outcome evaluation for various injuries began during World War I. At that time, mortality was calculated according to the organ injured. For example, all patients who

suffered an injury to the colon were determined to have lived or died. Accordingly, frequency of death was assigned to that particular organ, and it was assumed that any patient with a colonic injury had the same probability of dying. Because no other factors such as associated injuries or physiologic condition were considered, it is not surprising that any abdominal visceral injury was associated with a mortality rate of 50 to 60 percent. This practice continued during World War II and resulted in some remarkable conclusions that were subsequently shown to be incorrect. Perhaps the best example of this was the conclusion that performing colostomies for all colonic injuries during World War II resulted in a reduction of mortality for patients so treated from 60 percent during the first World War I to 30 percent in World War II. Civilian literature in the second half of this century recognized that the number of injured organs, major fractures, blood loss, and the presence of shock were all predictive of outcome but only in the crudest fashion. The quality of the local EMS system is another confounding factor. Regions with rapid response and transport are more likely to bring severely injured patients to the emergency room with signs of life than less efficient systems. As a result, mortality is paradoxically greater in regions with better EMS systems.

Today there are anatomic and physiologic grading systems. The anatomic systems are derived from the Abbreviated Injury Scale (AIS), which was developed during the 1950s. The AIS is a list that assigns a number from 1 (minor injury) to 6 (always fatal) for the various spectra of organ injuries. The AIS evaluates only solitary injuries; it cannot reflect the additional impact of multiple injuries. The Injury Severity Score (ISS) was devised to supplement this shortcoming. The ISS is calculated by squaring the AIS from the worst injured of three body compartments (head and neck, face, chest, abdomen and pelvis, pelvic girdle, and extremities) and adding them together. With the always-fatal score of 6 excluded, scores can range from 1 to 75. ISS is further characterized according to the mechanism of injury (blunt versus penetrating) and age (less than 55 years versus 55 years or greater). The ISS, however, suffers from the inability to consider multiple injuries in one compartment, the assumption that all compartments are of equal significance, and the lack of recognition of the patient's physiologic status. Several physiologic scoring systems have been developed. The Revised Trauma Score (RTS) is most commonly used. It is calculated from the Glasgow Coma Scale, blood pressure, and respiratory rate, with the GCS being most heavily weighted. The RTS is a purely physiologic score that is compromised by the relative insensitivity of these common clinical measurements. The TRISS method (Trauma and Injury Severity Score) was developed to incorporate the RTS and ISS, thereby combining physiologic and anatomic score and enhancing the importance of head injury. TRISS remains fundamentally flawed because of the limitations of RTS and ISS. Newer versions of the TRISS concept, such as ASCOT (A Severity Characterization of Trauma), also have failed to improve the prediction of postinjury mortality. Perhaps more important than mortality, outcome assessment must include the critical issues of total medical resource consumption (complications, hospital length of stay, cost for medical care, etc.) and capacity to return to preinjury functional status. Several functional outcome scales have been developed, but a standard has yet to be established.

BITES AND STINGS OF ANIMALS AND INSECTS

Rabies

In 1950 approximately 5000 cases of rabies were reported among dogs and 18 were reported in humans. Only 160 cases of rabies in dogs were reported in 1989. In 1991 there were three patients in the United States who died from rabies. Wild animals, therefore, constitute the most important potential source of infection for humans and domestic animals in the United States; however, the exposure that results from frequent contact between domestic dogs and humans continues to be the basis of most antirabies treatment.

Approximately 10,000 patients receive postexposure prophylaxis for rabies annually. Rabies among wild animals, especially skunks, foxes, raccoons, and bats, account for more than 85 percent of known cases of animal rabies. Although any mammal may carry rabies, rodents are seldom found to be infected with rabies and have not been known to cause rabies in human beings in the United States. Woodchucks accounted for 70 percent of rabies among rodents reported to the Centers for Disease Control and Prevention (CDC). In all cases involving rodents, the state or local health department should be consulted before a decision is made to initiate postexposure antirabies prophylaxis. Many of the cases of human rabies reported in the past 10 years have resulted from exposure outside of the United States; in much of the rest of the world the dog is the major species with rabies and the major source of rabies among human beings.

Circumstances surrounding the attack frequently furnish vital information as to whether or not vaccination is indicated. Most domestic animal bites are provoked attacks; if this history is obtained, rabies vaccine usually can be withheld if the animal appears healthy. Children are frequently bitten while attempting to separate fighting animals or while teasing or accidentally hurting the animal. Bites during attempts to feed or handle an apparently healthy animal are generally regarded as provoked. Postexposure prophylaxis combining local wound treatment, passive immunization, and vaccination is over 90 percent effective when appropriately applied. An unprovoked attack by a domestic animal is more likely than a provoked attack to indicate that the animal is rabid. A fully vaccinated dog or cat is unlikely to become infected with rabies, although rare cases have been reported.

Any penetration of the skin by teeth constitutes a bite exposure. Bites to the face and hands carry the highest risk, but the site of the bite should not influence the decision to begin treatment. Nonbites include scratches, abrasions, open wounds, or mucous membranes contaminated with saliva. If the material containing the virus is dry, the virus can be considered noninfectious. Other contact by itself, such as petting a rabid animal and contact with the blood, urine, or feces of a rabid animal, does not constitute an exposure and is not an indication for prophylaxis.

Most animal bites sustained by human beings are caused by dogs and cats, and in most instances it is possible to observe the biting animal for the development of rabies. Domestic animals that bite a person should be captured and observed for symptoms of rabies for 10 days. If none develop, the animal may be assumed to be nonrabid. If the animal dies or is killed, the head is sent promptly to a public health laboratory for examination. The tissue requires refrigeration, but not freezing, and transportation to the laboratory after the animal's death must be rapid. Clinical signs of rabies in wild animals cannot be interpreted reliably; therefore, any wild animal that bites or scratches a person should be killed at once (without unnecessary damage to the head) and the brain examined for evidence of rabies. Travelers to Asia, Africa, and Central and South America should be aware that more than 50 percent of the rabies cases among human beings in the United States result from exposure to dogs outside the United States (Table 6-9).

It is accepted that the incubation period for rabies in human beings ranges from 10 days to 1 year, with most cases occurring within 20 to 90 days of exposure. In cases of exposure of the head, neck, or upper extremities, the incubation period is potentially less than 30 days. Local care of the animal bite should consist of thorough irrigation, cleansing with soap solution, and debridement. Administration of tetanus toxoid and an antibiotic may be indicated.

Postexposure prophylaxis in addition to local wound treatment consists of *human rabies immune globulin (HRIG)* (Imogam Rabies) and vaccine. There are two rabies vaccines available in the United States: *human diploid cell rabies vaccine (HDCV)* or *rabies vaccine adsorbed (RVA)* (Imovax). Either is administered in conjunction with HRIG at the beginning of postexposure therapy. A regimen of five 1-mL doses of HDCV or RVA is given intramuscularly. The first dose of the five-dose course is given as soon as possible after exposure. Additional doses are given on days 3, 7, 14, and 28 after the first vaccination. For adults, the vaccine is always administered intramuscularly in the deltoid area. For children, the anterolateral aspect of the thigh also is acceptable. The gluteal area should never be used for HDCV or RVA injections because administration in this area results in lower neutralizing antibody titers.

Postexposure antirabies vaccinations should always include administration of passive antibody and vaccine, except for those who have previously received complete vaccine regimens with a cell culture vaccine or who have been vaccinated with other types of vaccines and have had documented rabies antibody titers; these persons should receive only vaccine. Because the antibody response after the recommended postexposure vaccination regimen has been satisfactory, routine postvaccination serologic testing is not recommended unless the patient is known to be immunosuppressed. The state health department can be contacted for recommendations.

HRIG is administered only once to provide immediate antibodies until the patient responds to the vaccine by actively producing antibodies. If HRIG was not given when vaccination was begun, it can be given through the seventh day after administration of the first dose of vaccine. Beyond the seventh day, HRIG is not indicated because the antibody response to cell culture vaccine is presumed to have occurred. The recommended dosage of HRIG is 20 IU/kg body weight. This formula is applicable for all age groups, including children. If anatomically feasible, up to one-half the dose of HRIG should be thoroughly infiltrated in the area around the wound, and the rest should be administered intramuscularly in the gluteal area. HRIG should never be administered in the same syringe or into the same anatomic site as vaccine. Because HRIG may partially suppress active production of antibody, no more than the recommended dose should be given (Table 6-10).

Local reaction such as pain, erythema, and swelling or itching at the injection site has been reported in 30 to 75 percent of recipients. Headache, nausea, abdominal pain, muscle aches, and dizziness have been reported from 5 to 40 percent of recipients. Cases of neurologic illness resembling Guillain-Barré syndrome that resolved have been reported. Local pain and low-grade fever may follow injections of *HRIG*. There is no evidence that hepatitis B virus, human immunodeficiency virus, or other viruses have ever been transmitted by commercially available HRIG in the United States. Corticosteroids can interfere with the development of active immunity after vaccination and may predispose the patient to rabies. When rabies postexposure prophylaxis is administered to persons receiving steroids or other immunosuppressive therapy, it is especially important that a serum sample be tested for rabies antibody to ensure that an acceptable antibody response has developed. Because of the potential consequences of inadequately treated rabies exposure, and because there is no indication that fetal abnormalities have been associated with rabies vaccination, pregnancy is not considered a contraindication to postexposure prophylaxis.

Manifestations and Treatment of Rabies

Symptoms of rabies include a 2- to 4-day prodromal period in which the patient reaches the excited stage. Paresthesia in the region of the bite is an important early symptom. Other symptoms include headaches, vertigo, stiff neck, malaise, lethargy, and severe pulmonary, symptoms including wheezing, hyperventilation, and dyspnea. The patient may have spasm of the throat muscles with dysphagia. The outstanding symptom of rabies is related to swallowing. Drooling, maniacal behavior, and convulsions ensue and are followed by coma, paralysis, and death. Intensive respiratory supportive care is essentially the only treatment to offer. Phenytoin (Dilantin) can be used for seizures.

Snakes

In North America all the poisonous snakes of medical importance are pit vipers, of the family Crotalidae, and the coral snake, of the Elapidae family. The pit vipers include rattlesnakes, cottonmouth moccasin, and the copperhead. Approximately 8000 persons are bitten each year by poisonous snakes, with over 98 percent of bites occurring on the extremities. Rattlesnakes are responsible for approximately 70 percent of deaths from snakebites, while death from the bite of a copperhead is extremely rare.

Pit Vipers

Pit vipers are named for the characteristic pit, a highly sensitive heat-sensing organ that is located between the eye and the nostril on each side of the head. These snakes may be identified by their elliptical pupil, as opposed to the round pupil of harmless snakes. Nonpoisonous snakes do not have pits. However, the coral snake does have a round pupil and lacks the facial pit. Pit vipers have two well-developed fangs that protrude from the maxillae, whereas most nonpoisonous snakes have rows of teeth without fangs. Pit vipers also may be identified by turning the snake's belly upward and noting the single row of subcaudal plates (Fig. 6-66). The coral snake is a small, brightly colored snake with red, yellow, and black rings. This color combination also occurs in nonpoisonous snakes, but the alternation of colors is different. Only the coral snake has a red ring next to a yellow ring. The nose of the coral snake is black.

The venoms of poisonous snakes consist of enzymatic complex proteins that affect all soft tissues. Venoms have been shown to have neurotoxic, hemorrhagic, thrombogenic, hemolytic, cytotoxic, antifibrinolytic, and anticoagulant effects. Most venoms contain hyaluronidase, which enhances the rapid spread of venom by way of the superficial lymphatics.

There may be considerable variation in the venom effect. Neurotoxic features such as muscle cramping, fasciculation, weakness, and respiratory paralysis or hemolytic characteristics may predominate, depending on the snake.

Pain from the bite of a pit viper is excruciating and probably the symptom that most easily differentiates poisonous from nonpoisonous snakebites. Pit vipers characteristically produce one or two fang marks. Hypotension, weakness, sweating and chills, dizziness, nausea, and vomiting are other systemic symptoms. Local signs and symptoms can include swelling, tenderness, pain, and ecchymosis and may appear within minutes at the site of venom injection. If no edema or pain is present within 30 min after injury, the snake probably did not inject any venom. Swelling may continue to increase for 24 h. Hemorrhage vesications, bullae, and petechiae may appear between 8 and 36 h, with thrombosis of superficial vessels and eventual sloughing of tissues. Systemic symptoms include paresthesias and muscle fasciculations. Muscle fasciculations are most common after a rattlesnake bite and often in the perioral region. Fasciculations rarely follow a copperhead bite or a cottonmouth bite. They are often seen in the face muscles and over the neck, back, and the involved extremity and can occur within 10 min.

The venom from rattlesnakes produces deleterious changes in the blood cells, defects in blood coagulation, injuries to the intimal linings of vessels, damage to the heart muscles, alterations in respiration, and to a lesser extent, changes in neuromuscular conduction. Pulmonary edema is common in severe poisoning, and hemorrhage into the lungs, kidneys, heart, and peritoneum can occur. Hematemesis, melena, changes in salivation, and muscle fasciculations may be seen. Urinalysis may reveal hematuria, glycosuria, and proteinuria. Red blood cells and platelets can decrease, and bleeding and clotting times usually are prolonged. Total afibrinogenemia is a hallmark of severe envenomation.

Blood should be immediately drawn for typing and crossmatching because hemolysis may later make this difficult. Because hemolysis and injury to kidneys and liver may occur, it is important to follow alterations in clotting mechanism, renal and liver function, and electrolyte status.

Coral Snakes

The coral snake contributes to only 3 percent of all bites and 1.5 percent of all deaths from poisonous snakes. Bites by the coral snake occasionally provoke blurred vision, ptosis, drowsiness, increased salivation, and sweating. The patient may notice paresthesia about the mouth and throat, sometimes slurring of speech, nausea, and vomiting. Pain is not a constant complaint, nor is edema a constant finding. Coral snake venom causes more extensive changes in the nervous system, and death may occur from inadequate ventilation.

Management of Snake Bites

Application of a tourniquet, incision, and suction are appropriate if used within 1 h of the time of the bite. The snake injects venom into the subcutaneous tissue, which is absorbed by capillaries and lymphatics. The tourniquet should be applied loosely to obstruct only venous and lymphatic flow. The tourniquet is not released once applied and may be left in place during the 30 min that suction is applied. The tourniquet may be removed after definitive treatment has been instituted and the patient is not in shock.

Incision and suction for 30 min may be beneficial if accomplished within 30 minutes after snakebite. The incision should be longitudinal and not cruciate. When two fang marks are seen, the depth of the venom injection is generally considered to be one-third of the distance between the fang marks. Severe bites may result in envenomations deep to the fascia, and surgical exploration may be indicated. Incisions made proximal to the bite are contraindicated.

The average snakebite does not require surgical excision. This procedure is reserved for the most severe envenomations. It has been shown that wide excision of the entire area around the snakebite within 1 h of the time of injection can remove most of the venom. Excision of the fang marks including skin and subcutaneous tissue should be considered in severe bites and in patients known to be allergic to horse serum and who are seen within 1 h of the bite. Most fatalities from snakebites do not occur for 6 to 48 h after the bite, giving time to institute other measures.

The most important treatment for a snakebite is antivenin, although many patients do not require it. Copperhead envenomation rarely necessitates antivenin. Most snakebite fatalities in the United States during the past 20 years have involved either delay in obtaining treatment, no antivenin treatment, or inadequate dosage. Because antivenin contains horse serum, before its administration skin testing is required. Epinephrine 1/1000 in a syringe should be available before antivenin is given.

Because the rattlesnake, cottonmouth moccasin, and copperhead belong to the same biologic family, their bites can be treated by the same antivenin (antivenin Crotalidae polyvalent). The coral snakebite is rare, and the antivenin is different from that for the pit viper. A North American coral snake (*Micrurus fulvius*) antivenin has been developed. It effectively treats *Micrurus* snakebites but is not effective in treating bites of *Micruroides*, the genus native to Arizona and New Mexico. Coral snake antivenin can be obtained from state health departments.

Information concerning identification of a snake or the proper antivenin frequently can be obtained from the nearest zoo herpetarium. A major problem with bites by exotic poisonous snakes is the choice and availability of suitable antiserum. Physicians confronted with this situation may obtain advice from the local poison center or the Antivenin Index Center of Oklahoma Poison Information Center, Oklahoma City, Oklahoma (405-271- 5454).

Antivenin should be withheld until a physician can determine whether it is indicated. Approximately 30 percent of all poisonous snakebites in the United States result in no envenomation. The indication for antivenin is governed by the degree of envenomation. With frequent observations using the classification presented in Table 6-11, the severity of the bite is often found to increase with time, and thus change in grade is observed. Most bites will have reached a final staging within 12 h. Antivenin usually is not required for grade 0 or I envenomation. Grade II may require 3 or 4 ampules, and grade III usually requires 5 to 15 ampules. If symptoms increase, several ampules may be required during the first 2 h. Proper dosage can be estimated by observing the clinical signs and symptoms. If systemic manifestations are severe, antivenin should be given rapidly, by intravenous drip, in large doses. The injection of antivenin locally around the bite is not advised.

If antivenin is indicated, 3 to 5 ampules are given by intravenous drip in 500 mL normal saline solution or 5% glucose solution. If severe systemic symptoms are already present, 6 to 8 ampules are given in addition. The dose of intravenously administered antivenin can be more easily titrated with response to treatment, and the amount administered is based on improvement in signs and symptoms, not on the weight of the patient. Antivenin is administered until severe local or systemic symptoms improve. When it is obvious that antivenin therapy will be instituted, the tourniquet should be left in place until antivenin is started.

If too much time has elapsed for excision to be effective and the patient is allergic to horse serum, a slow infusion of 1 ampule of antivenin in 250 mL of 5% glucose solution may be given in a 90-min period with constant monitoring of the blood pressure and electrocardiogram, depending on the seriousness of the bite. This is accomplished in an active emergency department or an intensive care unit. If an immediate reaction occurs, the antivenin is stopped, and a vasopressor and epinephrine may be required.

The incidence of serum sickness is directly related to the volume of horse serum injected. Of patients receiving 100 to 200 mL of horse serum, 85 percent have some degree of sensitivity within 8 to 12 days after injection.

Intravenous fluids are frequently required to replace the decreased extracellular fluid volume resulting from edema. Fascial planes may become very tense with obstruction of venous and later arterial flow, requiring fasciotomy. Adequate antivenin treatment usually makes surgical intervention unnecessary. These patients may need blood, since anemia can develop from the hematologic effects of envenomation. As afibrinogenemia has been reported, fibrinogen may be required. Vitamin K also may also be required. Bleeding and clotting abnormalities are treated with antivenin in addition to blood components. Antibiotics are recommended to prevent secondary infection, although their benefit is unproved. Tetanus toxoid is administered. The most common species of bacteria isolated from rattlesnake venom are *Pseudomonas aeruginosa*, *Proteus* species, *Clostridium* species, and *Bacteroides fragilis*.

Stinging Insects and Animals

Hymenoptera

The most important insects that produce serious and possibly fatal anaphylactic reactions are arthropods of the order Hymenoptera. This group includes the honeybee, bumblebee, wasp, yellow and black hornet, and the fire ant. The venom of these stinging insects is just as potent as that of snakes and causes more deaths in the United States yearly than are caused by snakebites.

Hymenopterans, except the bee, retain their stinger and are able to sting repeatedly, each time injecting some portion of the venom sac contents. The worker honeybee sinks its barbed sting into the skin and it cannot be withdrawn. As the bee attempts to escape, it is disemboweled. The stinger with the bowel, muscles, and venom sac attached is left behind. The muscles controlling the venom sac, although separated from the bee, rhythmically contract for as long as 20 min, driving the stinger deeper and deeper into the skin and continuing to inject the venom.

Symptoms consist of one or more of the following: localized pain, swelling, generalized erythema, a feeling of intense heat throughout the body, headache, blurred vision, injected conjunctivae, swollen and tender joints, itching, apprehension, urticaria, petechial hemorrhages of the skin and mucous membranes, dizziness, weakness, sweating, severe nausea, abdominal cramps, dyspnea, constriction of the chest, asthma, angioneurotic edema, vascular collapse, and possible death from anaphylaxis. Fatal cases may manifest glottal and laryngeal edema, pulmonary and cerebral edema, visceral congestion, meningeal hyperemia, and intraventricular hemorrhage. Death results from a combination of shock, respiratory failure, and central nervous system changes. Most deaths from insect stings occur within 15 to 30 min.

Early application of a tourniquet may prevent rapid spread of the venom. Bitten persons should be taught to remove the venom sac if present, being careful not to squeeze the sac. It may be necessary for some patients to carry an emergency kit, which is commercially available. Patients should be taught to give themselves an epinephrine injection. Patients having severe reactions should first receive 0.3 to 0.5 mL of a 1:1000 solution of epinephrine intravenously.

Stingrays

Approximately 750 persons each year are stung by stingrays. As the spine, which is curved and has serrated edges, enters the flesh, the sheath surrounding the spine ruptures, and venom is released. As the spine is withdrawn, fragments of the sheath may remain in the wound. The wound edges are often jagged and bleed freely. Pain usually is immediate and severe, increasing to maximum intensity in 1 to 2 h and lasting for 12 to 48 h. Treatment consists of copious irrigation with water to wash out any toxin and fragments of the spine's integumentary sheath. Venom is inactivated when exposed to heat. The area of the bite should be placed in water as hot as the patient can stand without injury for 30 min to 1 h. After soaking, the wound may be further debrided and treated appropriately. Patients treated in this manner have rapid and uncomplicated healing of the wound. Patients not treated with heat have tissue necrosis with prolonged drainage and chronically infected wounds.

Portuguese Man-of-War

After a severe sting by a Portuguese man-of-war there may be almost immediate severe nausea, gastric cramping, and constriction and tightness of throat and chest with severe

muscle spasm. There is intense, burning pain with weakness and perhaps respiratory distress. The most important emergency treatment is to inactivate the nematocysts immediately to prevent their continuous firing of toxins. This is accomplished by applications of a substance of high alcohol content, such as rubbing alcohol, followed by application of a drying agent, such as flour, baking soda, talc, or shaving cream. The tentacles may then be removed by shaving. An alkaline agent such as baking soda is then applied in order to neutralize the toxins, which are acidic. Demerol and Benadryl may dramatically relieve the pain and symptoms. Aerosol corticosteroid- analgesic balm is helpful.

Spiders Black Widow Spider

The most common biting spider in the United States is the black widow (*Latrodectus mactans*) (Fig. 6-67). The female spider has a reddish orange hourglass-shaped marking on its ventral surface. *Latrodectus mactans* venom is primarily neurotoxic in action and centers on the spinal cord. After a bite by the black widow spider, the majority of patients experience pain within 30 min, and a small wheal with an area of erythema appears. Nausea and vomiting occur in approximately one-third of patients, headache in one-fourth, and dyspnea may develop. The time of onset of symptoms after the bite is 30 min to 6 h. The severe symptoms last from 24 to 48 h. Generalized muscle spasm is the most prominent physical finding. Cramping muscle spasms occur in the thighs, lumbar region, abdomen, or thorax. Priapism and ejaculation have been reported. Most patients recover within 24 h.

Treatment consists of narcotics for the relief of pain and a muscle relaxant for relief of spasm. Methocarbamol (Robaxin) or 10 mL of a 10% solution of calcium gluconate relieves the symptoms. It is believed that calcium acts by depressing the threshold for depolarization at the neuromuscular junctions. Calcium gluconate may give instant relief of muscular pain, and methocarbamol can be administered intravenously 10 mL over a 5-min period, with a second ampule started in a saline solution drip. Although *Latrodectus mactans* antivenin is available, it is rarely required. The manufacturer recommends its use for patients with underlying cardiovascular disease. The antivenin is prepared from horse serum and is administered intramuscularly after appropriate skin tests. Hospitalization may be required for the young, the elderly, patients with significant chronic diseases, or those with severe signs and symptoms of envenomation.

Brown Recluse Spider

The distinguishing mark of the *Loxosceles reclusa* is the darker violin- shaped band over the dorsal cephalothorax (Fig. 6-68). The spider is native to the south central United States. The body ranges from 7 mm to 12 mm; including the legs, the spider's size ranges from 2 to 3 cm.

The initial bite may go unnoticed or be accompanied by a mild stinging sensation. Pain may recur 6 to 8 h afterward. A mild envenomation is associated with local urticaria and erythema that usually resolve spontaneously. More severe bites result in progression to necrosis and sloughing of skin with residual ulcer formation. A generalized macular and erythematous rash may appear in 12 to 24 h. Erythema develops, with bleb or blister formation surrounded by an irregular area of ischemia. A zone of hemorrhage and induration and a surrounding halo of erythema may develop peripherally. The central ischemia turns dark, and eschar forms by day 7; by day 14 the area sloughs, leaving an open ulcer. Approximately 3 weeks is required for the lesion to heal. The pain may be out of proportion with the size of the area involved. The progression from blue to black gives the bite a necrotic appearance, and the more severe bites develop within a few hours to 2 days. Systemically, the patient may have fever, nausea, vomiting, weakness, arthralgia, malaise, and even petechiae. The two principal systemic effects, hemolysis and thrombocytopenia, have been responsible for deaths. Hemoglobinemia, hemoglobinuria, leukocytosis, and proteinuria also may occur, and there may be eventual renal failure. *Loxosceles reclusa* venom is chiefly cytotoxic in action. Laboratory studies are obtained in patients with severe envenomation, including prothrombin time, partial thromboplastin time, platelet count, and urinalysis. In the pathophysiology of this spider bite, intravascular coagulations and the formation of microthrombi occur within the capillary, leading to capillary occlusion, hemorrhage, and necrosis.

Treatment is conservative because of the difficulty in predicting the severity of the bite. Various treatments have been advocated in addition to early excision, including treatment with corticosteroids, heparin, phentolamine, dextran, and infusion, but clinical studies have failed to identify the benefit of these agents. The dose for steroids has varied from 30 to 80 mg methylprednisolone daily tapered over a period of several days. A leukocyte inhibitor, dapsone (used in leprosy) is effective in reducing inflammation at the site of the brown recluse venom injection. Treatment with dapsone is 100 mg daily for 14 days before surgical excision, if required. The incidence of scarring and deformity was found to be much less in a dapsone-treated group than in a group treated with observation and subsequent surgical excision. There are significant side effects associated with dapsone treatment, including dose-dependent hemolytic anemia, methemoglobinemia, and rash. Conservative therapy usually is the preferred treatment. Excision of the necrotic area with skin grafting may be required at a later date.

Scorpions

Of the numerous species of scorpions in the United States, only one, *Centruroides exilicauda* or the bark scorpion, is medically significant. It is found primarily in the desert Southwest. Ranging in length from 1 to 7 cm, it is usually yellowish brown in color and may have vertical bands on its dorsum. A tubercle at the base of the stinger distinguishes the bark scorpion from other species. The venom is neurotoxic and causes the release of neurotransmitters from the autonomic nervous system and the adrenal glands. It also causes the depolarization of neuromuscular junctions.

The sting causes intense pain with few other local symptoms. Hyperesthesia persists at the site so that a light tap will reproduce the intense pain. The tap test reinforces the diagnosis. In addition to pain, other symptoms reflect the neurotoxic nature of the venom, including anxiety, blurred vision or temporary blindness, wandering eye movements, dyspnea, wheezing, dysphagia, involuntary urination and defecation, and opisthotonos. Somatic muscular contractions resembling seizures, hypertension, supraventricular tachyarrhythmias, and fever also are seen.

These stings have been of little significance in adults and are satisfactorily treated with cold compresses. In contrast, infants and small children have died from scorpion envenomation, though not since 1968. Small children with signs of envenomation should be admitted to the hospital and monitored. No special diagnostic tests are indicated. Treatment consists of airway management for excessive secretions, sedation, and treatment of arrhythmias and hypertension if indicated. Calcium gluconate has been used to treat muscle spasms. Narcotics should not be used because they aggravate the neurotoxic effects of the venom. A goat-derived antivenin is available but only in the state of Arizona.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 7 - Burns

Glenn D. Warden
David M. Heimbach

INTRODUCTION

Thermal burns and related injury are a major cause of death and disability in the United States. The introduction of burn centers in 1945 heralded a rapid improvement in survival and reduction of morbidity of burn patients and provided the basis for regional specialty treatment centers in other disciplines. The interactive multidisciplinary team has proved to be the least expensive and most efficient method of treating major burn injury, a long- term disease of which the initial acute care is only a small part of the total treatment. Burn patients often require years of supervised rehabilitation, reconstruction, and psychosocial support. Omission of any step in the treatment regimen by any of the burn team members, including the burn surgeon, the nursing staff, and rehabilitation, nutrition, and psychosocial support staff, can result in less than optimal outcome.

EPIDEMIOLOGY

In the United States approximately 2 million individuals annually are burned seriously enough to seek health care; about 70,000 of these require hospitalization, and about 5000 die. More than 90 percent of burns are caused by carelessness or ignorance and are completely preventable; nearly half are smoking- or alcohol-related. While prevention of burns is still the long-term solution to burn care, advances in the care of burned patients during the past 20 years are among the most dramatic in medicine.

The annual federal expenditure for research in cancer, heart disease, and stroke exceeds by 15 times that for trauma and burns, despite the fact that trauma and burns account for a loss of productive person years from injury greater than cancer, heart disease, and stroke combined. The number of burn deaths in the United States has decreased from 15,000 in 1970 to 5000 in 1996. Over the same period the size of burn associated with a 50 percent survival rate has increased from 30 percent of the *total body surface area (TBSA)* to over 80 percent TBSA in otherwise healthy young adults. Hospital stay has been cut in half. Ninety-six percent of patients admitted to burn centers survive, and eighty percent of them return to their preburn physical and social situation within a year of the injury.

The quality of burn care is no longer measured only by survival, but also by long-term function and appearance. Although small burns are not usually life-threatening, they need the same attention as larger burns to achieve functional and cosmetic outcome; in the largest burn centers the average burn size is less than 15 percent TBSA. The physician's goal for any burn is well-healed, durable skin with normal function and near-normal appearance. Scarring, a virtual certainty with deep burns, can be minimized by appropriate early surgical intervention and long-term scar management. These goals require individualized patient care plans, based on burn characteristics and host factors.

As with other forms of trauma, burns frequently affect children and young adults. The hospital expenses and the social costs related to time away from work or school are staggering. Most burns are limited in extent, but a significant burn of the hand or foot may keep manual workers away from work for a year or more, or permanently. The eventual outcome for the burned patient is related to the severity of the injury, the individual physical characteristics of the patient, the motivation of the patient, and the quality of the treatment.

Etiology

Cutaneous burns are caused by the application of heat, cold, or caustic chemicals to the skin. When heat is applied to the skin, the depth of injury is proportionate to the temperature applied, the duration of contact, and the thickness of the skin.

Scald Burns

In civilian practice, scalds, usually from hot water, are the most common cause of burns. Water at 140°F (60°C) creates a deep dermal or full- thickness burn in 3 s. At 156°F (69°C) the same burn occurs in 1 s. Freshly brewed coffee from an automatic percolator generally is about 180°F (82°C); boiling water always causes deep burns, and thick soups and sauces, which remain in contact longer with the skin, also invariably cause deep burns. Exposed areas tend to be burned less deeply than areas covered with thin clothing. Clothing retains the heat and keeps the liquid in contact with the skin for a longer period. Immersion scalds are always deep, severe burns. The liquid causing an immersion scald may not be as hot as with a spill scald, but the duration of contact is longer and these burns frequently occur in small children or elderly patients with thin skin.

Deliberate scalds are the commonest form of reported child abuse and are responsible for about 5 percent of admissions of children to burn centers. The physician should note any discrepancy between the history provided by the care giver and the distribution and probable cause of the burn. A suspicious burn must be reported promptly.

Scald burns from grease or hot oil are usually deep dermal or full-thickness burns. Cooking oil and grease may be in the range of 400°F. Tar and asphalt burns are a special kind of scald. The “mother pot” at the back of a roofing truck maintains tar at a temperature of 400 to 500°F (204 to 260°C). Burns caused by tar directly from the “mother pot” are invariably full-thickness burns. By the time the tar has been spread on the roof or street, its temperature has been lowered to the point where most burns caused by it are deep dermal in nature. The tar should be removed by application of a petroleum-based ointment under a dressing. The dressing may be removed and the ointment reapplied every 2 to 4 h until the tar has dissolved. Only then can the extent of the injury and the depth of the burn be estimated accurately.

Flame Burns

Flame burns are the next most common. Although the incidence of injuries caused by house fires has decreased with the use of smoke detectors, smoking-related fires, improper use of flammable liquids, automobile accidents, and ignition of clothing from stoves or space heaters still exact their toll. Patients whose bedding or clothes have been on fire rarely escape without some full-thickness burns.

Flash Burns

Flash burns are next in frequency. Explosions of natural gas, propane, gasoline and other flammable liquids cause intense heat for a very brief time. Clothing, unless it ignites, is protective against flash burns. Flash burns generally have a distribution over all exposed skin, with the deepest areas facing the source of ignition. Flash burns are mostly dermal, their depth depending on the amount and kind of fuel that explodes. These burns generally heal without requiring extensive skin grafting, but they may cover a large skin area and be associated with significant thermal damage to the upper airway.

Contact Burns

These burns result from contact with hot metals, plastic, glass, or hot coals; they are usually limited in extent, but invariably very deep. It is common for patients involved in industrial accidents to have associated crush injuries because these accidents are commonly caused by contact with presses or other hot, heavy objects. Automobile accidents may leave victims in contact with hot engine parts. The exhaust pipes of motorcycles cause a characteristic burn of the medial leg that, although small, usually requires excision and grafting. Toddlers who touch or fall against irons, ovens, and wood-burning stoves with outstretched hands are likely to suffer deep burns of the palms. Contact burns are often fourth-degree burns, especially those in unconscious or postictal patients, and those caused by molten materials.

Burn Prevention

The number of burns occurring annually in the United States remains unknown despite an extensive government-funded study; however, a number of states have passed legislation making a burn a reportable disease, so more accurate information may be forthcoming. One-third of the victims are children under the age of 15. In children under 8 years of age, the most common burns are scalds, usually from the spilling of hot liquids; in older children and adults, the most common burns are flame- related, usually the result of house fires or the ill-advised use of flammable liquids for burning brush or trash, lighting barbecues, etc., or smoking- or alcohol-related. Industrial accidents are most often caused by chemicals or hot liquids, followed by electricity and then molten or hot metal.

More than 90 percent of all burns are preventable by using common sense and taking ordinary precautions. Over the past 20 years several critical legislative actions, such as that mandating flame-resistant sleepwear for children, have decreased burns and burn mortality. Smoke detectors are required in all rental units and new construction, and are probably the most significant factor in the major decrease in burn mortality during the past decade. Many states have initiated legislation mandating that the maximum temperature for home and public hot water heaters be set to below 140°F (60°C). Individual burn centers, the *American Burn Association (ABA)*, and the International Society for Burn Injury have all produced multiple television public service announcements regarding hot water, carburetor flashes, barbecue burns, scalds, etc. Numerous programs are directed to schools; for example, the “Stop Drop and Roll” sequence is known by most school children. “Change your clock, change your smoke detector battery” is a national program reminding everyone to keep their detector batteries fresh.

HOSPITAL ADMISSION AND BURN CENTER REFERRAL

The need for hospital admission and specialized care is dictated by the severity of symptoms from smoke inhalation and the magnitude of associated burns. Any patient who is symptomatic with smoke inhalation and has more than trivial burns should be admitted to a hospital. If the burns cover more than 15 percent TBSA, the patient should be referred to a special care unit. In the absence of burns, admission depends on the severity of symptoms, the presence of preexisting medical problems, and the social circumstances of the patient. Otherwise healthy patients with mild symptoms (only a few expiratory wheezes, minimal sputum production, CO level <10, and normal blood gases) who have a place to go and someone to stay with them can be observed for an hour or two and then discharged. Patients with preexisting cardiovascular or pulmonary disease who have any symptom related to smoke inhalation should be admitted for observation. Patients with moderate symptoms (generalized wheezing, mild hoarseness, moderate sputum, CO levels 5 to 10, and normal blood gases) are admitted to a medical-surgical unit for close observation and treatment. Patients with severe symptoms (air hunger, severe wheezing, copious [usually carbonaceous] sputum) should be admitted to an intensive care unit or, preferably, a burn unit.

Burn Severity and Classification

The severity of injury caused by burns is proportionate to the size of the total burn, the depth of the burn, the age of the patient, and associated medical problems or injuries. Burns have been classified by the American Burn Association and the American College of Surgeons Committee on Trauma as minor, moderate, and severe. Minor burns are superficial burns of less than 15 percent TBSA. *Moderate* burns are defined as superficial burns of 15 to 25 percent TBSA in adults or 10 to 20 percent in children; full-thickness burns of less than 10 percent TBSA, and burns not involving the eyes, ears, face, hands, feet, or perineum. Because of the significant cosmetic and functional risks associated with burns, all but very superficial burns of the face, hands, feet, and perineum should be treated by a physician with an interest in burn care in a facility that is accustomed to dealing with burns. Major burns, as described above, and most full- thickness burns in infants and elderly patients, or patients with associated diseases or injuries, should also be cared for in a specialized facility. Moderate burns can be cared for in a community hospital by a knowledgeable physician, as long as the other members of the health care team have the resources and knowledge to ensure a good result. Newer techniques of early wound closure have made burn care more complex, and an increasing number of patients with small but significant burns are being referred to specialized care facilities.

The criteria for hospital admission of patients with minor or moderate burns vary according to physician preference, the patient's social circumstances, and the ability to provide close follow-up. In some circumstances superficial burns as large as 15 percent can be successfully managed on an outpatient basis. In other circumstances, burns as small as 1 percent may require admission because of the patient's inability or unwillingness to care for the wound. The physician should have low- threshold criteria for admission of elderly patients and infants. Any patient (child or adult) with suspicion of abuse must be admitted.

Burn Center Referral Criteria

The ABA has identified the following injuries as those requiring referral to a burn center after initial assessment and stabilization at an emergency department.

- (1) Second- and third-degree burns greater than 10 percent TBSA in patients under 10 or over 50 years of age.
- (2) Second- and third-degree burns greater than 20 percent TBSA in other age groups.
- (3) Second- and third-degree burns involving the face, hands, feet, genitalia, perineum, and major joints.
- (4) Third-degree burns greater than 5 percent TBSA in any age group.
- (5) Electrical burns, including lightning injury.
- (6) Chemical burns.
- (7) Inhalation injury.
- (8) Burn injury in patients with preexisting medical disorders that could complicate management, prolong the recovery period, or affect mortality.
- (9) Any burn patients with concomitant trauma (e.g., fractures) where the burn injury poses the greatest risk of morbidity or mortality. If the trauma poses the greater immediate risk, the patient may be treated initially in a trauma center until stable before being transferred to a burn center. The physician's decisions should be made with the regional medical control plan and triage protocols in mind.
- (10) Burn injury in children admitted to a hospital without qualified personnel or equipment for pediatric care.
- (11) Burn injury in patients requiring special social, emotional, and/or long-term rehabilitative support, including cases involving suspected child abuse, substance abuse, etc.

Transport and Transfer Protocols

Once an airway is established and resuscitation under way, burned patients are eminently suitable for transport. Resuscitation can continue en route because the patient will usually remain stable for several days. This was well demonstrated during the Viet Nam war: about 1000 burn victims were first transported from Viet Nam to Japan, and then from Japan to the military Burn Center in San Antonio, Texas; the transport was usually accomplished during the first 2 weeks postburn with very few complications.

Hospitals without specialized burn care facilities should have transfer agreements and treatment protocols with a burn center well in advance of the need for transfer. Definitive care begins at the initial hospital and continues without interruption during transport and at the burn center. Transfer should be from physician to physician, and contact should be established as soon as the patient arrives in the emergency room of the initial hospital.

The mode of transport depends on vehicle availability, local terrain, weather, and the distances involved. For distances of less than 50 miles (80 km), ground ambulance is usually satisfactory. Helicopter transport is often preferred when the distance is between 50 and 150 mi (80 and 240 km); however, patient monitoring, airway management, and effecting changes in therapy are very difficult to achieve in a helicopter. In all cases in which the time of transport will be long, it is the responsibility of the referring physician to ensure that the patient's condition will permit the transport. All patients transported by air should have a nasogastric tube inserted and placed on dependent drainage, because nausea and vomiting inevitably result during the flight. Two large-bore intravenous lines are mandatory.

For distances over 150 mi (240 km), fixed-wing aircraft are most satisfactory. Modern air ambulances are completely equipped intensive care units, and the personnel are usually well trained for critical care.

The airway must be secured. At 30,000 ft the planes are pressurized to an altitude of about 5500 ft. Supplemental oxygen can be given in flight, but if the patient's oxygenation is marginal, it may be best to intubate and place the patient on a ventilator for the transport. Intubation is difficult en route; if there is question of upper airway edema, the patient should be intubated prior to transport. Burned patients have difficulty maintaining body temperature, and they should be warmly wrapped prior to transport. Bulky dressings, a blanket, and a mylar sheet (usually available from the flight team) can help maintain body temperature. If the patient has any cardiac irregularities, the plane must be equipped with monitoring capability, because noise and vibrations in-flight make clinical monitoring difficult. If there is danger of compromised circulation due to circumferential full- thickness burns, escharotomies should be done at the referring hospital, unless the total hospital-to-hospital time will be less than 2 h.

Burn Center Verification and a National Burn Registry

In 1995, in conjunction with the American College of Surgeons Committee on Trauma (ACS-COT), the ABA initiated a program of Burn Center Verification. A resource document outlines the resources and process necessary to provide optimal care of the burn patient. Burn centers may be reviewed to verify that they provide state-of-the-art care for burn patients, a process involving a lengthy questionnaire, a site visit, a written report, and approval by the joint verification committees. By 1996, over 25 centers had gone through this process. A national burn registry is kept by U.S. burn centers to provide national statistics regarding incidence, epidemiology, and outcome of burn cases.

EMERGENCY CARE

Care at the Scene

Airway

Once flames are extinguished, initial attention must be directed to the airway. Immediate cardiopulmonary resuscitation is rarely necessary, except in electrical injuries that have induced cardiac arrest or in patients with severe carbon monoxide poisoning with hypoxic cardiac arrest. Any patient rescued from a burning building or exposed to a smoky fire should be placed on 100% oxygen by tight-fitting mask if there is any suspicion of smoke inhalation. If the patient is unconscious, and appropriately trained personnel are present, an endotracheal tube should be placed and attached to a source of 100% oxygen. If the airway has to be supported by a tight mask, there is a significant danger of aspiration of gastric contents, because air forced into the stomach will distend it and cause vomiting. The mask prevents expulsion of the fluid, and gastric contents can flood the tracheobronchial tree.

Other Injuries and Transport

Once an airway is secured, the patient is assessed for other injuries and then transported to the nearest hospital. If a burn center is within a 30- min drive and the burn is severe, the patient may be taken directly to that facility. Patients should be kept flat and warm and be given nothing by mouth. If appropriately trained, the emergency medical technicians should place an intravenous line and begin fluid administration of crystalloid solution at a rate of approximately 1 L/h. For transport, the patient should be wrapped in a clean sheet and blanket. Sterility is not required. Before or during transport, constricting clothing and jewelry should be removed from burned parts, because local swelling begins almost immediately. Constricting objects increase swelling, and removing tight jewelry in the presence of distal edema is time-consuming and difficult.

Cold Application

Smaller burns, particularly scalds, are treated with immediate application of cool water. It has been mathematically demonstrated that cooling cannot reduce skin temperature enough to prevent further tissue damage, and that histologic damage is similar with or without cooling, but there is evidence in animals that cooling delays edema formation, probably by reducing initial thromboxane production. After several minutes have elapsed, further cooling does not alter the pathologic process. Iced water should never be used even on the smallest of burns. If ice is used on larger burns, systemic hypothermia may follow, and the associated cutaneous vasoconstriction can extend the thermal damage.

Emergency Room Care

The primary rule for the emergency physician is, "Forget about the burn." As with any form of trauma the ABC protocol—airway, breathing, circulation—must be followed. Although a burn is a dramatic injury, a careful search for other life-threatening injuries is the first priority. Only after making an overall assessment of the patient's condition should attention be directed to the burns. The assessment of patients who have not been thermally injured is discussed in Chap. 6; the following sections apply specifically to problems encountered in the burn patient.

Emergency Assessment of Inhalation Injury

The history is important. Inhalation injury should be suspected in anyone with a flame burn, and assumed, until proved otherwise, in anyone burned in an enclosed space. The acrid smell of smoke on a victim's clothes should raise suspicion. The rescuers are the most important historians and should be questioned carefully before they leave the emergency facility.

Careful inspection of the mouth and pharynx should be done early. Hoarseness and expiratory wheezes are signs of potentially serious airway edema or smoke poisoning. Copious mucus production and carbonaceous sputum are sure signs, but their absence does not rule airway injury out. Carboxyhemoglobin levels should be obtained; elevated carboxyhemoglobin levels or any symptoms of carbon monoxide poisoning are presumptive evidence of associated smoke poisoning.

A decreased P/F ratio, the ratio of arterial P_{o2} to percentage of inspired oxygen (Fi_{o2}) in arterial blood gases is one of the earliest indicators of smoke inhalation. A ratio of 400 to 500 is normal; patients with impending pulmonary problems have a ratio of less than 300 (e.g., a Pa_{o2} of less than 120 with an Fi_{o2} of 0.40). A ratio of less than 250 is an indication

for aggressive pulmonary support rather than for increasing the inspired oxygen concentration.

Fiberoptic bronchoscopy is inexpensive, is quickly performed in experienced hands, and is very useful for accurately assessing edema of the upper airway. Although bronchoscopy documents tracheal erythema, it does not materially influence the treatment of pulmonary injury.

Fluid Resuscitation in the Emergency Room

As burns approach 20 percent TBSA, local inflammatory cytokines escape into the circulation and result in a systemic inflammatory response. The capillary leak, permitting loss of fluid and protein from the intravascular compartment into the extravascular compartment, becomes generalized. Cardiac output decreases as a result of marked increased peripheral resistance, hypovolemia secondary to the capillary leak, and the accompanying increase in blood viscosity. The resulting intense sympathetic response leads to decreased perfusion to the skin and viscera. Decreased flow to the skin may convert the zone of stasis to one of coagulation, thereby increasing the depth of burn. Decreased cardiac output may depress CNS function, and, in extreme cases, lead to severe cardiac depression with eventual cardiac failure in healthy patients, or to myocardial infarction in patients with preexisting coronary artery atherosclerosis. Impairment in CNS function manifests as restlessness, followed by lethargy, and finally by coma. If resuscitation is inadequate, burns of 30 percent TBSA frequently leads to acute renal failure, which, in the case of severe burn, almost invariably results in a fatal outcome.

Resuscitation begins by starting intravenous Ringer's lactate solution at a rate of 1000 mL/h in adults and 20 mL/kg in children. Burn patients requiring intravenous resuscitation (generally those with burns over 20 percent TBSA) should have a Foley catheter placed and urine output monitored hourly.

Patients with burns less than 50 percent TBSA can usually be resuscitated with a single large-bore peripheral intravenous line. Because of the high incidence of septic thrombophlebitis lower extremities should not be used as portals for intravenous lines. Upper extremities are preferable, even if the intravenous line must pass through burned skin. Patients with burns over 50 percent TBSA, who have associated medical problems, who are very young or very old, or who have concomitant smoke inhalation should have additional central venous pressure monitoring. Because of the hemodynamic instability in patients with burns over 65 percent TBSA, these patients should be transferred as quickly as possible so they can be monitored in an intensive care setting where Swan-Ganz catheters for measuring pulmonary wedge pressures and cardiac outputs can be placed.

Tetanus

Burns are tetanus-prone wounds. The need for tetanus prophylaxis is determined by the patient's current immunization status. Previous immunization within 5 years requires no treatment, immunization within 10 years a tetanus toxoid booster, and unknown immunization status hyperimmune serum (Hyper-Tet).

Gastric Decompression

Many burn centers begin tube feeding on admission, to protect the stomach from stress ulceration and the patient from a paralytic ileus and decreasing catabolism. If the patient is to be transported, the safest course is usually to decompress the stomach with a nasogastric tube.

Pain Control

During the shock phase of burn care, medications should be given intravenously. Subcutaneous and intramuscular injections are absorbed variably depending on perfusion and should be avoided. Pain control is best managed with small intravenous doses of morphine, usually 2 to 5 mg, given until pain control is adequate, without affecting blood pressure.

Psychosocial Care

Psychosocial care should begin immediately. The patient and family must be comforted and given a realistic assessment regarding the prognosis of the burns. In house fires, patients' loved ones, pets, and possessions may have been destroyed. If the family is not available, some member of the team, usually the social worker, should determine the extent of the damage. If the patient is a child, and if the circumstances of the burn are suspicious, physicians in all states are required by law to report any suspected case of child abuse to local authorities.

Care of the Burn Wound

After all other assessments are complete, attention should be directed to the burn. If the patient is to be transferred during the first postburn day, which is almost always the case, the burn wounds can be left alone. However, the size of the burn should be calculated to establish the proper level of fluid resuscitation, and pulses distal to circumferential full-thickness burns should be monitored. The patient can be wrapped in a clean sheet and kept warm until arriving at the definitive care center.

Escharotomy and Fasciotomy

Chest Escharotomy

The adequacy of respiration must be monitored continuously throughout the resuscitation period. Early respiratory distress may be due to compromise of the ventilatory function caused by a cuirass effect related to a deep circumferential burn wound of the chest. Pressures required by ventilation increase and arterial P_{co2} rises. Inhalation injury, pneumothorax, or other causes may also result in respiratory distress.

When escharotomy is required in a patient with a circumferential chest wall burn, it is performed in the anterior axillary line bilaterally. If there is significant extension of the burn onto the adjacent abdominal wall, the escharotomy incisions should be extended to this area and should be connected by a transverse incision along the costal margin (Fig. 7-1).

Escharotomy of Extremities

Edema formation in the tissues under the tight unyielding eschar of a circumferential burn of an extremity may produce significant vascular compromise that, if unrecognized and untreated, will lead to permanent, serious neurologic and vascular deficits. All rings, watches, and other jewelry must be removed from injured limbs to avoid distal vascular ischemia. Skin color, sensation, capillary refill, and peripheral pulses must be assessed hourly in any extremity with a circumferential burn. The occurrence of any of the following signs or symptoms may indicate poor perfusion of the distal extremity: cyanosis, deep tissue pain, progressive paresthesia (loss of sensation), progressive decrease or absence of pulse, or sensation of cold extremities. An ultrasonic flowmeter is a reliable means for assessing arterial blood flow, and the need for an escharotomy, and can also be used to assess adequacy of circulation after an escharotomy.

Direct monitoring of intramuscular compartment pressure provides objective evidence for adequacy of circulation. It is useful for assessing not only the need for, but also the adequacy of, escharotomy and/or fasciotomy. An 18-gauge needle is attached to the transducer tubing that is used for monitoring arterial pressure and inserted into the compartment; pressure is measured on the monitor. In burn patients the threshold pressure for performing escharotomy or fasciotomy is 30 mmHg. While elevation and manipulation of the extremity may relieve minor deviations of pressure, escharotomy is necessary for 30 mmHg or higher.

Both escharotomies and fasciotomies may be done as bedside procedures with a sterile field and scalpel. Local anesthesia is unnecessary because third-degree eschar is insensate; small doses of intravenous narcotics may be utilized to control anxiety. The incision, which must avoid major nerves, vessels, and all tendons, should be placed along the mid-medial or mid-lateral aspect of the extremity and, to permit adequate separation of the cut edges for decompression, should extend through the eschar down to the subcutaneous fat. The incision should extend through the length of the constricting third-degree burn, and should be carried across involved joints (see Fig. 7-1).

When a single escharotomy incision in an extremity does not result in adequate distal perfusion, a second escharotomy incision on the contralateral aspect of the extremity should be performed. A finger escharotomy is seldom required, and should be performed only after consultation with the receiving burn center physician.

Escharotomy and/or fasciotomy is rarely required within the first 6 h postburn and should only be done under the supervision of a trained surgeon. Because burn patients are at risk for developing a compartment syndrome up to 72 h postinjury, any involved extremity should be reassessed continually for signs of the dangerous elevation in compartment pressures that can occur after initial decompression.

BURN SEVERITY

The severity of any burn injury is related to the size and depth of the burn and to the part of the body that has been burned. Burns are the only quantifiable form of trauma. The single most important factor in predicting burn-related mortality, need for specialized care, and the type and likelihood of complications is the overall size of the burn as a proportion of the patient's total body surface. Treatment plans, including initial resuscitation and subsequent nutritional requirements, are directly related to the size of burn.

Burn Size

A general idea of burn size can be made by using the *Rule of Nines*. Each upper extremity accounts for 9 percent of TBSA, each lower extremity accounts for 18 percent, the anterior and posterior trunk each account for 18 percent, the head and neck account for 9 percent, and the perineum accounts for 1 percent. Although the Rule of Nines is reasonably accurate for adults, a number of more precise charts have been developed. Most emergency rooms have a chart available comparable to the one shown in Fig. 7-2. A diagram of the burn can be drawn on the chart, and, theoretically, a precise calculation of burn size can be made from the accompanying TBSA estimates given.

Children under 4 years old have much larger heads and smaller thighs in proportion to body size than do adults. In infants the head accounts for nearly 20 percent of the TBSA; infant body proportions do not fully reach adult percentages until adolescence. Even when using precise diagrams, individual observer variation may differ by as much as ±20 percent. An observer's experience with burned patients rather than his or her educational level appears to be the best predictor of accuracy of estimation. To increase accuracy in burn size estimation, especially when burns are in scattered body areas, the observer should calculate the unburned areas on a separate diagram. If the calculations of the unburned areas and the burned areas do not add up to 100 percent, the observer should begin again with a new diagram.

For smaller burns an accurate assessment of burn size can be made by using the patient's hand, which amounts to 2.5 percent TBSA (see Fig. 7-2). The dorsal surface, including the fingers, accounts for 1 percent, the palmar surface, including the fingers, for 1 percent, and the vertical surface for 0.5 percent.

Burn Depth

Along with burn extent and patient age, depth of burn is a primary determinant of mortality. Burn depth is also the primary determinant of the patient's long-term appearance and function.

Burns not extending all the way through the dermis leave behind epithelium-lined skin appendages: sweat glands, and hair follicles with attached sebaceous glands. When dead dermal tissue is removed, epithelial cells swarm from the surface of each appendage to meet swarming cells from neighboring appendages, forming a new, fragile epidermis on top

of a thinned and scarred dermal bed. Skin appendages vary in depth, and the deeper the burn the fewer the appendages that contribute to healing and the longer the burn takes to heal. The longer the burn takes to heal, the less dermis remains, the greater the inflammatory response, and the more severe the scarring.

When nonoperative treatment is the routine, as it is in many developing countries, an accurate assessment of burn depth is of little importance except for predicting mortality. On the other hand, with aggressive surgical treatment, an accurate estimation of burn depth is crucial. Burns that heal within 3 weeks usually do so without hypertrophic scarring or functional impairment, although long-term pigmentary changes are common. Burns that take longer than 3 weeks to heal often produce unsightly hypertrophic scars; frequently lead to functional impairment; and provide only a thin, fragile epithelial cover for many weeks or months. State-of-the-art burn care, at least in patients with small and moderate burns, involves early excision and grafting of all burns that will not heal within 3 weeks. The challenge of course is to determine which burns *will* heal within 3 weeks and are thus best treated by daily local care. Other burns should be treated surgically.

An understanding of burn depth requires an understanding of skin thickness. The living epidermis is an intensely active layer of epithelial cells under layers of dead keratinized cells and is superficial to the active structural framework of the skin, the dermis. The thickness of skin varies with the age and sex of the individual and the area of the body. The thickness of the living epidermis is relatively constant, but keratinized (dead and cornified) epidermal cells may reach a height of 0.5 cm on the palms of hands and the soles of feet. The thickness of the dermis varies from less than 1 mm on eyelids and genitalia to more than 5 mm on the posterior trunk. The proportional thickness of skin in each body area in children is similar to that in adults, but infant skin thickness in each specific area may be less than half that of adult skin; the skin does not reach adult thickness until puberty. Similarly, in patients over 50 years of age dermal atrophy begins; all areas of skin become thin in elderly patients.

The depth of burn is dependent on the heat of the burn source, the thickness of the skin, the duration of contact, and the heat-dissipating capability of the skin (blood flow). A scald in an infant or elderly patient will be deeper than an identical scald in a young adult. A diabetic with impaired sensation, or an inebriated patient with an impaired sensorium, who lies on a heating pad (even of temperature less than 106°F (41°C) all night may sustain full-thickness burns because of the long duration of contact with the pad and the pressure of the body weight which occludes cutaneous blood flow and prevents heat dissipation.

Burns are classified according to increasing depth as first-degree, second- degree (superficial dermal and deep dermal), third-degree (full-thickness), and fourth-degree. Because most deep burns are removed surgically and grafted, such a precise characterization is not necessary for non-life- threatening burns. A more pertinent classification might be “shallow burns” and “deep burns.” Nevertheless, distinguishing between deep burns that are best treated by early excision and grafting and shallow burns that heal spontaneously is not always straightforward, and many burns have a mixture of clinical characteristics, making precise classification difficult.

Shallow Burns

First-Degree Burns

First-degree burns involve only the epidermis. First-degree burns do not blister, but become erythematous because of dermal vasodilation, and are quite painful. Over 2 to 3 days the erythema and the pain subside. By about day 4, the injured epithelium desquamates in the phenomenon of “peeling,” which is well known after sunburn.

Superficial Dermal Burns (Second-Degree)

Superficial dermal burns include the upper layers of dermis, and characteristically form blisters with fluid collection at the interface of the epidermis and dermis. Blistering may not occur for some hours after injury, and burns originally appearing to be first-degree may be diagnosed as superficial dermal burns after 12 to 24 h. When blisters are removed, the wound is pink and wet and currents of air passing over it cause pain. The wound is hypersensitive, and the burns blanch with pressure. If infection is prevented, superficial dermal burns heal spontaneously in less than 3 weeks, and do so with no functional impairment. They rarely cause hypertrophic scarring, but in pigmented individuals the healed burn may never completely match the color of the surrounding normal skin.

Deep Burns

Deep Dermal Burns (Second-Degree)

Deep dermal burns extend into the reticular layers of the dermis. Deep dermal burns also blister, but the wound surface is usually a mottled pink and white color immediately after the injury, because of the varying blood supply to the dermis (white areas have little to no blood flow; pink areas have some blood flow). The patient complains of discomfort rather than pain. When pressure is applied to the burn, capillary refill occurs slowly or may be absent. The wound is often less sensitive to pinprick than the surrounding normal skin. By the second day, the wound may be white and is usually fairly dry. If infection is prevented, these burns will heal in 3 to 9 weeks, but invariably do so with considerable scar formation. Unless active physical therapy is continued throughout the healing process, joint function can be impaired, and hypertrophic scarring, particularly in pigmented individuals and children, is common.

Full-Thickness Burns (Third-Degree)

Full-thickness burns involve all layers of the dermis and can heal only by wound contracture, by epithelialization from the wound margin, or by skin grafting. Full-thickness burns appear white, cherry red, or black, and may or may not have deep blisters. Full-thickness burns are described as being leathery, firm, and depressed when compared with adjoining normal skin, and are insensitive to light touch or pinprick. The difference in depth between a deep dermal burn and a full-thickness burn may be less than a millimeter. The clinical appearance of full-thickness burns can resemble that of deep dermal burns. Like deep dermal burns they may be mottled in appearance. They rarely blanch on pressure, and they may have a dry, white appearance. In some cases the burn is translucent, with clotted vessels visible in the depths. Some full-thickness burns, particularly immersion scalds, have a red appearance, and may be confused with superficial dermal burns. They can be distinguished, however, because they do not blanch with pressure. Full-thickness burns develop a classic burn eschar: the structurally intact but dead and denatured dermis that, over days and weeks, separates from the underlying viable tissue.

Fourth-Degree Burns

Fourth-degree burns involve not only all layers of the skin, but also subcutaneous fat and deeper structures. These burns almost always have a charred appearance, and frequently only the cause of the burn gives a clue to the amount of underlying tissue destruction. Electrical burns, contact burns, some immersion burns, and burns sustained by patients who are unconscious at the time of burning may all be fourth-degree.

The Assessment of Burn Depth

The standard technique for determining burn depth has been clinical observation of the wound. The difference in depth between a burn that heals in 3 weeks, a deep dermal burn that heals only after many weeks, and a full-thickness burn that will not heal at all may be only a matter of a few tenths of a millimeter. A burn is a dynamic process for the first few days; a burn appearing shallow on day 1 may appear considerably deeper by day 3. Further, the kind of topical wound care used can dramatically change the appearance of the burn. For these reasons, and because of the increasing importance of an accurate assessment of burn depth for planning definitive care of burn wounds, there is considerable interest in technology that will help determine burn depth more precisely than clinical observation.

Evaluation by an experienced surgeon as to whether an intermediate-depth dermal burn will heal in 3 weeks is about 50 percent accurate. In experienced hands, however, early excision and grafting provides better results than nonoperative care for such indeterminate burns.

Other techniques involve (1) the ability to detect dead cells or denatured collagen (through biopsy, ultrasound, vital dyes), (2) assessment of changes in blood flow (through fluorescein, laser Doppler, and thermographic assays), (3) analysis of the color of the wound (through light reflectance methods), and (4) evaluation of physical changes, such as edema (through *nuclear magnetic resonance (NMR)* techniques.

Biopsy

Biopsy and histologic examination would seem the most accurate technique for determining burn depth, but biopsies are expensive, leave permanent scars in wounds that would not be excised, and require an experienced pathologist to distinguish live from denatured collagen and cells. In addition, wound depth changes during the first 48 h, and Jackson reported that 7 days were necessary to get reproducible results from burn biopsies. Whatever the timing of biopsy, it requires another day or two to get permanent sections, and there is no guarantee that specimens from areas adjacent to the biopsy are the same depth as the biopsy specimens. For these reasons biopsies are rarely used in clinical practice, and to the authors' knowledge no studies have demonstrated a correlation between biopsy findings and healing within 3 weeks.

Vital Dyes

Theoretically, a vital dye directly applied to the burn wound would be useful in detecting dead tissue and also in determining the needed depth of surgical excision. Davies, in 1980, described important characteristics of such a dye. It should stain only dead tissue, not be removable with wound treatment, be nontoxic, provide a sharp demarcation between living and dead tissue, penetrate all dead tissue, and be compatible with topical treatments usually used in burn care. In a rat model methylene blue, toluidine blue, trypan blue, Evans blue, and sulpham blue were evaluated. Methylene blue, which is metabolized to a colorless compound by living cells, was selected for preliminary testing in patients. When methylene blue was mixed with silver sulfadiazine and applied topically within 48 h, a significant blue discoloration appeared that remained even after vigorous washing. Excision was carried down to dermis that was unstained. The authors reported encouraging results in this preliminary study. However, casual use of this technique in our burn centers did not produce a satisfactorily sharp demarcation to guide excision.

Fluorescein Fluorometry

Fluorescein, injected systemically, is delivered through a patent circulation and fluoresces under ultraviolet light. It has been widely used to determine viability of flaps, intestine, and even whole extremities. The use of fluorescein fluorescence to determine burn depth was first reported in 1943, but this technique was unused until more precise instrumentation was developed—a fiberoptic perfusion fluorometer—which could measure the magnitude of fluorescence. Gatti studied 63 burns with the fluorometer after intravenous administration of sodium fluorescein. The fluorescein kinetics were monitored for 1 h within the first 48 h, and again between the third and sixth days postburn, and compared to the kinetics in adjacent normal skin. Depth of burn was confirmed by biopsy and healing characteristics. Fluorometric analysis during both study periods consistently distinguished between partial-thickness and full-thickness burns. Partial-thickness burns uniformly exhibited fluorescence within 10 min; full-thickness burns showed no fluorescence. However, when this technique was reported in 1982, most burn surgeons only excised full- thickness burns, leaving partial-thickness burns to heal on their own, and there was therefore an advantage to distinguishing between partial- and full-thickness burns.

Gatti's results were confirmed in 1984. However, a 1986 report by Black could not confirm significant differences between partial- and full-thickness burns using a similar technique because standard deviations in both categories were too large to distinguish between groups, let alone be predictive in any given burn. Our own experience with fluorescein indicates that it confirms the clinical diagnosis in very deep and very shallow burns, areas where there is little confusion, but that it cannot distinguish between intermediate and deep dermal burns.

Laser Doppler Flowmetry

Laser Doppler flowmetry has been used since 1975 for monitoring cutaneous circulation. Light from a helium-neon laser is carried by a fiberoptic cable to the skin where it interacts with stationary structures and moving blood cells within a sample volume of approximately 1 mm. Back- scattered light from the moving cells is shifted in velocity using the Doppler principle, while back-scattered light from stationery objects remains at its original frequency. The mixing of these lightwaves is translated into an electrical signal, and mathematical estimations of blood flow can be made in normal versus study areas of skin. This technique is easy to use and noninvasive (although the probe must be held against the skin), and provides immediate results.

Sørensen's group in Denmark first reported the use of a laser Doppler flowmeter on a burn unit. Initial studies in our burn center showed excellent correlation with full-thickness burns (no flow) and shallow burns (normal or increased flow), but considerable variation from patient to patient with comparable burns, and, at different times, in the same area in the same patient with moderate and deep dermal burns. Readings varied particularly with: (1) temperature (immediately after bathing, a warm room, etc.), (2) the patient's state of anxiety (catecholamine response), and (3) elevation of an extremity (e.g., cutaneous blood flow essentially disappeared when the hands were elevated).

In further studies, the cutaneous circulation over time was followed in burned patients and rats to determine whether measuring changes in cutaneous blood flow would help predict the ultimate fate of indeterminate burns that were not obviously shallow or deep. A laser Doppler flowmeter was used to study cutaneous perfusion for at least 72 h in partial-thickness wounds on patients with burns of less than 15 percent TBSA, and in experimental wounds of similar size on rats. Clinical wounds that healed without grafting within 3 weeks consistently showed elevated perfusion levels that continued to increase over 72 h. Wounds eventually requiring grafting demonstrated lower perfusion levels with no obvious pattern of increase. The observed trends of increased flow for healing burns and flat flow for nonhealing burns were constant. Thus for patients to be followed for several days, this method has merit in deciding, after day 3, which patients will benefit from excision and grafting.

As instrumentation has improved, so have the results of serial measurements of burn blood flow. A differential analysis of multiple parameters during measurement has brought accuracy to the 80 to 90 percent range in predicting healing within 3 weeks.

Thermography

Diminished blood flow to deep dermal and full-thickness burns makes them cooler to touch, a finding confirmed by thermography in 1974 by Hackett. Initial studies of thermography as a tool for predicting the need for excision and grafting of deep dermal wounds in 30 patients presented by Mason in 1981 suggested that thermography might be more accurate than clinical judgment. Thermographic findings, like laser Doppler flowmetry, are highly dependent on room and patient temperature, the patient's anxiety and stress level, and the area of the body being considered, as well as the cooling effect of evaporative water loss. Despite these drawbacks, Cole, in 1990, compared thermography to clinical assessment of 32 burned hands. Superficial and deep partial-thickness burns were treated conservatively, with excision and grafting only of those that had not healed 2 to 3 weeks after injury. The group that had undergone the delayed surgical procedure and the healed group were analyzed retrospectively to determine the predictive value of the initial clinical evaluation and thermography as assessments of the depth of the burns. Obvious full-thickness burns were excised and grafted within 5 days and were not included in the study. Initial thermographic assessment correctly predicted the outcome (whether healed or excised and grafted) in 33 of 36 burns. This relationship was highly significant, while initial clinical assessment of depth had no significant relationship with the time taken to heal.

Ultrasound

Moserová, using an industrial ultrasound device in 1982, was able to detect differences between normal pig skin and pig skin that had been scalded for 5 and 15 s. His initial work was expanded as technology improved. Cantrell was able to detect denatured from normal collagen in pig skin. One problem with this technique is that collagen denatures at 149°F (65°C) while epidermal cells, from which the burn must heal, are killed at about 117°F (47°C), so ultrasonic apparent depth is likely to be underestimated. Further refinements in instrumentation enabled Brink to demonstrate a significant correlation between ultrasound burn depth and histology in pigs. Wachtel, in five burn patients, reported that ultrasound comparisons with clinical evaluation and histopathologic studies of burn wound biopsy specimens of the same burned areas failed to show any substantive improvement in predicting the depth of burn by the ultrasonic scanning techniques. Technology in this area has dramatically improved, so this may still be a technique of the future.

Nuclear Magnetic Resonance

Full-thickness burns result in slower resorption of wound edema than partial-thickness burns. Proton NMR parameters correlate with tissue water content, and Koruda tested whether proton NMR could distinguish between full-thickness and partial-thickness burns. Early after burning, NMR could distinguish higher water content in partial- and full-thickness rat burns than in adjacent normal skin. By 48 h, the partial-thickness burn had returned to control values, while the full-thickness burn remained edematous. However, the rat skin had to be excised in order to use NMR, and so there is no useful clinical application for the technique. In partial- and full-thickness scald wounds, reductions in PCr/Pi ratios correlated with burn depth and improved over time postinjury. Future advances in technology might enable more precise measurements in the clinical situation.

Light Reflectance

The skin is relatively transparent to short-wavelength infrared light, and reduced hemoglobin absorbs more of the light than oxygenated hemoglobin. Anselmo and Zawacki, in 1973, reasoned that thrombosed vessels in full-thickness burns would become visible in infrared light, and could be distinguished from the open vessels of partial-thickness burns. Computer analysis of photographs taken with red, green, and infrared filtered light generated ratios of the green/infrared, red/infrared, and red/green images point by point over the entire surface that accurately distinguished shallow, deep dermal, and full-thickness burns. The technique, however, was too expensive, time-consuming, and slow for clinical decision making. Heimbach and Afromowitz devised a portable, noninvasive (the wound is not touched) electronic device that could instantaneously measure the spectral characteristics of red, green, and infrared light reflected from the burn. The device was essentially 100 percent accurate in differentiating shallow and full-thickness wounds. For intermediate wounds, using the endpoint of wound healing in less than or more than 3 weeks, clinical assessment by two experienced surgeons was compared to readings from the device. In about one-third of cases, the surgeons were unwilling to commit to a prediction. When surgeons were willing to make a prediction, they were incorrect about 25 percent of the time. The reflectance device was significantly more accurate than the clinical assessment in those burn wounds predicted not to heal; an accuracy of 79 percent was achieved in the clinically indeterminate wounds. The dynamic qualities of the burn wound are emphasized in these studies because the light reflectance device clearly showed changes each day for the first 3 to 4 days, and it was most accurate on day 3.

PHYSIOLOGICAL RESPONSE TO BURN INJURY

Burn patients with or without inhalation injury commonly manifest an inflammatory process involving the entire organism; the term *systemic inflammatory response syndrome (SIRS)* was introduced to summarize that condition. The most common cause of SIRS is the burn sepsis. SIRS with infection or bacteremia is a major factor determining morbidity and mortality in thermally injured patients. Pathologic alterations of the metabolic, cardiovascular, gastrointestinal, and coagulation systems occur with hypermetabolism, an increase in cellular, endothelial, and epithelial permeability, typical hemodynamic alterations, and often extensive microthrombosis. The circulatory manifestations of the systemic inflammatory response largely disappear within 24 h, but the patient remains in a hypermetabolic state until wound coverage is achieved.

Burn Shock

Burn shock is a complex process of circulatory and microcirculatory dysfunction, not easily or fully repaired solely by fluid resuscitation. Hypovolemic shock and tissue trauma result in formation and release of local and systemic mediators, which produce an increase in vascular permeability or an increase in microvascular hydrostatic pressure. Most mediators act to increase permeability by altering venular membrane integrity. The early phase of burn edema, lasting from minutes to an hour, is attributed to mediators such as histamine, bradykinin, and vasoactive amines, products of platelet activation, and the complement cascade of hormones, prostaglandins, and leukotrienes. Vasoactive amines also may act directly by increasing microvascular blood flow or vascular pressures, accentuating the burn edema.

Histamine is probably responsible for the early phase of increased capillary permeability after burn injury, because it is released in large quantities from mast cells in burned skin immediately after injury. Histamine increases leakage of fluids and proteins from systemic microvessels; its major effect is on the venules, in which an increase in intravascular junction space is characteristic. Serum histamine peaks in the first several hours postinjury, suggesting that histamine is involved only in the very early increase in permeability.

Serotonin is released immediately upon postburn platelet aggregation and acts directly to increase pulmonary vascular resistance and indirectly to amplify the vasoconstrictive effects of norepinephrine, histamine, angiotensin II, and prostaglandin.

Prostaglandins, vasoactive products of arachidonic acid metabolism, are released in burn tissue and contribute to formation of burn edema. These substances do not directly alter vasopermeability, but increased levels of vasodilator prostaglandins, such as PGE₂, and prostacyclin, PGI₂, result in arterial dilatation in burn tissue, increasing blood flow and hydrostatic pressure in the injured microcirculation, accentuating the edema process. Increased concentrations of PGI₂ and the vasoconstrictor thromboxane A₂ have been demonstrated in burn tissue, burn blister fluid, lymph, and wound secretion.

The activation of the proteolytic cascades, including those of coagulation, fibrinolysis, the kinins, and the complement system, occurs immediately after burn injury. Kinins, specifically the bradykinins, increase vascular permeability, primarily in the venule. In addition to the loss of capillary integrity, thermal injury also causes changes at the cellular level. Baxter demonstrated a generalized decrease in cellular transmembrane potential involving nonthermally injured and thermally injured cells. Platelet activating factor is released after burn injury and increases capillary permeability. The reduction of cardiac output after burn injury is a result of hypovolemic and cellular shock, increased systemic vascular resistance due to sympathetic stimulation and hypovolemia, with release of catecholamines, vasopressin, angiotensin II, and neuropeptide Y. After successful resuscitation, cardiac output normalizes after 18 to 24 h and increases to supernormal levels during the wound-healing phase of burn management.

Metabolic Response to Burn Injury

Hypermetabolism

Resting energy expenditure (REE) after burn injury can be as much as 100 percent above predictions based on standard tables for size, age, sex, and weight. Some debate persists regarding the genesis of this phenomenon, but increased heat loss from the burn wound and increased b-adrenergic stimulation are probably primary factors. Radiant heat loss is increased from the burn wound secondary to high blood flow. Measurement of resting energy expenditure is helpful in assessing nutritional status. On average, resting energy metabolism equals approximately 1.3 times the predicted basal energy expenditure obtained from the Harrison-Benedict equation. Glucose metabolism is elevated in almost all critically ill patients, including those with burn injuries, but studies have focused particularly on burn patients because their relatively stable condition allows reproducible experimental conditions. Gluconeogenesis, particularly from alanine, and glycogenolysis are increased. Protein is excreted primarily in the urine as urea, contributing to the progressive depletion of body protein stores. Proteolysis in burn patients is increased as compared to normal individuals who are fed the same amount of protein and calories. This results in increased efflux of amino acids from the muscle, including glucogenic amino acids. Alanine, in particular, is released at an increased rate. Increased gluconeogenesis from amino acid renders these amino acids unavailable for reincorporation into the body protein. Plasma insulin levels that are usually normal are elevated in burn patients. The basal rate of glucose production is elevated despite a normal or elevated plasma insulin level, which can be defined as hepatic insulin resistance. Fatty acids are released at a rate in excess of requirements of fatty acids and energy substrates. In burn patients, over 70 percent of released fatty acids are not oxidized, but rather reesterified into triglyceride,

resulting in fat accumulation in the liver. This is unfortunate because utilization of fat for energy would decrease dependence on proteolysis.

A number of studies have been designed to assess the optimal form of carbohydrate and fat. Both xylitol and fructose have been posed as alternative carbohydrates, and a wide variety of fats have been advocated, including fish oil, medium-chain triglycerides, and structured lipids in which medium- and long-chain fatty acids are incorporated into the same triglyceride molecule. Xylitol and fructose require energy for transposition to glucose before utilization for energy. Xylitol in large quantities is hepatotoxic, and fructose may cause lactic acidosis.

Major trauma, burns, and sepsis have in common a rapid net catabolism of body protein, as well as a redistribution of the nitrogen pool within the body. Muscle protein breakdown is accelerated while “acute phase” proteins are produced at an increased rate in the liver. Wound repair requires amino acid protein synthesis and increased immunologic activity, and may require accelerated protein synthesis. Protein intake over 1 g/kg/day has been recommended for thermally injured patients; for thermally injured patients with normal renal function, the recommended protein intake is 2 g/kg/day. The importance of glutamine intake in critically ill patients has been investigated. Current oral and enteral formulations have largely omitted glutamine, but some beneficial effects have been found with enteral glutamine. Including glutamine in intravenous formulations results in improved nitrogen balance, which, while statistically significant, probably is not biologically significant because of its insufficient magnitude.

Neuroendocrine-Mediator Response

Catecholamines appear to be the major endocrine mediators of the hypermetabolic response in thermally injured patients. Pharmacologic blockade of beta receptors diminishes the intensity of postburn hypermetabolism. Thyroid hormonal serum concentrations are not elevated in patients with large burns. Total thyronine (T₃) and thyroxine (T₄) concentrations are reduced, and reverse T₃ concentrations are elevated while cellular concentrations are likely normal. Concentrations of free T₃ and T₄ fall markedly in the presence of sepsis in burned patients.

Burn injury abolishes the normal diurnal variation in glucocorticoid concentrations, but these hormones do not appear to influence metabolic activity directly and only have a permissive role in catecholamine stimulation. They are, however, primarily responsible for increased proteolysis. Glucagon concentrations are related directly to metabolic rate and cortisol concentrations and may modulate resting metabolism through anti-insulin effects.

Immune Response to Burn Injury

The immune status of the burn patient has a profound impact on outcome in terms of survival, death, and major morbidity. The greatest difficulty in attempting to decipher the body's response to injury is the complex interaction of the cytokine cascade, the arachidonic acid cascade, and the neuroendocrine axis.

Cytokine Cascade

Cytokines were considered originally to be regulatory chemicals secreted by cells of the immune system, and growth factors were seen as chemicals originating from inflammatory and reparative tissue. The distinction between growth factors and peptides, hormones, and cytokines is no longer distinct. Many of these cytokines and growth factors are released from damaged tissues at the wound site, where they exert local and systemic effects.

After injury, a number of cytokines are induced rapidly, including *tumor necrosis factor (TNF)*, interleukin-1 (IL-1), and interleukin-6 (IL-6). The timetable of induction is similar in burned and/or injured patients. Tumor necrosis factor alpha (TNF-α) is detectable early in the period of burn shock; the maximum level of TNF-α throughout the course is of prognostic significance. Because the physiological effects of TNF are almost indistinguishable from endotoxin, the induction of TNF has been held responsible for the clinical effects of endotoxemia. After initial induction of IL-1, synthesis is impaired significantly for several days after injury. By contrast, there is up-regulation of local production of IL-1 and IL-6 in inflammatory sites, inducing polymorphonuclear neutrophil chemoattraction.

Interleukin-2 (IL-2) is a key cytokine in the mediation of the cellular immune response, and patients with large burns have significantly suppressed production of IL-2, which correlates with the length of time from injury. Investigations of pulmonary production of IL-8 after respiratory distress syndrome are not inconclusive. Interleukin-6 (IL-6) is a ubiquitous cytokine that is produced by a variety of cells, and is detected in increased concentration in blood or tissues after disturbances of physiological homeostasis, including thermal injury. Its most important role is the induction of hepatic synthesis of acute phase response proteins that include antibacterial products, such as α₂-glycoprotein, C₃, and fibronectin.

Arachidonic Cascade

The major product of arachidonic acid cascade after thermal injury is *prostaglandin E₂ (PGE₂)*, produced by macrophages and partially mediated by endotoxin. PGE₂ exerts its immunosuppressive effect primarily by inhibition of lymphocyte IL-2 production and T-cell activation and down- regulation of IL-6. The induction of the arachidonic acid cascade is complex, and is apparently dependent at least on two pathways, one calcium-channel- dependent and one calcium-independent. There are massive increases of another arachidonic acid derivative, thromboxane B₂, in the plasma of burn patients, particularly immediately postburn and during septic episodes. Leukotriene B₄, another arachidonic acid product, is a potent neutrophil chemotactin that is produced after thermal injury.

Cell-Mediated Immunity

Cell-mediated immunity is impaired after burn injury, including documented delays in allograft rejection, impairment in mitogenic and anogenic responsiveness of lymphocytes, burn-size-related suppression of graft- versus-host activity, suppression of delayed cutaneous sensitivity tests, and diminution of peripheral lymphocytes and thoracic duct lymphocyte concentration. There is agreement that the functional capacity of thymodependent lymphocytes (T cells) to perform their normal physiological response is impaired. Whether this failure is the result of “overuse” or indirectly the result of down-regulation by cytokine cascades and other products of the inflammatory reaction is unclear.

Macrophages

Macrophage function is impaired after thermal injury. Macrophage products suppress mitogenic responsiveness in normal lymphocytes. Proinflammatory cytokines are produced by macrophages in short bursts, probably inhibited by a feedback loop with decreased receptor expression. Activation includes pulmonary macrophages and may provide the background for the development of the adult respiratory distress syndrome seen in burn patients.

B Cells

The function of bone marrow–derived thymocytes, or B cells, after thermal injury is less well documented than that of the macrophages or T cells. The B-cell population is subject to the same nonspecific activation as the rest of the lymphocyte population.

Neutrophils

Neutrophil dysfunction after thermal injury has been extensively studied; it is manifested by a decreased Fc receptor expression, depressed intracellular killing capacity, and leukocyte chemotaxis that is accompanied by a brief increase in neutrophil respiratory burst. In addition, expression of CD16 (FcR, Fc, IgG receptors) and CD11 (adhesion molecules) on neutrophils is impaired after thermal injury, and this reduction seems to be directly related to the appearance of bacteremia and pneumonia. Baseline granulocyte oxidative activity in burn neutrophils is increased. Induction of neutrophil activation probably requires several different stimulants, but it is known that TNF and endotoxin can both activate neutrophils; IL-6 is also a potent inducer of superoxide production by the neutrophil.

Humoral Immunity

After thermal injury there is a marked diminution of total serum IgG concentration and all subclasses. These levels return to normal between 10 and 14 days postburn. Extremely low levels of IgG on admission are predictive of a poor prognosis. These changes have been ascribed to a combination of leakage through the burn wound, protein catabolism, and a relative diminution in synthesis of IgG. IgM and IgA levels appear to be relatively unaffected.

The classical and the alternative complement pathways are depleted, but the alternative pathway is more profoundly altered. Complement inactivation by heat appears to ameliorate cell-mediated immunosuppression, suggesting that some of the impairment of the cell- mediated immunosuppression postburn may be due to a complement- associated mechanism. The production of granulocyte colony-stimulating factor (GCSF) and of granulocyte-macrophage colony-stimulating factor (GM-CSF) is also impaired.

Unidentified or partially identified immunosuppressive factors are present in burn serum and in burn subeschar tissue fluid. A low-molecular-weight immunosuppressive peptide has been found in blood that is capable of suppressing neutrophil chemotaxis, T-cell blastogenesis, and B-cell blastogenesis.

FLUID MANAGEMENT

Proper fluid management is critical to survival in major thermal injury. In the 1940s, hypovolemic shock or shock-induced renal failure was the leading cause of death after burn injury. Mortality related to burn-induced volume loss has decreased considerably with increased knowledge of the massive fluid shifts and hemodynamic changes that occur during burn shock. A vigorous approach to fluid therapy has led to reduced mortality rates in the first 48 h postburn, but 50 percent of the deaths occur within the first 10 days after burn injury from a multitude of causes. One of the most significant causes is inadequate fluid resuscitation therapy. Fluid management after burn shock resuscitation is also important.

Historical Perspective

The necessity for fluid resuscitation after burn injury was appreciated over a century ago, but the magnitude of the fluid loss was not apparent until the studies of Frank Underhill of the victims of the Rialto Theater fire in 1921. That burn shock was due to extravascular fluid loss was further elucidated by Cope and Moore, who conducted studies on patients from the Coconut Grove disaster in 1942. They developed the concept of burn edema and introduced the body weight–burn formula for fluid resuscitation. In 1952, Evans developed a burn surface area–weight formula that became the first simplified means of calculating fluid resuscitation needs for burn patients. Surgeons at the Brooke Army Medical Center modified the original Evans formula and this became the standard for the next 15 years.

Pathophysiology of Burn Shock

Burn shock is hypovolemic and cellular in nature, and is characterized by specific hemodynamic changes including decreased cardiac output, extracellular fluid, and plasma volume, and oliguria. As with other forms of shock, the primary goal is to restore and preserve tissue perfusion. In burn shock, resuscitation is complicated by obligatory burn edema, and the voluminous transvascular fluid shifts that result from a major burn are unique to thermal trauma.

One major component of burn shock is the increase in total body capillary permeability. Direct thermal injury results in marked changes in the microcirculation. Most of the changes occur locally at the burn site; maximal edema formation occurs between 8 and 12 h postinjury in smaller burns, and 12 and 24h postinjury in major thermal injuries. The rate of

progression of tissue edema is dependent upon the adequacy of resuscitation.

Multiple mediators have been proposed to explain the changes in vascular permeability. The proposed mediators produce an increase in vascular permeability or an increase in microvascular hydrostatic pressure. The end result of the changes in the microvasculature due to thermal injury is disruption of normal capillary barriers separating intravascular and interstitial compartments, and rapid equilibrium between these compartments. Plasma volume is severely depleted, clinically manifested as hypovolemia, with a marked increase in extracellular fluid.

Thermal injury also causes changes at the cellular level. Baxter has demonstrated that in burns greater than 30 percent TBSA, there is a systemic decrease in cell transmembrane potential, involving nonthermally injured cells. This decrease in cell-transmitting potential, defined by the Nernst equation, results from an increase in intracellular sodium concentration secondary to a decrease in sodium ATPase activity responsible for maintaining the intracellular-extracellular ionic gradient. Resuscitation only partially restores the membrane potential and intracellular sodium concentrations to normal levels, demonstrating that hypovolemia, with its attendant ischemia, is not totally responsible for the cellular swelling seen in burn shock. Membrane potential may not return to normal for many days postburn despite adequate resuscitation. If resuscitation is inadequate, cell membrane potential progressively decreases, resulting ultimately in cell death.

Moyer, Baxter, and Shires established the role of crystalloid solutions in burn resuscitation, and delineated the fluid volume changes in the early postburn period. Moyer's studies, in 1965, demonstrated that burn edema sequesters enormous amounts of fluid, resulting in the hypovolemia of burn shock. Baxter and Shires in 1968, using radioisotope dilution techniques, demonstrated that edema fluid in the burn wound is isotonic with respect to plasma, and contains protein in the same proportions as that found in blood. This was further evidence of the complete disruption of the normal capillary barrier in major burns, with free exchange between plasma and extravascular extracellular compartments. In a canine model, they also established the end points of crystalloid resuscitation as optimal cardiac output and restoration of ECF at the end of 24 h. Clinical studies confirmed the efficacy of restoring ECF to within 10 percent of controls within 24 h. This became the basis for the Baxter (Parkland) formula. The associated mortality rate was comparable to that obtained with a colloid-containing resuscitation formula.

Moncrief and Pruitt characterized the hemodynamic alterations in burn shock with and without fluid resuscitation. Their efforts culminated in the Brooke formula modification, which was based on 2 mL per kilogram of body weight per percentage of TBSA burned (2 mL/kg body weight/% burn) during the first 24 h. Fluid needs were estimated initially according to the modified Brooke formula, but the actual volume for resuscitation was based on clinical response. In their study, resuscitation permitted an average decrease of about 20 percent in both extracellular fluid and plasma volume, but no further loss accrued in the first 24 h. In the second 24 h postburn, plasma volume restoration occurred with the administration of colloid. Cardiac output, initially low, rose over the first 18 h postburn, despite plasma volume and blood volume deficits. Peripheral vascular resistance rose during the initial 24 h, but decreased as cardiac output improved. When plasma volume and blood volume loss ceased, cardiac output rose to supranormal levels where it remained until healing or grafting occurred.

Moylan and associates in 1973, using a canine model, defined the relationships between fluid volume, sodium concentration, and colloid in restoring cardiac output. No significant colloid effect on cardiac output was noted in the first 12 h postinjury. In addition, 1 meq of sodium was found to exert an effect on cardiac output equal to 13 times that of 1 mL of salt- free fluid volume. Thus, any combination of sodium and fluid volume within the broad limits of the study would effectively resuscitate a thermally injured patient.

Arturson's 1979 studies characterized the nature of the "leaky capillary" in the postburn period. In a canine model, increased capillary permeability was found locally and in remote nonburned tissue when the TBSA burned exceeded 25 percent. He proposed that the burn wound is characterized by rapid edema formation due to dilatation of the resistance vessels (precapillary arterioles), increased extravascular osmotic activity due to the products of thermal injury, and increased microvascular permeability to macromolecules. The increased permeability permits molecules with molecular weights of up to 350,000 to escape from the microvasculature, a size that allows essentially all elements of the vascular space, except red blood cells, to escape. Studies by Demling and co-workers have demonstrated that in 50 percent TBSA burns, one-half of the initial fluid resuscitation requirement may end up in nonthermally injured tissues.

Resuscitation from Burn Shock

The primary goal of fluid resuscitation is to replace fluid sequestered as a result of thermal injury. The critical concept in burn shock is that massive fluid shifts can occur even though total body water remains unchanged. What actually changes is the volume of each fluid compartment, with intracellular and interstitial volumes increasing at the expense of plasma volume and blood volume. The edema is accentuated by the resuscitation process. The National Institutes of Health consensus summary on fluid resuscitation in 1978 was not in agreement in regard to a specific formula, but there was consensus on two major issues: general guidelines to be used during the resuscitation process, and type of fluid. The volume infused should be the least amount of fluid necessary to maintain adequate organ perfusion and should be continually titrated to avoid under- or overresuscitation. Replacement of the extracellular salt lost into the burned tissue and into the cell is essential for successful resuscitation.

Crystalloid Resuscitation

Crystalloid, in particular lactated Ringer's solution with a sodium concentration of 130 meq/L, is the most popular resuscitation fluid. Proponents of the use of crystalloid solution argue that other solutions, specifically colloids, are no better, and certainly more expensive, than crystalloid for maintaining intravascular volume after thermal injury. Even large proteins leak from the capillary after thermal injury, negating any theoretical advantage from colloid. Capillaries in nonburned tissues may maintain relatively normal protein permeability characteristics.

The quantity of crystalloid needed is dependent upon the parameters used to monitor resuscitation. If a urinary output of 0.5 mL/kg of body weight/h indicates adequate perfusion, approximately 3 mL/kg/% burn will be needed in the first 24 h. If 1 mL/kg body weight/h is optimal, considerably more fluid will be needed, and more edema will result. The Parkland formula recommends 4 mL/kg/% burn in the first 24 h, with one-half of that amount administered in the first 8 h (Table 7-1). The modified Brooke formula recommends beginning burn shock resuscitation at 2 mL/kg/% burn in the first 24 h (Table 7-1). In major burns, severe hypoproteinemia usually develops with these resuscitation regimens. The hypoproteinemia and interstitial protein depletion may result in more edema formation.

Hypertonic Saline

The resuscitation of burn patients with salt solution of 240 to 300 meq/L rather than lactated Ringer's solution results in less edema because of the smaller total fluid requirements. Urine output is used as the indicator of adequate resuscitation. Demling and colleagues demonstrated in an animal model that the net fluid intake was less if burned animals were resuscitated with hypertonic saline to the same cardiac output as compared with lactated Ringer's. Urine output was much higher with hypertonic solution. Soft tissue interstitial edema in burned and nonburned tissue, as reflected by lymph flow, was increased with hypertonic saline to a similar extent as with lactated Ringer's. A shift of intracellular water into extracellular space occurs as the result of the hyperosmolar solution. Extracellular edema increases as intracellular fluid decreases, giving the external appearance of less edema. Several studies have reported that this intracellular water depletion does not appear to be deleterious, but the issue is controversial. Current recommendations are that the serum sodium levels should not be allowed to exceed 160 meq/dL. Gunn and associates in a prospective randomized study of patients with 20 percent TBSA burns evaluated hypertonic sodium lactate versus lactated Ringer's solution and were not able to demonstrate decreased fluid requirements, improved nutritional tolerance, or decreased body weight gain percentage.

We have used a modified hypertonic solution in major thermal injuries of greater than 40 percent TBSA. The resuscitation fluid contains 180 meq Na^+ (lactated Ringer's + 50 meq NaHCO_3), and is used until reversal of metabolic acidosis has occurred, usually by 8 h postburn. Resuscitation is begun at 4 mL/kg/% burn, but the administered volume is titrated to maintain urine output at 30 to 50 mL/h. After 8 h, lactated Ringer's is administered to maintain urine output at 30 to 50 mL/h. Hyponatremia is avoided in infants and in the elderly.

Colloid Resuscitation

Plasma proteins generate the inward oncotic force that counteracts the outward capillary hydrostatic force. Without protein, plasma volume could not be maintained. Protein replacement was an important component of early formulas for burn management. The Evans formula used 1 mL/kg body weight/% burn each for colloid and lactated Ringer's, over the first 24 h. In the original Brooke Army Hospital formula, 0.5 mL/kg/% burn was administered as colloid, and 1.5 mL/kg/% burn as lactated Ringer's. Moore's formula proposed a substantial amount of colloid. Considerable confusion exists concerning the role of protein (albumin) in a resuscitation formula. There are three approaches:

(1) Protein solutions are not given in the first 24 h because during this period they are no more effective than salt containing crystalloid water in maintaining intravascular volume. Protein may also promote accumulation of lung water when edema fluid is absorbed from the burn wound.

(2) Proteins, specifically albumin, should be given from the beginning of resuscitation with crystalloid.

(3) Protein should not be given between 8 h to 12 h postburn because of the massive fluid shifts during this period, after which they should be used.

Demling demonstrated that restoration and maintenance of plasma protein content are not effective until 8 h postburn, when adequate levels can be maintained with infusion. Because nonburned tissues appear to regain normal permeability shortly after injury and because hypoproteinemia may accentuate the edema, the first alternative is the least appropriate.

Heat-fixed plasma protein solutions, e.g., Plasmanate, contain some denatured and aggregated protein, which decreases the oncotic effect. Albumin solutions are clearly the most oncologically active solutions. Fresh frozen plasma contains all the protein fractions that exert the oncotic and the nononcotic functions. The optimal amount of protein remains undefined. Demling uses between 0.5 and 1 mL/kg/% burn of fresh frozen plasma during the first 24 h, beginning at 8 to 10 h postburn, and also argues that while all major burns require large amounts of fluid, the following groups of patients develop less edema and better maintain hemodynamic stability if fresh frozen plasma is used during resuscitation: older patients with burns, patients with burns and concomitant inhalation injury, and patients with burns in excess of 50 percent TBSA.

Slater and co-workers used fresh frozen plasma during burn shock, lactated Ringer's, 2 L/24 h, and fresh frozen plasma, 75 mL/kg/24 h (Table 7-1). The volume of fresh frozen plasma is calculated, but the volume infused is titrated to maintain an adequate urine output. These authors use colloid early in the burn shock period, but most burn patients have received lactated Ringer's in significant volumes during field management. In the young pediatric burn patient with major burn injury, colloid replacement is frequently required because serum protein concentration rapidly decreases.

Dextran

Dextran is a colloid consisting of glucose molecules that have been polymerized into chains to form high-molecular-weight polysaccharides. This compound is available commercially in a number of molecular sizes. Dextran, with an average molecular weight of 40,000, is referred to as low- molecular weight-dextran. British dextran has a mean molecular weight of 150,000, but the dextran used in Sweden has a molecular weight of 70,000. Dextran is excreted by the kidneys, with 40 percent removed within 24 h, and the

remainder is slowly metabolized. Demling and associates used dextran 70 in a 6% solution to prevent edema in nonburned tissues. Dextran 70 is associated with some risk of allergic reaction, and can interfere with blood typing. Dextran 40 improves the microcirculatory flow by decreasing red blood cell aggregation. The net requirements for maintaining vascular pressure at the baseline levels with dextran 40 are about half those noted with lactated Ringer's alone during the first 24 h postburn, with an infusion rate of dextran 40 and saline of 2 mL/kg/h with sufficient lactated Ringer's to maintain adequate perfusion. At 8 h, an infusion of fresh frozen plasma at 0.5 to 1.0 mL/kg/% burn over 18h is instituted with necessary additional crystalloid (see Table 7-1).

Special Considerations in Burn Shock Resuscitation

Fluid Resuscitation in the Thermally Injured Pediatric Patient

The burned child represents a special challenge, because resuscitation therapy must be more precise than that for an adult with a similar burn. Children have a limited physiological reserve. Children require proportionately more fluid for burn shock resuscitation than adults with similar thermal injury; fluid requirements for children average 5.8 mL/kg/% burn. Children commonly require intravenous resuscitation for relatively small burns of 10 to 20 percent TBSA, a finding confirmed by Baxter. Graves and associates give children 6.3 ± 2 mL/kg/% burn. The Cincinnati Shriners Burns Institute uses the Parkland formula, adding maintenance fluid to the resuscitation fluid volume, to begin burn shock resuscitation, 4 mL/kg/% burn, plus 1500 mL maintenance fluid per square meter of burn surface area (BSA), per 24 h (Table 7-2). Graves found that if maintenance fluids were subtracted from the resuscitation fluid requirements, the resulting resuscitation volume approached 4 mL/kg/% burn. At the Galveston Shriners Burns Institute, fluid requirements are estimated according to a formula based on TBSA, and burned BSA, in square meters. Total fluid requirements for the first day are estimated as follows: $5000 \text{ mL/m}^2 \text{ \% burn/24 h} + 2000 \text{ mL/m}^2 \text{ BSA/ 24 h}$.

Inhalation Injury

Inhalation injury increases the fluid requirements for resuscitation from burn shock after thermal injury. Patients with documented inhalation injury require 5.7 mL/kg/% burn, as compared to 3.98 mL/kg/% burn in patients without inhalation injury. Inhalation injury accompanying thermal trauma increases the magnitude of total body injury and requires increased volumes of fluid and sodium to achieve resuscitation.

Choice of Fluids and Rate of Administration

All the solutions reviewed are effective in restoring tissue perfusion. Most patients with burns under 40 percent TBSA with no pulmonary injury can be resuscitated with isotonic crystalloid fluid. In patients with burns of over 40 percent TBSA, and/or in patients with pulmonary injury, hypertonic saline can be used in the first 8 h postburn, after which lactated Ringer's is infused to complete resuscitation. In pediatric and elderly burn patients, using less concentrated but hypertonic concentrations of sodium (e.g., 180 meq/L) gives the benefits of hypertonic resuscitation without the potential complications of excessive sodium retention and hypernatremia.

In patients with massive burns, young pediatric patients, and burns complicated by severe inhalation injury, a combination of fluids can be used to achieve the desired goal of tissue perfusion while minimizing edema. In these patients, the regimen of modified hypertonic (lactated Ringer's + 50 meq NaHCO₃) saline fluid containing 180 meq Na/L is used for the first 8 h. After correction of the metabolic acidosis, which usually requires 8 h, the patients are given lactated Ringer's only for the second 8 h. In the last 8 h, 5% albumin in lactated Ringer's solution completes the resuscitation. The resuscitation solution used in Galveston for pediatric patients is an isotonic glucose-containing solution to which a moderate amount of colloid (human serum albumin) is added. The solution is prepared by mixing 50 mL of 25% human serum albumin (12.5 g) with 950 mL of 5% dextrose in a lactated Ringer's solution.

The volume of infused fluid should maintain a urine output of 30 to 50 mL/h in adults and 1 mL/kg/h in children. In children weighing more than 50 kg, the urine volume should not exceed 30 to 50 mL/h. Heart rate and blood pressure are not indicative of fluid volume status in the burn patient; therefore, fluid volume status and cardiac output should be measured directly via thermodilution pulmonary artery catheterization, but a low measured filling pressure with evidence of adequate perfusion is common. Placement of a Swan-Ganz catheter to monitor burn shock resuscitation should be reserved for burn patients with limited cardiac reserve, such as the elderly or patients with significant concomitant disease, or burn patients who require large volumes.

None of the resuscitation formulas can be more than general guidelines for burn shock resuscitation. The Parkland formula, for instance, decreases the volume administered by 50 percent at 8 h postburn. The relationship between the fluid volume required and time postburn depicted by the smooth curve in Fig. 7-3 represents the influence of temporal changes in microvascular permeability and edema volume on fluid needs. The gentle changes depicted by that curve are in sharp contrast with the abrupt changes in fluid infusion rate prescribed by the formula.

Resuscitation is considered successful when there is no further accumulation of edema fluid, usually between 18 and 30 h postburn, and the volume of infused fluid needed to maintain adequate urine output approximates the maintenance fluid volume, which is the patient's normal maintenance volume plus evaporative water loss.

Fluid Replacement Following Burn Shock Resuscitation

Heat-injured microvessels may manifest increased vascular permeability for several days, but the rate of fluid loss is considerably less than that seen in the first 24 h. Burn edema at 24 h postburn is near maximal, and the interstitial space may well be saturated with sodium. Additional fluid requirements depend on the type of fluid used during the initial resuscitation. If hypertonic salt resuscitation was used during the entire burn shock period, a hyperosmolar state is produced, and the addition of free water is required to restore the extracellular space to an iso-osmolar state.

If colloid was not been used during burn shock and the serum oncotic pressure is low because of intravascular protein depletion, protein repletion frequently is needed. Protein requirement varies with the resuscitation used. The Brooke formula proposes 0.3 to 0.5 mL/kg/TBSA burn of 5% albumin during the second 24 h. The Parkland formula replaces the plasma volume deficit, which varies from 20 to 60 percent of the circulating plasma volume, with colloid. We have used colloid replacement based on a 20 percent plasma volume deficit during the second 24 h (circulating plasma volume $\times$ 20 percent).

In addition to colloid, the patients should receive maintenance fluids. The total daily maintenance fluid requirement in the adult patient is calculated by the following formula, where m₂ is square meters of TBSA:

Total maintenance fluid = (1500 mL/m²) +

evaporative water loss [(25 + % burn) $\times$ m² $\times$ 24]

This fluid may be given intravenously or via enteral feeding. The solution infused intravenously should be 50% normal saline with potassium supplements. Because of the loss of intracellular potassium during burn shock, the potassium requirement in adults is about 120 meq/day.

After the initial 24 to 48 h postburn period of resuscitation, urinary output is an unreliable guide to sufficient hydration. Respiratory water losses; osmotic diuresis secondary to glucose intolerance; high-protein, high- calorie feedings; and derangements in antidiuretic hormone (ADH) mechanisms contribute to increased fluid losses, despite an adequate urine output. Adult patients with major thermal injuries require a urine output of 1500 to 2000 mL/24 h; children require 3 to 4 mL/kg/h.

Measurement of serum sodium concentration is not only a means of diagnosing dehydration, but the best guide for managing successful fluid replacement. Other useful laboratory indices of the state of hydration include body weight change, serum and urine nitrogen concentrations, serum and urine urea glucose concentrations, the intake and output record, and clinical examination.

For very large burns, and in the pediatric burn patient, continuous colloid replacement may be required to maintain colloid oncotic pressure Maintaining serum albumin levels above 2.0 g/dL is desirable. Electrolytes, calcium, magnesium, and phosphate should also be monitored and maintained within normal limits.

RESPIRATORY INJURY

Of the nearly 50,000 fire victims admitted to hospitals each year, smoke or thermal damage to the respiratory tree may occur in as many as 30 percent. Carbon monoxide poisoning, thermal injury, and smoke poisoning are three distinctly separate aspects of clinical inhalation injury, and though symptoms and treatment are distinct, they may coexist and require concomitant treatment.

Carbon Monoxide Poisoning

As many as 60 to 70 percent of deaths from house fires can be attributed to carbon monoxide poisoning. Carbon monoxide is a colorless, odorless, tasteless gas that has an affinity for hemoglobin 200 times greater than oxygen. When inhaled and absorbed, carbon monoxide binds to hemoglobin to form carboxyhemoglobin (COHb). COHb interferes with oxygen delivery to tissues by at least four mechanisms. First, it prevents reversible displacement of oxygen on the hemoglobin molecule. Second, COHb shifts the oxygen-hemoglobin dissociation curve to the left, thereby decreasing oxygen unloading from normal hemoglobin at the tissue level. Third, carbon monoxide inhibits the cytochrome oxidase a₃ complex, resulting in less effective intracellular respiration. Fourth, carbon monoxide may bind to cardiac and skeletal muscle, causing direct toxicity, and act in the central nervous system in a poorly understood fashion, causing demyelination and associated neurologic symptoms. The degree of enzymatic and/or muscle impairment may not be directly correlated with the levels of blood carboxyhemoglobin. Levels of carboxyhemoglobin are measured easily. COHb levels less than 10 percent do not cause symptoms, although patients with exercise-induced angina may show a decreased exercise tolerance. At levels of 20 percent, healthy persons complain of headache, nausea, vomiting, and loss of manual dexterity. At 30 percent, patients become weak, confused, and lethargic. In a fire, this level can be fatal because the victim loses the ambition and the ability to flee the smoke. At levels between 40 and 60 percent, the patient lapses into coma, and levels above 60 percent are usually fatal. In very smoky fires, carboxyhemoglobin levels of 40 to 50 percent can be reached after only 2 to 3 min of exposure.

Carbon monoxide is reversibly bound to the heme pigments and enzymes and, despite its intense affinity, readily dissociates according to the laws of mass action. The half-life of carboxyhemoglobin, when breathing room air, is between 4 and 5 h. On 100% oxygen, the half-life is reduced to 45 to 60 min. In a hyperbaric oxygen chamber at 2 atm, it is about 30 min, and at 3 atm it is about 15 to 20 min.

The importance of carbon monoxide poisoning in victims of isolated smoke inhalation injury during fires was dramatically demonstrated in the 1979 MGM Grand Hotel and 1981 Hilton Hotel fires in Las Vegas. Although only a small number of burn injuries occurred, 123 people died at the scene, primarily from carbon monoxide intoxication. At the same time, the efficacy of prompt assessment and treatment of smoke inhalation was demonstrated. Of more than 400 individuals who received hospital evaluation for smoke inhalation injury, mortality was less than 1 percent, and the rate of significant complications (myocardial infarction, respiratory failure, or pneumonia) was 1 percent each.

Patients burned in an enclosed space or having any suggestion of neurologic symptoms should be placed on 100% oxygen while awaiting measured carboxyhemoglobin levels. The use of *hyperbaric oxygen (HBO)* is controversial. Only two randomized studies have been conducted. In a group of patients with CO exposure but no unconsciousness, half received normobaric oxygen, and the other half received one treatment in a hyperbaric chamber. There was no difference in immediate or delayed neurologic sequelae. Group two included patients who lost consciousness and who received either one or two HBO treatments. Again there were no differences. Hyperbaric oxygen was not useful in patients with moderate CO exposure, whether or not they lost consciousness, regardless of their COHb levels at admission. Nor were two sessions of HBO in patients with a brief loss of consciousness better than a single session. It is not clear that hyperbaric oxygen is better than normobaric oxygen.

While hyperbaric oxygen treatment may be innocuous in isolated CO poisoning, in the presence of associated burns and smoke poisoning, the rate of potentially fatal complications during treatment in the chamber is extremely high. Patients who have not lost consciousness and who have a normal neurologic examination on admission will almost always recover completely without treatment beyond administration of 100% oxygen. Patients who remain comatose in the emergency room have a poor prognosis and rarely awaken.

Thermal Airway Injury

The term “pulmonary burn” is a misnomer. True thermal damage to the lower respiratory tract and lung parenchyma is extremely rare, unless live steam or exploding gases are inhaled. The air temperature near the ceiling of a burning room may reach 540°C (1000°F) or more, but air has such poor heat-carrying capacity that most of the heat is dissipated in the nasopharynx and upper airway. The heat dissipation in the upper airway, however, may cause significant thermal injury.

Thermal injury to the respiratory tract is usually immediate, and consists of mucosal and submucosal edema, erythema, hemorrhage, and ulceration. Thermal injury is usually limited to the upper airway (above the vocal cords) and trachea for two reasons: (1) the nasopharynx and oropharynx provide a very effective mechanism for heat exchange because of their relatively large surface area and associated air turbulence, as well as their mucosal water lining that acts as a heat reservoir, and (2) sudden exposure to hot air may trigger reflex closure of the vocal cords, reducing the potential for lower airway injury. Animal experiments have demonstrated that significant heat exchange also occurs in the airway segment between the vocal cords and the tracheal bifurcation, protecting the lower airway. Thus, the lower airway is rarely exposed to hot, ambient gas at a fire scene. An exception is superheated steam inhalation, where because of energy released in the respiratory tract as the steam condenses to water, severe injury has been reported in the lower airways with measurable injury in the alveoli. In these patients the lower airway is rapidly obstructed, and they usually die from untreatable asphyxiation.

Patients with greatest risk of upper airway obstruction are those injured in an explosion (gasoline vapor, propane, butane, or natural gas) with burns of the face and upper torso, and those who have been unconscious in a fire. Mucosal burns of the mouth, nasopharynx, and larynx result in edema formation and may lead to upper airway obstruction at any time during the first 24 h postburn. Any patient with burns of the face should have a careful visual inspection of the mouth and pharynx, and, if these are abnormal, the larynx should be visualized immediately. Red or dry mucosa or small mucosal blisters raise the possibility of airway obstruction; in the presence of a closed space fire significant smoke poisoning may be present. The presence of significant intraoral and pharyngeal burns is a clear indication for early endotracheal intubation, as progressive edema can make later intubation extremely hazardous, if not impossible. Mucosal burns are rarely full-thickness, and can be successfully managed with good oral hygiene. Once the patient is intubated, the tube should remain in place for 3 to 5 days, until the edema subsides.

Smoke Inhalation

A nightclub fire that claimed 48 lives in Dublin in 1981 added a great deal to our understanding of smoke inhalation injury because the disaster site was meticulously reconstructed and the event reenacted for scientific study. Within minutes, visibility was reduced to less than 1 m and ambient temperatures reached 1160°C. Near the fire, dramatic changes in inhaled gas concentrations were noted; oxygen was reduced to less than 2 percent, carbon monoxide increased to greater than 3 percent, hydrogen cyanide was measured at 250 ppm, and hydrogen chloride was measured at 8500 ppm.

Hydrogen cyanide, a common product of polyurethane combustion, is a more effective inhibitor of cellular respiration than carbon monoxide, and also interferes with normal oxygen utilization at the tissue level. By inhibiting the final step of oxidative phosphorylation at the cytochrome a_3 level, cyanide halts aerobic metabolism, inducing lactic acidosis and cellular asphyxia. The effects of cyanide are specifically injurious to tissues with little anaerobic reserve, i.e., the central nervous system. Combined exposure to cyanide and carbon monoxide results in a deadly synergistic decrease in tissue oxygen utilization.

A vast array of toxic products are released during combustion (flaming) or pyrolysis (smoldering), depending on the type of fuel that is burned, whether burning occurs in a high- or low-oxygen environment, and the actual heat of combustion. Some 280 toxic products have been identified in wood smoke. Petrochemical science has produced a wealth of plastic materials in homes and automobiles that, when burned, produce nearly all of these and many other products not yet characterized. Prominent by-products of incomplete combustion are oxides of sulfur, nitrogen, and many aldehydes. One aldehyde, acrolein, causes severe pulmonary edema in concentrations as low as 10 ppm.

After inhalation, the highly soluble organic aldehydes produced by combustion and pyrolysis (e.g., formaldehyde, acetaldehyde, acrolein) rapidly dissolve in water lining the mucosa of the upper and lower airways, causing direct epithelial injury. The result is epithelial necrosis, edema, and submucosal hemorrhage, primarily in the lower airways. Ammonia is also highly soluble, and reacts with respiratory tract water to form ammonium hydroxide, a strong alkali. Inhaled sulfur dioxide becomes similarly hydrated and then oxidizes to form sulfurous and sulfuric acid. These caustic acids (including hydrogen chloride) and bases cause significant mucosal coagulation and liquefaction necrosis, producing an injury pattern similar to that of the aldehydes.

Smoke inhalation can cause direct epithelial damage at all levels of the respiratory tract, from oropharynx to alveolus. The anatomic level at which the damage occurs is dependent on the ventilatory pattern, the smoke constituents (e.g., particulate concentration, particulate size, and chemical components), and the anatomic distribution of particulate deposition. Although the chemical mechanisms of injury may be different with different toxic products, the overall end-organ response is reasonably well defined. There is an immediate loss of bronchial epithelial cilia and decreased alveolar surfactant. Microatelectasis, and sometimes macroatelectasis, results and is compounded by mucosal edema in small airways, with immediate development of atelectasis that is only slowly reversible by normal ventilation. The regional hypoventilation results in significant alveolar atelectasis, intrapulmonary shunt, and subsequent hypoxemia. Chemical irritation of the respiratory tract, particularly the upper and lower airways, causes an acute inflammatory response.

The initial response is an approximately tenfold increase in bronchial blood flow. Concurrent with this is the stimulation of alveolar macrophages, the release of chemotactic factors, the activation of circulating neutrophils that localize to the site of injury, and the release of oxygen radicals and tissue proteases resulting in changes in vascular permeability. Development of airway edema, combined with sloughing of necrotic epithelial mucosa and impairment of mucociliary clearance of secretions, produces airway obstruction in small and large airways, and the result is ventilatory inhomogeneity. The mismatch between ventilation and perfusion leads to hypoxemia.

Wheezing and air hunger are common early symptoms of smoke inhalation. In a few hours, tracheal and bronchial epithelium begins to slough, and a hemorrhagic tracheobronchitis develops. The pulmonary parenchymal injury appears to be dose-dependent. Alveolar macrophages are activated and attract other white blood cells to the pulmonary endothelium where their release of inflammatory mediators and oxygen free radicals further disrupts the endothelial-epithelial barrier. Pulmonary lymph flow increases with an increased protein content after the double insult of smoke inhalation and cutaneous burns. The severe impairment of chemotaxis of pulmonary alveolar macrophages undoubtedly contributes to the high incidence of late pneumonia seen in patients with associated cutaneous burns.

In very severe cases, the hemorrhagic tracheobronchitis and small airway plugging result in severe ventilatory difficulty during the first 48 h and patients succumb to a severe respiratory acidosis because of their inability to clear CO₂. In moderately severe cases with associated extensive burns interstitial edema becomes prominent, resulting in *adult respiratory distress syndrome (ARDS)*, with difficulty in oxygenation.

Concomitant cutaneous burn injury results in the systemic release of inflammatory mediators, including prostaglandins and oxidants that can aggravate pulmonary injury independent of smoke inhalation. Thromboxane A₂ released from burned tissue causes a variety of changes in the lung, including pulmonary hypertension, reduced dynamic compliance, and increased lipid peroxidation. Oxidants generated as a consequence of neutrophil activation and increases in xanthine oxidase contribute to lung injury. Decreased plasma oncotic pressure from the loss of plasma protein through increasingly permeable vessels in both burned and unburned tissue creates an abnormal oncotic pressure gradient in the lung that, when combined with pulmonary hypertension, results in transient hydrostatic pulmonary edema. These changes help explain the degree of comorbidity in cases of combined inhalation and burn injuries.

Early in the course of smoke poisoning pulmonary function is variable. Typically, decreased lung volume (*functional residual capacity [FRC]*), decreased vital capacity, and evidence of obstructive disease with reduction in flow rates, an increase in dead space, and a rapid decrease in compliance occur. Much of the variability in pulmonary response appears to be related more to the severity of the associated cutaneous burn than to the degree of smoke inhalation. Without associated cutaneous burns, the mortality from smoke poisoning is very low, the disease rarely progresses to ARDS, and symptomatic treatment usually leads to complete resolution of symptoms in a few days. In the presence of burns, smoke poisoning approximately doubles the mortality from burns of any size. Pulmonary symptoms are usually present on admission, but they may be delayed for 12 to 24h. The earlier the onset, the more severe the disease.

Diagnosis

The incidence of smoke inhalation injury in victims of fire varies with diagnostic criteria. The incidence may be as low as 2 to 15 percent when single or restrictive criteria based on history and physical examination are used, but as high as 20 to 30 percent when based on tests such as fiberoptic bronchoscopy. The overall incidence of smoke inhalation in the United States has fallen in the past two decades, primarily because of the use of home smoke detectors.

Anyone with a flame burn and anyone burned in an enclosed space should be assumed to have smoke poisoning until proved otherwise. The acrid smell of smoke on the victim's clothes should raise suspicion. In obtaining a history, emphasis should be placed on data specific to the smoke exposure and to the type of therapy instituted prior to hospitalization. When exposure occurs in a closed space, such as a building or an automobile, the smoke is less diluted by ambient air, resulting in greater pulmonary exposure to carbon monoxide and smoke constituents than in an open-space exposure. The duration of exposure correlates with the severity of lung injury.

An examination should be performed, including evaluation of the face and oropharyngeal airway (edema, stridor, or soot impaction suggesting smoke inhalation), chest auscultation (wheezing or rhonchi suggesting injury to lower airways), level of consciousness (decreased with hypoxemia, carbon monoxide poisoning, or cyanide poisoning), and testing for the presence of specific neurologic defects that might be associated with carbon monoxide. A careful inspection of the mouth and pharynx should be done early. Hoarseness and expiratory wheezes are signs of potentially serious airway edema or smoke poisoning. Copious mucus production and carbonaceous sputum are signs of injury, but their absence does not indicate that injury is absent. Carboxyhemoglobin levels should be obtained; elevated carboxyhemoglobin levels or any symptoms of carbon monoxide poisoning are presumptive evidence of associated smoke poisoning.

Anyone suspected of smoke poisoning should have a set of arterial blood gases drawn. One of the earliest indicators is a falling P/F ratio, the ratio of arterial P_{O_2} to the percentage of inspired oxygen (FiO_2). A ratio of about 400 is normal; patients with impending pulmonary problems have a ratio of less than 350 (e.g., an arterial P_{O_2} of less than 140 with an FiO_2 of 0.40). A ratio of less than 250 is an indication for vigorous pulmonary therapy, not an indication for increasing the inspired oxygen concentration.

The early need for bronchoscopy remains controversial. Some recommend the routine use of fiberoptic bronchoscopy, noting that it is inexpensive, quickly performed in experienced hands, and useful in assessing edema of the upper airway. Aside from documenting the presence of tracheal erythema and carbon deposits, however, it does not materially influence the treatment for smoke poisoning. To determine the role of this examination in patients with smoke inhalation, fiberoptic bronchoscopy was performed in 100 consecutive patients admitted to a regional burn unit presenting with at least one warning sign of inhalation injury (closed-space smoke exposure, facial burn, singed nasal vibrissae, perioral burn, pharyngeal edema, hoarseness, carbonaceous sputum, bronchorrhea, or wheezing). A 96 percent correlation was found between positive bronchoscopic findings and the triad of closed-space fire, COHb levels >10 percent, and carbonaceous sputum. If two items of the triad were present, the correlation dropped to 70 percent, and if only one was present, the correlation was less than 30 percent. No other positive correlations were detected. Upper airway edema could best be correlated with an explosion (flash burn) that involved both the face and the upper torso. Nearly 50 percent of these patients had significant upper airway edema and underwent prophylactic airway intubation. Based on these observations and the limitations of fiberoptic bronchoscopy, it is recommended that a history, clinical examination, and laboratory studies be used to make the diagnosis of inhalation injury, and the use of fiberoptic bronchoscopy be reserved for exceptional cases (e.g., expansion of lobar atelectasis or removal of obstructing intrabronchial secretions).

Two clinical studies of burn patients with clinical signs of smoke inhalation found poor correlation of bronchoscopic findings with the need for ventilatory support or the subsequent development of ARDS, concluding that immediate bronchoscopy neither indicates the level of respiratory support that will be required nor predicts its duration. Flow volume loops may be useful, but are complex to perform and difficult to interpret. The simple admission chest x-ray is notoriously insensitive in detecting severely injured lung early after smoke exposure, with false-negative rates as high as 92 percent. Computed tomography (CT) may be useful in demonstrating early atelectasis and bronchial edema, but it is expensive and delays treatment, and, like bronchoscopy, has no practical implications for therapy.

Treatment

Upper Airway

No standard treatment has evolved to ensure survival after smoke poisoning. In the presence of increasing laryngeal edema, nasotracheal or orotracheal intubation is indicated. A tracheostomy is never an emergency procedure and should not be used as the initial step in airway management in patients with burns to the face and neck. Instead, a soft-cuffed endotracheal tube should be placed, and left in place for about 72 h or until the generalized oropharyngeal edema subsides. An adult patient's ability to breathe around the tube with the cuff deflated is an indication for removal of the tube. This assessment is difficult in children due to their smaller anatomy, the use of uncuffed endotracheal tubes, the increased incidence of postextubation stridor, and the frequent need for reintubation. The incidence of postextubation stridor in burn victims is as high as 47 percent, compared to 4 percent in elective surgical patients. The treatment of postextubation stridor includes the administration of racemic epinephrine and helium-oxygen (Heliox) mixtures.

Lower Airway and Alveolar Damage

Tracheobronchitis, commonly seen in smoke and toxic gas inhalation victims, produces wheezing, coughing, and retained secretions. The ventilation-perfusion mismatch present in these patients can result in mild to moderate hypoxemia, depending on the degree of underlying lung disease; therefore, supplemental oxygen should be administered routinely. Increased airway resistance is more often the result of decreased airway caliber (from mucosal and/or submucosal edema and retained secretions) than true bronchospasm. Although a trial of bronchodilators is indicated in those with preexisting bronchospastic disease, its efficacy is questionable. Nebulized or metered dose therapy with β_2 agonists or racemic epinephrine, subcutaneous terbutaline, or intravenous aminophylline is used most commonly.

The presenting sign of lower airway damage is hypoxemia, diagnosed by pulse oximetry or, preferably, arterial blood gas analysis. Because most inhalation victims receive supplemental oxygen, significant alveolar-arterial oxygen gradients can be missed by pulse oximetry; saturations less than 95 percent do not occur until arterial P_{O_2} is less than 80 mmHg. Arterial blood gas analysis is the monitor of choice in assessing oxygenation after inhalation injury. Initial therapy should always include the administration of high-flow oxygen, to supplement oxygenation and to reduce carboxyhemoglobin in cases of carbon monoxide inhalation. Upper airway patency must be assured, and airway resistance minimized with chest physiotherapy and/or bronchodilators. Central hypoventilation caused by carbon monoxide or cyanide poisoning should be treated immediately with endotracheal intubation and assisted ventilation, and efforts to reverse intoxication.

Treatment for smoke poisoning is supportive, with the goal of maintaining adequate ventilation and oxygenation until the lung heals itself. Mild cases of smoke poisoning are treated with highly humidified air, vigorous pulmonary toilet, and bronchodilators as needed.

The need for mechanical ventilation to supplement oxygenation is determined by repeated blood gas measurements. As a guideline, the Pa_{O_2}/FiO_2 (P/F) ratio may be calculated without mixed venous blood sampling and used as an approximation of the shunt fraction. A P/F ratio between 200 and 400 indicates mild to moderate injury, usually requiring only supplemental oxygen therapy. A P/F ratio below 200 is evidence of serious parenchymal lung injury, and usually indicates a need for intubation and ventilation with high inspired oxygen fractions or the use of *positive end-expiratory pressure (PEEP)*. Because of the frequent presence of atelectasis after alveolar exposure to smoke, PEEP can be useful, provided pressures are not excessive, subsequent to the initial use of large tidal volumes (10 to 15 mL/kg) initially in a standard volume-control ventilator setting. The tidal volume should be monitored by measuring total respiratory compliance over a range of tidal volumes, with avoidance of those volumes associated with a reduction in compliance (and increased risk of lung rupture).

High-frequency percussive ventilation has been attempted in patients with inhalation injury. This mode of ventilation provides adequate oxygenation at a lower FiO_2 , with lower peak and mean airway pressures. The benefit to inhalation injury is the apparently enhanced clearance of bronchial secretions achieved with this technique. Initial reports suggest potential superiority to standard ventilation techniques in patients with smoke inhalation. Mechanical causes for poor ventilation include restricted chest-wall motion from full-thickness burns, pneumothorax from high ventilator pressures, and mechanical difficulties with the endotracheal tube.

Establishing a tracheostomy in burn patients is controversial. If the upper airway is in danger of imminent obstruction and endotracheal intubation attempts are unsuccessful, emergent cricothyroidotomy is indicated. The indications for nonemergent tracheotomy have changed. After a period in the 1970s when tracheostomy was the standard method of securing the upper airway after severe burn injury, several reports associated the procedure with mortality rates ranging from 52 to 100 percent due to a greater incidence of overwhelming pulmonary infection. Improvement in endotracheal tube construction resulted in specific efforts to avoid tracheotomy in burn patients in the 1980s. Currently, mortality rates for burn patients with tracheostomy are not different from those for patients treated with endotracheal tubes, despite the fact that the former group is more likely to include patients who are burned so severely as to preclude survival. For patients requiring prolonged endotracheal intubation, tracheostomy should be performed between 3 and 30 days after intubation. Patients with anterior neck burns who require tracheostomy should undergo excision and grafting of the area 5 to 7 days prior to creation of the tracheostomy. This minimizes pulmonary and burn wound infectious complications associated with the tracheostomy.

Prophylactic antibiotics are not valuable in burn-related chemical pneumonitis, and subsequent burn management and treatment of eventual bacterial pneumonia can be made more difficult if the early use of antibiotics leads to the selection of resistant organisms. Steroids are commonly used in patients with severe asthma. There has been a tendency when dealing with smoke poisoning to use steroids for their spasmolytic and anti-inflammatory action. In a prospective blind study of patients with smoke poisoning and associated major burns, Moylan and Chan demonstrated that mortality and infectious complications were higher in the patients treated with steroids. Robinson found that steroids did not alter the hospital course of patients without associated burns admitted to hospital after the MGM Grand and Hilton hotel fires in 1981. A trial of genetically engineered surfactant in patients with smoke poisoning was unsuccessful.

The decision on whether to admit a patient to the hospital, as well as the need for specialized care, depends on the severity of symptoms from the smoke and the presence of associated burns. A patient who has symptoms associated with smoke inhalation and who has more than trivial burns should be admitted. If the burns are greater than 15 percent TBSA, the patient should be referred to a burn center intensive care unit. In the absence of burns, admission depends on the severity of symptoms, the presence of preexisting medical problems, and the social circumstances of the patient. Otherwise healthy patients with mild symptoms (only a few expiratory wheezes, minimal sputum production, CO level <10, and normal blood gases) who have a place to go and someone to stay with them can usually be watched for a few hours and then discharged. Patients with preexisting cardiovascular or pulmonary disease should be admitted for observation if they have symptoms related to the smoke. Patients with moderate symptoms (generalized wheezing, mild hoarseness, moderate sputum, CO levels of 5 to 10 percent) and normal blood gases can be admitted for close observation and treatment. Severe symptoms (air hunger, severe wheezing, copious, usually carbonaceous, sputum), require intubation, possible therapeutic bronchoscopy, ventilator support, and placement in an intensive care unit.

WOUND MANAGEMENT

Early Excision and Grafting

For many years, burns were treated by daily washing, removal of loose dead tissue, and topical application of saline-soaked dressings until they healed by themselves or granulation tissue appeared in the base of the wound. Superficial dermal burns healed within 2 weeks and deep dermal burns healed over many weeks if infection was prevented. Full-thickness burns lost their eschar in 2 to 6 weeks through bacterial collagenase production and daily mechanical debridement. When the granulating bed became free of debris and relatively uninfected, split-thickness skin grafts were applied, usually some 3 to 8 weeks after injury, and a 50 percent graft take was considered acceptable. Repeated graftings eventually closed the wound. The prolonged and intense inflammatory response made hypertrophic scar and contractures part of normal burn treatment. Vigorous physical therapy, nutritional support, psychological support, and pain management were required on a daily basis for many weeks to yield a satisfactory result.

This is no longer standard procedure. For deeper burns, rather than waiting for spontaneous separation, the eschar is surgically removed and the wound closed, with grafting techniques and procedures for immediate placement of flaps tailored to meet patients' individual needs. Several technical advances have made this possible. There is "safer" blood, better monitoring equipment and methods, and a better understanding of the altered physiology and increased metabolic demands of patients with major burns. The ability to stabilize the patient within a few days of the injury has enabled the surgeon to remove deep burn wounds before invasive infection occurs. An aggressive surgical approach to large and small burns has produced a number of advantages. Early wound closure shortens hospital stay and duration of illness. Early studies did not demonstrate dramatic differences in cosmetic and functional results, but as surgeons have become more experienced, both improved function and appearance have resulted. This is particularly true with burns of the face, hands, and feet.

Historical Perspective

The surgical principle that "clean wounds should be closed" has been applied since the days of Hippocrates. For technical reasons, burns have been the exception to this rule. Early in the twentieth century, primary burn wound closure was attempted in patients with major burns, but systemic instability, massive hemorrhage, graft loss, malnutrition, and infection resulted in such high mortality rates that major burn excisions were abandoned. Because of their relatively limited extent, electrical burns remained an exception. Cope was the first to report a series of 58 primary excision and graftings in 38 patients after a 1947 nightclub fire. From 1961 to 1974, Janzekovic treated 2615 of 4370 burned patients with early tangential (or sequential) excision and sheet autografting; most of the burns were small (under 20 percent TBSA). The hospital stay dramatically decreased, pain was less, need for

reconstructive procedures decreased, and “aesthetic disability” was greatly reduced. The early 1980s produced more clinical and experimental data suggesting that prompt wound closure produced better results and improved metabolic response. It is more difficult to prove that early excision improves mortality in patients with extensive burns (greater than 60 percent TBSA). Until a useful “artificial skin” is available, the goal of early complete wound closure in patients with massive burns cannot be achieved. In these patients, donor sites are limited and excised burns must be covered with temporary biologic dressings until the donor sites heal and can be recropped. Recent FDA approval of a permanent “off the shelf” dermal substitute made of collagen matrix combined with a glycosaminoglycan (chondroitin 6-sulfate) that acts as a template for endogenous cells to reproduce a new dermis is a major step in the development of a permanent skin substitute.

Elderly patients present unique problems. Continued stress from the burn injury results in a high incidence of cardiac and cerebrovascular catastrophes. The atrophic skin of elderly patients presents problems in burn wound and donor site healing. No substantial decrease in mortality for elderly patients has occurred. The mortality rate for patients over 65 years of age with burns less than or equal to 20 percent TBSA remains at approximately 10 percent. This increases logarithmically as the size of the burns increases, and there are few elderly survivors with burns over 50 percent TBSA.

More burn centers are practicing early excision and grafting. When experience has made these procedures routine, early excision and grafting will become the treatment of choice for all deep dermal and full-thickness burns. The procedure is still limited by difficulty in diagnosing burn depth, by limited donor sites, and the difficulties involved in excision of three- dimensional areas, such as the perineum, ears, and nose.

Current Status of Wound Care

Evidence supports the following conclusions:

- (1) Small (less than 20 percent) full-thickness burns, and burns of indeterminate depth (deep partial versus full thickness), if treated by an experienced surgeon, can be safely excised and grafted with a decrease in hospital stay, cost to the patient, and time away from work or school.
- (2) Early excision and grafting dramatically decreases the number of painful debridements required by all patients.
- (3) Patients with burns between 20 and 40 percent TBSA will have fewer infectious wound complications if treated with early excision and grafting.
- (4) In animals with experimental burns, the depressed immune response and hypermetabolism associated with burns can be ameliorated by early burn wound removal.

Clinical impressions without hard data supporting them include the following:

- (1) Scarring is less severe in wounds closed early, leading to better appearance and fewer reconstructive procedures. There is no good measure of acceptable “cosmetic appearance ” and comparative studies await an acceptable scale to measure results.
- (2) Mortality from wound infection is lower in patients with major burns after early excision. Because wounds exceeding the donor sites cannot permanently be closed completely until donor sites can be reharvested, proof will come only when a durable permanent cover can be applied in a timely fashion.
- (3) Mortality from other complications of major burns may be lower with early excision and grafting. Decreasing stress, hypermetabolism, and the overall bacterial load of the patients enables them to resist other complications. The only data to support this conclusion come from animal studies.

Technical Considerations

Excision of more than 10 percent of the total body surface should be done in a highly structured environment. Without tourniquets, blood loss can be massive. Graft loss can be catastrophic. Excellent monitoring, nursing, physical therapy, nutritional support, anesthesia, and 24-h physician coverage are mandatory. Smaller burns in important areas (hands, face, and feet) also require considerable experience.

Excisional procedures should be performed as early as possible after the patient is stabilized. This allows the wound to be closed before infection occurs and, in extensive burns, allows donor sites to be recropped as soon as possible. Cosmetic results are better if the wound can be excised and grafted before the intense inflammatory response associated with burns becomes well established. Any burn projected to take longer than 3 weeks to heal is a candidate for excision within the first postburn week. Wound excision is adaptable to all age groups, but infants, small children, and elderly patients require close perioperative monitoring.

Excision can be performed to include the burn and subcutaneous fat to the level of the investing fascia (fascial excision), or by sequentially removing thin slices of burned tissue until a viable bed remains (sequential excision). Fascial excision assures a viable bed for grafting, but takes longer, sacrifices potentially viable fat and lymphatics, and leaves a permanent cosmetic defect. Sequential excision can create massive blood loss and risks grafting on a bed of uncertain viability, but sacrifices minimal living tissue and leads to a far superior cosmetic result than fascial excision. Current practice reserves fascial excision for patients with fourth-degree burns and patients with such massive burns that they can afford no graft loss.

Tangential (Sequential) Excision

The principle of tangential excision is to shave very thin layers of burn eschar sequentially until viable tissue is reached. The burn can be removed with a variety of instruments, usually power- or hand-driven dermatomes. Relatively shallow burns and some burns of moderate depth will bleed briskly from thousands of capillaries after one slice. If the bed does not bleed briskly, another slice of the same depth is taken until a viable bed of dermis or subcutaneous fat is reached. If inspection of the dermal or fatty bed reveals a surface that appears gray or dull rather than white and shiny, or if there is evidence of clotted vessels, the excision should be carried deeper. Any fat that has a brownish discoloration, has blood staining, or contains clotted blood vessels will not support a skin graft and must be excised until the bed contains uniformly yellow fat with briskly bleeding vessels. Bleeding is controlled with sponges soaked in 1:10,000 epinephrine solution applied to the excision bed for 10 min. Continued bleeding is then controlled with an electrocautery. Major bleeding is rare; when bleeding has occurred, it has invariably been associated with inadequate cauterization of a vessel with pulsatile flow.

Areas on the extremities may be excised using a tourniquet. The cadaver- like appearance of the dermis and the lack of brisk bleeding can easily mislead the surgeon into sacrificing normal tissue by carrying the excision deeper than necessary.

Fascial Excision

Fascial excision is reserved for patients with very deep burns (charred flame burns, prolonged contact burns, molten metal burns, and electrical burns), or for patients with very large, life-threatening, full-thickness burns. The most common technique uses an electrocautery with cutting and coagulating capabilities.

The advantages of fascial excision include:

- (1) It results in a reliable bed of known viability.
- (2) Tourniquets can be routinely used for extremities.
- (3) Operative blood loss is less than with sequential excision.
- (4) Less experience is required to ensure an optimal bed.

The disadvantages include:

- (1) The operative time is longer.
- (2) There may be severe cosmetic deformity, especially in obese patients.
- (3) There is a higher incidence of distal edema when excision is circumferential.
- (4) There is greater danger of damage to superficial nerves and tendons.
- (5) Cutaneous denervation, which may or may not be permanent, may occur.
- (6) Skin graft loss from the relatively vascular fascia over joints (elbow, knee, ankle) can lead to an ungraftable bed and require eventual flap coverage.

Early Reconstruction

A potential advantage to excision and grafting is to provide a closed wound before the intense inflammatory response begins. If careful attention is given to sound principles of plastic surgery, the risk that there will be a need for subsequent reconstruction can be decreased. Skin graft junctures should be avoided over joints, and grafts should be placed transversely when possible. Thick skin grafts yield a better appearance than thin skin grafts. If the burn is well excised, and the skin can be spared, thick skin grafts should be used on the face, neck, and other cosmetically important areas. The resultant donor sites can be overgrafted with thin skin grafts to minimize hypertrophic scarring of the donor site. Whenever possible, cosmetically important areas should be grafted with sheet skin grafts. Although meshed skin grafts provide cover with excellent function, the meshed pattern persists as a permanent reminder of the burn.

Adjacent pieces of skin graft should be approximated carefully. While staples are adequate for areas in which cosmetics is not an issue, for critical areas, such as the face, suturing

the edges together is preferred. If the wound can be left open or dressed with a dry bandage after sheet skin grafting, steri-strips can be used effectively. They will not remain in place if the wound is covered with moist dressings.

Keeping these early reconstructive principles in mind during the first operations may avert the need for later procedures entirely, and will help convert what could be major reconstructive efforts into minor ones.

Donor Sites

In previous years, when only full-thickness burns were skin-grafted, and patients endured many weeks of daily debridement, donor sites were treated superficially. They were covered with either dry fine mesh gauze, or gauze impregnated with a dye or other antimicrobial agent. They were left to desiccate, and the gauze usually separated from the wound in 2 to 3 weeks, sometimes removing substantial areas of new epithelium. As aggressive programs of early excision and grafting developed, donor sites became a priority. With early excision, the patient was spared the painful daily debridement, and with burn pain diminished, patients concentrated on donor site pain.

There are hundreds of dressings available for donor site and after-grafting wound care. Reports indicate that there is no optimal donor dressing. All dressings seem to work, and differences in healing times are only 1 or 2 days. Comfort levels and ease of care are the most significant determinants.

Healed donor sites are still not free of complications. In addition to hypertrophic scarring and changed pigmentation, patients may be troubled by blistering for several weeks. Blisters are self-limiting and are usually treated with bandages or ointments until they reepithelialize. Infections occur in about 5 percent of patients. Infection is treated with systemic antibiotics and continuously moist dressings or silver sulfadiazine.

SKIN SUBSTITUTES

The next major step in burn care is likely to be an artificial skin that will be readily available, perform barrier function (epidermis), and provide the structural durability and flexibility of the dermis. It must be permanent, affordable, not susceptible to hypertrophic scarring, provide normal pigmentation, and grow with developing children. Progress toward this goal has been substantial over the past decade.

Cultured Epidermal Autograft

In the earliest grafting procedures, successful in vitro culturing of epidermal cells (keratinocytes) produced a permanent skin cover whereby small bits of unburned skin could provide seed keratinocytes that could be grown to form sheets of cells and grafted onto a burn wound bed, closing massive wounds when donor sites were limited. The first successful grafting was reported in children in 1986. In 3 weeks' time, an initial skin biopsy could be enlarged 1000 times, producing hundreds of 5- by 5-cm sheets of epidermis, three to eight cell layers thick. Placing these fragile sheets on an excised burn wound could provide a permanent wound cover. Considerable experience has been gained with *cultured epidermal autografts (CEAs)*. Areas of concern and controversy remain with respect to efficacy of take, durability, necessity, and cost-effectiveness.

Variability in take may be related to the bed on which the CEAs are applied, with native or allogenic dermis (from cadaver allografts) being the most successful. Bacterial contamination and colonization of the grafted bed results in rapid CEA disappearance from the site.

Engrafted CEAs are poorly adherent and extremely fragile for months after application. Care must be taken to avoid blistering and graft loss with movement, rubbing, scratching, and physical therapy because of the lack of a dermis in excisions that reach fat or fascia. The histologic events leading to CEA maturation also appear inconsistent. One study reported that within 6 days "flat epidermis with all normal strata had regenerated, and the process of de novo dermal-epidermal junction formation had begun" and "within three to four weeks, the dermal-epidermal junction was complete, but full maturation of anchoring fibrils required more than a year." Compton reported that "the subjacent connective tissue initially healed to form normal scar, but it remodeled dramatically, regenerated elastin, and resembled a true dermis within four to five years." Putland, however, indicated that "the CEA interface with underlying bed remained flat for up to three years in three of four patients. CEA epidermal rete ridges, if formed subsequently, were fewer, thinner, and shorter, whereas expanded split-thickness skin grafts had well-defined rete ridges after one year." It was concluded that the persistent fragility of CEAs is related to the delayed formation of rete ridges.

When the patient is admitted, a 1-cm² skin biopsy specimen is usually sent to a commercial laboratory for culturing. Three weeks later 5- by 5-cm² sheets of cultured cells are delivered. CEAs are expensive. Rue reported that the average cost of CEAs covering only 4.7 percent TBSA was \$43,000.

The following conclusions can be drawn:

- (1) Cultured keratinocytes can be grown within 3 weeks and can be grafted successfully on a viable, noninfected excised burn wound bed. For a massive burn with few or no available donor sites, this may be life-saving, especially in children.
- (2) Take varies from poor to fair, with an optimistic range of 30 to 40 percent.
- (3) Because epidermal cells lack a dermis, in early stages of wound healing, CEAs can provide only a barrier function to prevent fluid exudation and bacterial invasion.
- (4) The cover is fragile and must be protected from mechanical disruption for months, limiting daily activities and vigorous physical and occupational therapy.
- (5) The cost is approximately \$6000 to \$10,000 for each 1 percent TBSA covered.
- (6) CEAs are a temporary measure, permitting survival in the patients with massive burns. It is unlikely that they can ever be a permanent solution because they provide only one-half of the necessary bilaminar constituent of skin.

Dermal Substitutes

Many investigations have looked for an acceptable dermal matrix onto which CEA or thin epidermal grafts could be placed. The only FDA-approved synthetic dermal substitute, a material developed by Burke and Yannis, is a bovine collagen matrix with fiber size and distance similar to that of dermis, with a ground substance, chondroitin 6-sulfate, filling the pores. It provides a template on which native fibroblasts, endothelial cells, and macrophages can replace the collagen with a dermal matrix resembling dermis more than disorganized scar. In a multicenter trial to determine whether results with ultrathin epidermal grafts over this material could provide cover equal to conventional meshed grafts, the results were excellent. The material is now in use, offering better cosmetic and functional results using ultra-thin epidermal grafts so that donor sites can be recropped frequently in the treatment of large burns.

NUTRITIONAL SUPPORT

The nutritional effects of the hypermetabolic response to thermal injury are manifested as exaggerated energy expenditure and massive nitrogen loss. Nutritional support is directed primarily toward provision of calories to match energy expenditure and provision of nitrogen to replace or support body protein stores.

Changes in metabolism are triggered by drastic changes in the hormonal profile, including elevated levels of catecholamines, glucocorticosteroids, and glucagon. These hormones, together with other circulating peptides, such as IL-1, TNF, and probably IL-6, accelerate protein catabolism, gluconeogenesis, and lipolysis. Insulin levels are usually in the normal range or elevated, but they are low in relation to the increased glucagon concentrations in plasma. Catecholamines and glucocorticoids antagonize the action of insulin, the key anabolic hormone that promotes storage of the metabolic fuels within the cells. These metabolic and hormonal consequences have important effects on nutritional status.

Caloric Requirements

Malnutrition in patients who have undergone a surgical procedure was first characterized in terms of negative energy balance. Hypermetabolism and hypercatabolism are universal consequences of injury. The cause of hypermetabolism seems to be dictated by the neurohumoral and cytokine stress response. The magnitude of the increase in metabolic rate is directly proportional to the size of burn injury. The total energy expenditure may be elevated from 15 to 100 percent of basal needs, exceeding those of other injuries and directly proportional to burn size. Energy needs must be evaluated carefully in formulating a parenteral or enteral diet therapy program. If the regimen is deficient in calories, protein synthesis will not be optimal, and nitrogen balance will continue to be negative. Mathematical derivations exist for the calculation of daily caloric needs in burns. The formula most used is the Long's modification of the Harris-Benedict equation (Table [7-3](#)). The Harris-Benedict equation estimates *basal metabolic rate (BMR)* with reasonable accuracy. Long proposed that the BMR be multiplied by various stress factors depending upon type of injury. The current revision for burn patients uses a multiplication factor of 1.3. The more severely ill the patient, the less accurate standard formulas are for estimating calorie expenditure.

Routine determination of *resting energy expenditure (REE)* from the measurement of oxygen consumption and carbon dioxide production is conducted at least twice weekly on burn patients for proper adjustments of caloric needs. REE determinations should not be construed as equivalent to the 24-h calorie requirement. Compensations must be made for daily energy fluctuations that occur with physical therapy, stress, temperature spikes, dressing changes, and other influences on metabolic rate. The patient's calorie goal should be calculated at 120 to 130 percent of the measured REE. Urinary nitrogen excretion is relatively easy to measure, but nitrogen losses in dressings and skin make accurate measurements of nitrogen balance difficult to obtain.

Carbohydrates

Carbohydrates, primarily in the form of glucose, appear to be the best source of nonprotein calories in the thermally injured patient. Certain tissues, including the burn wound, neural tissues, and the formed elements of the blood, utilize glucose in an obligatory fashion. Provision of glucose to these tissues occurs at the expense of lean body mass if adequate nutrition is not provided. In the unaliminated state, the major sources of three carbon precursors for new glucose production by the liver are the wound and skeletal muscle. The wound uses glucose by anaerobic glycolytic pathways, producing large amounts of lactate as an end product. The wound meets its high glucose requirements by means of high glucose delivery rates, which are made possible by the enhanced circulation to the wound.

In the liver, lactate is extracted and utilized for new glucose production by the Cori cycle. Concomitantly, alanine, glutamine, and other glycogenic amino acids contribute to increased gluconeogenesis. Increased ureagenesis, with urea ultimately derived from body protein stores, parallels the rise in hepatic glucose output. Peripheral amino acids and wound lactate account for approximately one-half to two-thirds of new glucose produced by the liver. The mild hyperglycemia observed in hypermetabolic burn patients is a

consequence of accelerated glucose flow arising from increased hepatic glucose production, not from decreased peripheral utilization.

Because glucose that is obtained by gluconeogenic pathways is ultimately derived from protein stores, depletion of body protein during periods of starvation leads to energy deficits and malfunctioning of glucose- dependent energetic processes at the cellular level. Active transport mechanisms responsible for maintaining transmembrane ionic gradients in erythrocytes are deranged in catabolic, thermally injured patients. The abnormal sodium and potassium gradients in red blood cells can be reversed by providing these patients with high caloric levels of carbohydrate as glucose. Hepatic clearance of indocyanine green, an energy-dependent active transport process, is decreased in severely injured patients when energy normally supplied as glucose is replaced by an isocaloric glucose-free source. Glucose-insulin solutions correct the “sick cell syndrome” in burned patients who exhibit a prompt natriuresis and nonosmotic diuresis when metabolic requirements are met by glucose.

Protein

Combining glucose and nitrogen-containing nutrients improves nitrogen balance and allows more calories to be used for the restoration of nitrogen balance than would be the case if either nutrient group were used alone. Energy and protein cooperatively contribute to the improvement in protein conservation. After injury, the individual effects of glucose and amino acids on nitrogen equilibrium operate by at least two different mechanisms. Amino acid administration promotes synthesis of visceral and muscle protein without affecting the rate of protein breakdown. Glucose retards whole- body protein breakdown and decreases the total amino acid pool, but exerts little effect on protein synthesis. Both mechanisms improve nitrogen balance, and glucose and nitrogen should be components of the nutritional regimen for the severely burned catabolic patient.

The unique importance of glutamine as a fuel source has been recognized. The gastrointestinal tract uses glutamine as a respiratory energy source, and disposes of the majority of glutamine as ammonia, urea, and citrulline. The alanine generated from glutamine in the intestine and kidney is used for gluconeocentesis. During critical illness, circulating concentrations of glutamine fall, and supplemental glutamine is required to meet gastrointestinal tract energy requirements. While glutamine is easily supplied by the enteral route, and all agree on its efficacy, parenteral preparations are not routinely available. There is controversy as to whether parenterally administered glutamine is efficacious.

Administration of arginine after injury has become increasingly important. Increased dietary arginine may diminish protein catabolism, reducing urinary nitrogen secretion in trauma or stress and improving immune function. Another beneficial effect of arginine is its secretagogue activity on pituitary and pancreatic hormones. Dietary arginine supplementation was shown to increase plasma insulin levels after glucose administration in protein-depleted rats. Tube feedings designed to enhance immune function have demonstrated the beneficial effects of arginine on anabolic hormone secretion.

Fat

The role of fat as a source of nonprotein calories is dependent on the extent of injury and the associated hypermetabolic response. When hypercaloric diets that do not contain nitrogen are administered, carbohydrate alone is more effective in sparing body protein than is fat alone. Fat appears to be a poor calorie source for the maintenance of nitrogen equilibrium and lean body mass in hypermetabolic patients with large burns. Patients with only moderate elevations of metabolic rate can use lipid calories efficiently, but these patients rarely require parenteral nutrition; most table foods and defined diets contain all necessary fat nutrients.

Linoleic acid was established as “essential” in the 1930s. It cannot be synthesized and has a specific role in maintaining cellular integrity. Consumption of 1 to 3 percent of total calories as linoleic acid is sufficient to prevent deficiency in humans. Studies suggest that fatty acids of the omega-3 family (particularly a-linolenic and eicosapentaenoic acid) are also important in dietary constituents. Eicosapentaenoic acid is the primary source of the triene prostaglandins. Omega-3 fatty acids form a series of compounds similar to the omega-6 series, but have different biologic effects. These effects are less catabolic and less injurious than prostaglandins derived from the omega-6 fatty acids.

Not only the amount of total fat, but also the structure of the fatty acids used in nutritional support programs, can have divergent effects on metabolism, morbidity, and mortality. Enteral tube feeding products are largely devoid of omega-3 fatty acids; a-linolenic deficiency has been demonstrated in patients on long-term tube feedings.

Vitamins and Minerals

Vitamin requirements in critically ill hypermetabolic burn patients remain poorly defined. The fat-soluble vitamins (A, D, E, and K) are stored in fat depots and are slowly depleted during prolonged feeding of solutions that do not contain any vitamin formulations. The water-soluble vitamins (B- complex and C) are not stored in appreciable amounts, and are depleted rapidly. All vitamins should be supplemented. The dosage guidelines recommended by the *National Advisory Group/American Medical Association (NAG/AMA)* are reasonable for burn patients unless symptoms of deficiency occur. Ascorbic acid has an essential role in wound repair, and plasma levels are frequently depressed in burn patients. It is prudent to supplement the NAG/AMA formulation with 250 to 500 mg of vitamin C daily. Larger doses may cause diarrhea and formation of renal stones, and will interfere with laboratory studies. Excessive doses of vitamins A and D produce toxic symptoms, and monitoring of serum levels in critically ill patients is often misleading, since the concentrations of the vitamin carrier proteins are commonly decreased in these patients.

Mineral nutrients are important because of their role in metabolic processes. Frequent determinations of serum sodium, potassium, chloride, calcium, magnesium, and phosphorus are the best guides to electrolyte replacement. Less is known about trace metal requirements after thermal injury. Zinc is an important cofactor in enzymatic function and wound repair, and zinc deficiency has been documented in burn patients. After injury in animal models, zinc and other trace metals seem necessary for nitrogen retention, but the level of these metals may merely reflect nitrogen balance and have little direct implication. Periodic measurements of zinc, copper, manganese, and chromium are the best way to determine replacement-dosage guidelines. Trace elements are present in varying concentrations as contaminants in amino acid parenteral solutions, and contribute to satisfying daily requirements.

Route of Administration

The route of administration of nutrients is important because it seems to influence outcome. Total parenteral nutrition is used only when the patient's needs cannot be wholly met by the enteral route.

Patients with burns under 25 percent TBSA that are not complicated by facial injury, inhalation injury, or malnutrition, and are not associated with psychological difficulties, including possible abuse, can usually be maintained on high-calorie, high-protein diets ingested orally. The nutritional requirements of patients with large burns cannot be met by the oral route alone, and these patients should be fed gastrointestinally or nasointerally. A functionally intact alimentary tract always should be used.

In severely burned patients, gastric ileus may limit the stomach's role in nutritional support, at least in the early postburn phase, but the small bowel usually maintains normal mobility and absorption. The safest route for infusion of nutrients is distal to the ligament of Treitz. The placement of a small bowel feeding tube, during resuscitation or during surgical treatment, can be the first step in providing nutritional support. A protective nasogastric tube is used to measure gastric residual content. Enteral feeding has advantages over parenteral feedings. Enteral nutrients seem to maintain the integrity of the gastrointestinal tract, and increased hepatic protein synthesis may reduce the incidence of bacterial translocation from the gut. An oral diet preserves gut mucosal mass and maintains digestive enzyme content; parenteral feeding results in decreased mucosal cell turnover. Studies have verified that oral feeding stimulates the gut to elaborate trophic hormones, particularly gastrin. Enteral calories initiate greater insulin release than parenteral nutrition, and insulin appears to promote anabolism. Studies have shown that institution of enteral feeding immediately after admission of patients with burn trauma is beneficial. This feeding technique blunts the intensity of the hypercatabolic response and more effectively maintains preinjury weight. Associated findings are decreased circulating concentrations of the counterregulatory hormones epinephrine, glucagon, and corticosteroids.

Total parenteral nutrition should be instituted when enteral feedings alone cannot provide adequate nutritional support. Prolonged postresuscitation ileus, overuse of narcotics, and constipation are frequently causes of failure of successful enteral alimentation. Sepsis is associated with ileus and severe glucose intolerance, and these symptoms may be the only evidence of this complication. Previously tolerated feedings must often be discontinued while hyperglycemia is being controlled and the patient is resuscitated. Ileus commonly persists, and nutrition is reinstituted by the parenteral route, often requiring large doses of insulin. If total support cannot be supplied enterally, it is useful to continue enteral nutrition even if the majority of nutrition is supplied parenterally. The benefits of enteral nutrition are realized even when 20 percent of needs are given enterally.

Composition of Enteral Nutrition

Three decades ago, burn and trauma patients requiring nutritional support beyond that provided by a hospital diet received tube feedings consisting of food solutions prepared in a blender. These blended meals have been standardized, and are available in commercially prepared formulas. The goal is to provide nutritional support tailored to meet the needs of critically ill burn patients. Nutritional support should provide substrates in proportions based on the specific metabolic derangements in burns. Standard meal replacement products formulated for nonstressed or minimally stressed patients do not meet the unique nutritional requirements of moderately or severely metabolically stressed patients. Because these regimens can be grossly incompatible with needs, they are often counterproductive, exacerbating nutritional inadequacies and increasing the chance of hepatic, pulmonary, and gastrointestinal tract complications, as well as other metabolic dysfunction. Studies of patients with renal failure, pulmonary insufficiency, hepatic encephalopathy, and burns indicate that specialized feeding regimens improve disease-related metabolic derangements while enhancing nutritional status.

In burn patients, significant amounts of omega-3 fatty acids should be used. The modular tube feeding at the Shriners Burns Institute is a high- protein (20 percent of calories), low-fat (15 percent of nonprotein kilocalories [kcal]) nucleic acid–restricted formulation (only enough to supply essential fatty acid), which is enriched with omega-3 fatty acid (half of lipid calories), arginine (2 percent of kcal), cysteine (0.5 percent of kcal), histidine (0.5 percent of kcal), vitamin A (5000 IU/L), zinc sulfate (220 mg/day), and ascorbic acid (1 g/day). The composition is structured to improve immune function, optimize wound healing, and lower production of the proteolytic and immunosuppressive dienoic prostaglandins. Commercially available immune function–enhancing enteral feedings have been shown to improve critically ill patients. Monitoring enteral and tube feeding regimens to determine the tolerance and effectiveness of the dietary program is as important as selection and initiation of the formula. Careful nutrition and metabolic assessment can help ensure optimal support with minimal complications. Table [7-4](#) outlines suggested clinical and laboratory parameters that should be continuously monitored. The most effective nutritional support regimens are effected by adherence to sound principles of early implementation of immune function–enhancing enteral feeding, with individual balancing and continuous monitoring.

Ancillary Nutritional Support Measures

Metabolic expenditure can be minimized by blunting stressful stimuli. Thermally injured patients, particularly children, have difficulty maintaining body temperature in cold environments. Because of the apparent change in the hypothalamic set point of thermal neutrality, burn patients require higher ambient temperatures for comfort. The temperature of thermoneutrality is approximately 38.2°C (100.7°F), 4 degrees higher than that of normal subjects. Warming burn patients to this level decreases the metabolic rate and corresponding energy requirements. Thermal blankets, radiation reflectors, and heat lamps may be required to maintain the patient's temperature above 37°C (98.6°F).

Pain that accompanies wound manipulation and other patient care procedures accentuates metabolic expenditure, and administration of narcotics reduces the metabolic rate. Adequate analgesia and sedation should be provided so that patients have periods of uninterrupted rest. Hypovolemia, dehydration, and sepsis are potent stimuli of catecholamine secretion, and appropriate regimens for volume replacement and antibiotic administration should be followed. Systemic infection exacerbates erosion of body mass, and additional

calories must be supplied to maintain nitrogen balance at the same level obtained before infection.

Human growth hormone increases nitrogen retention when administered with adequate calories and nitrogen. Improved nitrogen balance is reflected by increased retention of potassium, phosphorus, and amino acids. The actions of exogenous growth hormone appear to be mediated by the effects of increased insulin secretion on carbohydrate metabolism and perhaps increased secretion of insulin-like growth factor. Lack of activity promotes muscle wasting and atrophy. Vigorous physical therapy promotes preservation of muscle bulk and must be provided on a daily basis to all patients requiring prolonged hospitalization. Patients in skeletal traction or air-fluidized beds are relatively immobile and lose lean body mass as a result; simple isometric exercises can usually be done by these patients. Wound care and expeditious wound closure are the most effective measures for limiting the injury and its metabolic sequelae.

INFECTION

Most morbidity and mortality in severely burned patients are related to infection. Thermal injury causes severe immunosuppression that is directly related to the size of the burn wound. A direct relationship between specific immune defects and infection has yet not been established, but it is likely that this global immunosuppression makes the burn patient susceptible to infection. Sepsis occurs when the balance of interaction between the host and opportunistic organisms is altered unfavorably. Important determinants of sepsis in burn patients are factors such as the creation of new portals of entry, altered host defenses, and exposure to potential pathogenic and opportunistic organisms.

Predictors of Infection

Reliable predictors of infection in patients with severe burns would allow more timely intervention with surgical excision and/or anti-infective agents. Many aspects of the hypermetabolic response of uninfected burn patients are similar to those of infected and septic patients without large inflammatory wound surfaces. The extent of burn injury is one of the major demographic predictors of outcome. The incidence of infection and sepsis rises as burn size increases. Children appear to be more susceptible to systemic infection for a given burn size. The presence of inhalation injury correlates highly with infection and mortality. Patients with severe inhalation injury and no burn can have a fatal outcome. Burns involving less than 10 to 20 percent TBSA in otherwise healthy burn patients are almost never associated with life-threatening infection.

The search for a laboratory study to facilitate the early diagnosis of infection has led to intense examination of postinjury alterations of hormones, acute-phase proteins, and fluorescent substances in the blood and plasma of burn patients. Most laboratory studies are nonspecific and cannot distinguish between inflammation and infection.

Clinical Manifestations

Many of the physiological criteria that have been claimed to reflect sepsis are noninfectious manifestations of postinjury hypermetabolism. Hyperthermia, tachycardia, increased ventilation, and high cardiac output are present routinely in otherwise healthy patients with large burns.

Body temperature in burn patients is dependent partially on environmental conditions. Hyperthermia (39°C [102.2°F] or greater) occasionally represents a febrile response to infection, particularly in children, but episodic elevations in temperature are common in uninfected burn patients. This diagnosis remains one of exclusion, and a definitive diagnosis cannot be made without a workup designed to detect infection. Hypothermia, however, commonly indicates sepsis, usually due to gram-negative organisms. Leukocytosis is also nonspecific. As long as large wounds remain open, moderate elevations in leukocyte counts are common. Thrombocytopenia is caused by several factors, including infection and sepsis. Normal to high platelet counts almost always occur in burn patients who are stable, and are not an indication of the imminent likelihood of sepsis. Thrombocytopenia is one of the major manifestations of infection.

Other systemic manifestations are even more nonspecific. Decreasing mental status can be caused by excessive sedation, histamine receptor blocking agents, and cerebrovascular disease. Hyperglycemia may be due to irregular administration of high-calorie nutrient solutions or hypokalemia. Increased food requirements, hypotension, and oliguria may be related to underreplacement of evaporative water loss or unrecognized diarrhea. The most important observations are related to the temporal association of these physiological events. A precipitant onset of hyperglycemia, fall in blood pressure, and decrease in urinary output should suggest the possibility that the patient is becoming unstable. If these findings are associated with development of hypothermia, leukopenia, and a falling platelet count, the patient is probably developing sepsis, and it is important to do an immediate infection evaluation and administer the appropriate antibiotics.

The most common location of lethal infection is the respiratory tract. Awareness of the variety of infections commonly encountered in burn patients allows an orderly evaluation of the potentially infected patient.

Wound Infection

A change in the pattern of burn wound infections over the past few decades is probably related to the proliferation of broad-spectrum antibiotics. Before the availability of penicillin, streptococci and staphylococci were the predominant infecting organisms. By the late 1950s, gram-negative bacteria (*Pseudomonas* species) had emerged as the dominant organism causing fatal wound infections in burn patients.

All burn wounds become contaminated soon after injury with the patient's endogenous flora or with resident organisms in the treatment facilities. Microbial species colonize the surface of the wound and may penetrate the avascular eschar. This event is without clinical significance. Bacterial proliferation may occur beneath the eschar at the viable tissue–nonviable tissue interface, leading to subeschar separation. In a few patients, microorganisms may breach this barrier and invade the underlying viable tissue, producing systemic sepsis.

The essential pathological feature of burn wound sepsis is invasion of the organisms into viable tissue. The organisms then spread to the perivascular structures and directly invade the vessel wall, causing capillaritis and vascular occlusion. Hemorrhagic necrosis follows; subsequently, organisms invade the bloodstream, producing metastatic lesions. Any organisms capable of invading tissue can produce burn wound sepsis. The predominant organisms causing burn wound infection vary depending on the treatment facility. Burn wound infection can be focal, multifocal, or generalized. The likelihood of septicemia increases in proportion to the size of the burn wound. Since the introduction of effective topical therapy, fungal burn wound infection, primarily involving highly invasive *Phycomycetes* and *Aspergillus* species, has increased.

Pneumonia

One result of the prolonged survival of severely burned patients in critical care units, made possible by modern patient support techniques, is that the respiratory tract has become the most common locus of infection. Bronchopneumonia has replaced hematogenous pneumonia as the most common form of pulmonary infection in burn patients. A diagnosis of pneumonia is confirmed by the presence of characteristic chest radiograph patterns, and the presence of offending organisms and inflammatory cells in the sputum. After inhalation injury, early infiltrates usually represent chemical pneumonitis and not infectious pneumonia, although this damaged lung tissue may become infected. Prophylaxis with antibiotics should not be used, as it selects resistant organisms and does not reduce the incidence of pneumonia. Colonization of the upper airway of patients requiring intubation and mechanical ventilation should not be confused with a respiratory tract infection. For the diagnosis of bronchopneumonia, analysis of sputum samples may be adequate. If there is concern about the identity of the organism, bronchoscopy should be used.

Suppurative Thrombophlebitis

Suppurative thrombophlebitis is a major cause of sepsis in burn patients, occurring in up to 5 percent of patients with major burns. It is associated with the use of intravenous catheters, especially if the catheters have been inserted by cut-down techniques; the incidence increases with the duration of vein cannulation. The nidus of infection is usually located in the vein at the site of the catheter tip, where there is endothelial damage, injury, and fibrin clot formation. The fibrin mesh is subsequently seeded during episodes of bacteremia, which may occur at any time during the hospital stay. This complication can be eliminated by the placement of catheters in high-flow veins, such as the femoral, subclavian, or internal jugular veins, and by changing insertion sites every 48 to 72 h, dislodging the fibrin clot.

Bacterial Endocarditis

Endocarditis is occasionally the cause of occult sepsis in burn patients, and its incidence continues to rise with the increasing use of intravenous catheters for hemodynamic monitoring. Endocarditis should be suspected in patients with positive blood cultures and no other identifiable source of bacteremia. These patients should be examined repeatedly by biplanar, transthoracic, and transesophageal echocardiography until the source of the septicemia is identified. Most lesions are found on the right side of the heart, and over 85 percent of patients have had central venous or pulmonary artery catheters placed in the right atrium or through the right ventricle. Systemic antibiotic therapy should be instituted and continued for at least 4 weeks.

Urinary Tract Infections

Most patients with burns greater than 20 percent TBSA require indwelling urinary catheters to guide fluid resuscitation. Aseptic techniques of insertion and catheter care, the use of a closed drainage system, and the removal of the catheter at the earliest clinically indicated time are effective measures for preventing urinary tract infections. In the absence of an inflammatory response (less than 10 white blood cells per high-power field), the majority of patients with positive urine cultures do not require antimicrobial treatment. Candiduria in the absence of signs of systemic infection can be treated with bladder irrigations with amphoterin B. Burns of the penis usually do not require bladder catheter drainage unless they are severe. Full-thickness burns of the penis should be treated with excision and grafting.

Chondritis of the Ear

The pinna of the ear is composed almost entirely of cartilage with minimal blood supply and is vulnerable to infection. It is a rare complication. When chondritis does occur, a conservative approach with drainage of the helix centrally, in an attempt to preserve the outer cartilages, is usually successful.

Treatment of Infection

The definitive treatment of the septic burn wound is the expeditious excision of the wound. Many of the other infections common to burn patients require surgical intervention. Most infections of burn patients acquired in the hospital involve the organisms that originally colonized the burn wound.

Topical Antimicrobial Therapy

Before the introduction of effective topical antimicrobial agents, up to 60 percent of the deaths in specialized burn treatment facilities were caused by burn wound sepsis. The three agents with proved wide-spectrum antimicrobial activity when applied to the burn wound are silver nitrate, mafenide acetate, and silver sulfadiazine (Table 7-5). Silver sulfadiazine is the most common agent used in burn centers. Only mafenide acetate is able to penetrate the eschar, and it is the only agent capable of suppressing dense bacterial proliferation beneath the eschar surface. The main disadvantage of mafenide acetate is the strong carbonic anhydrase inhibition, which interferes with renal buffering mechanisms. Bicarbonate is wasted, chloride is retained, and the resulting hyperchloremia is compensated for by an increase in ventilation and subsequent respiratory alkalosis. Silver nitrate must be used after injury before bacteria have penetrated the wound. Its disadvantages are the associated electrolyte imbalances, which are common, and methemoglobinemia formation, which

is unusual.

Subeschar Clysis and Surgical Treatment

When burn wound sepsis has developed, the probability of survival is less than 10 percent. Subeschar infusion of antibiotics has been utilized to prevent or treat burn wound invasion in burn patients that escaped topical chemotherapeutic control. Injection of semisynthetic penicillins beneath the infected eschar is associated with markedly improved survival. Subeschar clysis is best used as adjuvant therapy in preparation of patients for eschar excision or as primary treatment for patients who are too unstable hemodynamically to tolerate surgery. Generalized wound sepsis in stable patients should be treated by surgical excision.

Antibiotics

The number and types of antibiotics used in a burn center should be restricted, and the criteria for documenting infections should be well- defined. An infection caused by an identified organism is treated by a single antibiotic. Controlled trials in burn patients have not demonstrated any improvement in survival rates achieved by the routine use of antibiotic combinations to treat serious infection. The indiscriminate use of multiple agents promotes overgrowth of *Candida* species, enterococci, and multiple-antibiotic-resistant organisms in the patient and in the burn center. The current problem of increase in staphylococcal and enterococcal resistance to Vancomycin emphasizes the importance of using the least complex antibiotic shown to be effective against the organisms.

Numerous studies have demonstrated that altered pharmacokinetics of antibiotics in burn patients result in lowered serum drug levels when the usual recommended dose is used. In most cases, these doses are subtherapeutic in the burn patient, especially with antibiotics that are predominantly renally excreted. Serum levels should be monitored frequently and early in the course of therapy. Inappropriate peak levels should prompt alterations in the dosage, while inadequate trough levels should prompt shortening of the dosage interval.

ELECTRICAL AND CHEMICAL BURNS

Electrical Burns

Care at the Scene

Electrical burns are particularly dangerous. If the patient remains in contact with the source of electricity, the rescuer must avoid touching the patient until the current can be turned off or the wires cut with properly insulated wire cutters. Once away from the source of current, the ABCs (airway, breathing, circulation) must be checked. Ventricular fibrillation, or standstill, is common; cardiopulmonary resuscitation should be instituted if carotid or femoral pulses are not palpable. If pulses are present, but the patient is apneic, mouth-to-mouth resuscitation alone may be life-saving. CPR should continue until a cardiac monitor can indicate further treatment. Once an airway is established and pulses return, a careful search must be made for associated life-threatening injuries. Electrocuted patients frequently fall from heights and may have serious head or neck injuries. The intense tetanic muscle contractions associated with electrocution can fracture vertebra or cause major joint dislocations.

Acute and Definitive Care

Electrical burns are thermal burns from very high intensity heat and from electrical disruption of cell membranes. As electric current meets the resistance of body tissues, it is converted to heat in direct proportion to the amperage of the current and the electrical resistance of the body parts through which it passes. The smaller the size of the body part through which the electricity passes, the more intense the heat and the less the heat is dissipated. Fingers, hands, forearms, feet, and lower legs are frequently totally destroyed; areas of larger volume, like the trunk, usually dissipate enough current to prevent extensive damage to viscera unless the entrance or exit wound is on the abdomen or chest. Arc electrical burns are common in addition to the usual entrance and exit wounds. These deep and destructive wounds occur when current takes a direct path, often between joints in close apposition to one another at the time of injury. Burns of the volar aspect of the wrist, the antecubital fossa when the elbow is flexed, and the axilla are most common. While cutaneous manifestations of electrical burns may appear limited, the skin injury is only the tip of the iceberg, and massive underlying tissue destruction may be present. Resuscitation needs are usually far in excess of what would be expected on the basis of the cutaneous burn size, and associated flame and/or flash burns often compound the problem. Myoglobinuria frequently accompanies severe electrical burns. Disruption of muscle cells releases cell fragments and myoglobin into the circulation to be filtered by the kidney. If this condition is untreated, the consequence can be permanent kidney failure.

Cardiac damage, for example, myocardial contusion or infarction, may be present. The conduction system may be deranged, and in some cases, there can be actual rupture of the heart wall or rupture of a papillary muscle, leading to sudden valvular incompetence and refractory cardiac failure. Household current at 110 V either does no damage or induces ventricular fibrillation. If there are no cardiac abnormalities present in a patient in the emergency room after shocks of 110 to 440 V, the likelihood that they will appear later is small. Even with injuries resulting from high- voltage currents, normal cardiac function on admission generally means that subsequent cardiac dysrhythmia is unlikely. Studies confirm that commonly measured cardiac enzymes bear little correlation to cardiac dysfunction, and elevated enzymes may be from noncardiac muscle damage. Monitoring ECG and isoenzymes in an ICU setting for 48 h may be unnecessary in patients with electrical burns who have stable cardiac rhythms on admission.

The nervous system is particularly sensitive to electricity. The most severe brain damage occurs when current passes through the head, but spinal cord damage is possible whenever current has passed from one side of the body to the other. Myelin-producing cells are susceptible, and delayed transverse myelitis can occur days or weeks after injury. Conduction remains normal through existing myelin, but as the old myelin wears out, it is not replaced and conduction stops. Damage to peripheral nerves is common, and may cause permanent functional impairment. Every patient with an electrical injury must have a thorough neurological exam as part of the initial assessment. Persistent neurologic symptoms may lead to chronic pain syndromes, and so-called posttraumatic stress disorders are much more frequent after electrical burns than after thermal burns.

Cataracts are a well-recognized complication of electrical contact burns. They occur in 5 to 7 percent of patients followed, they are frequently bilateral, and they can occur even in the absence of contact points on the head. They often occur within a year or two of injury. Electrically burned patients should undergo a thorough ophthalmologic examination during the admissions phase of acute care. In addition, electrical burns are frequently job-related, and a thorough baseline examination will help workers get job-related insurance if they develop cataracts in the future.

Wound Management

There are two situations in which early surgical treatment is indicated for patients with electrical burns. Rarely, massive deep tissue necrosis will lead to acidosis or myoglobinuria, which will not clear up with standard resuscitation techniques, and major debridement and/or amputation may be necessary on an emergency basis. More commonly, the deep tissues undergo swelling, and the risk of compartment syndrome further compromising damaged tissue is real. Careful monitoring, including measurement of compartment pressures, is mandatory, and escharotomies and fasciotomies should be performed at the slightest suggestion of progression. Routine compartment pressure measurements may be helpful, but any of the signs of impending compartment syndrome (increased pain, pallor, absence of pulseless, decreased sensation, and tense swelling) mandate prompt compartment release in the operating theater. Any progression of median or ulnar nerve deficit in a hand that has been electrically burned is an indication for median and ulnar nerve release at the wrist.

If immediate decompression or debridement is not required, definitive surgical procedures can be done between days 3 and 5, before bacterial contamination occurs and after the tissue necrosis is delineated. Vascular grafts to replace clotted arteries are sometimes an option. However, they may actually increase morbidity and prolong recovery, and amputation and one of the newer prostheses might provide better function than a hand or foot with poor sensation and motor function.

Chemical Burns

Emergency Care

In cases involving chemical burns, whenever possible, involved clothing should immediately be removed and the burns should be thoroughly flushed with copious amounts of water at the scene of the accident. Chemicals will continue to burn until removed, and washing for at least 15 min under a running stream of water may limit the overall severity of the burn. No thought should be given to searching for a specific neutralizing agent. Delay deepens the burns, and neutralizing agents may cause burns themselves; they frequently generate heat while neutralizing the offending agent, adding a thermal burn to the already potentially serious chemical burn. Powdered chemicals should be brushed off skin and clothing.

Chemical burns, usually caused by strong acids or alkalis, are most often the result of industrial accidents, assaults, or the improper use of harsh solvents and drain cleaners. In contrast to a thermal burn, chemical burns cause progressive damage until the chemicals are inactivated by reaction with the tissue, or diluted by flushing with water. Individual circumstances vary, but acid burns may be more self-limiting than alkali burns. Acid tends to “tan” the skin, creating an impermeable barrier that limits further penetration of the acid. Alkalis combine with cutaneous lipids to create soap and thereby continue “dissolving” the skin until they are neutralized. A full-thickness chemical burn may appear deceptively superficial, causing only a mild brownish discoloration of the skin. The skin may appear to remain intact during the first few days postburn, and only then begin to slough spontaneously. Chemical burns should be considered deep dermal or full-thickness, until proved otherwise.

Some chemicals, such as phenol, cause severe systemic effects, and others, such as hydrofluoric acid, may cause death from hypocalcemia even after moderate exposure. Unless the characteristics of the chemical are well known, the treating physician is advised to call the local poison control bureau for specifics in treatment.

OUTPATIENT MANAGEMENT OF THERMAL INJURIES

Minor burns comprise approximately 95 percent of all burns treated in the United States. These burns are usually superficial, do not exceed 10 to 15 percent of the body surface area, and rarely require hospitalization. Many moderate and even major thermal injuries, as classified by the American Burn Association Injury Severity Grading System, are amenable to outpatient management after initial evaluation and stabilization. Although survival is usually not an issue and most minor burns will ultimately heal regardless of therapy, undertreatment and (more commonly) overtreatment can result in infection or delay in healing with discomfort and prolonged morbidity. The goals of outpatient burn management include wound healing, patient comfort, and rapid rehabilitation.

Management of Minor Burns

Treatment at the Scene of the Accident

After elimination of the heat source, areas with minor burns should be placed in tepid water rather than ice water. The potential benefits of cooling burn wounds are controversial (Table 7-6). It has been suggested that the beneficial effects, if they do exist, last only for the first 2 or 3 minutes after thermal injuries, and the application of ice water after this initial period may result in prolonged edema and impairment of healing, and may convert a partial-thickness to a full-thickness injury. The burn area should then be wrapped in a clean cloth and the victim taken to an emergency facility.

Chemical burns should be irrigated with copious amounts of water. Tar burns should be cooled with water, but the tar should not be removed at the scene of the accident.

Initial Medical Management

A protocol for immediate wound care in the emergency room or outpatient facility is mandatory for efficiency in treating thermal injury, which is too often characterized by panic, confusion, and a multiplicity of approaches. A thorough history taking should elicit when and where the accident occurred, and the burning agent (flame, hot liquid, electricity, etc.). The history should also determine whether there is any possibility of smoke inhalation injury. Depending upon the circumstances of the accident, evaluation for possible associated injuries is also important. Pertinent past medical history, including drug allergies, medication history, and history of systemic illnesses, should also be obtained.

Tetanus prophylaxis is the same for minor burns as for any other injury. A tetanus toxoid booster is given to any patient who has not received one for 5 years or who cannot recall the date of last immunization. Patients not previously immunized should receive 250 units of tetanus human immune globulin, and the first of a series of active immunizations with tetanus toxoid.

Burn wounds may be soaked in tepid water or covered with saline-soaked sponges that decrease the pain until the physician can evaluate the patient. Ice water should not be used. The wounds should be washed with mild soap and warm water, excessive debris trimmed, and hair shaved within margins of at least 1 in around the burn wound.

The removal of tar and asphalt is best accomplished with the use of Medisol, a citrus and petroleum distillate with hydrocarbon structure that consists of 70% petroleum distillate (base oil), 25 to 27% limonene (orange oil), 2 to 3% lanolin, and 1% surfactant (dioctyl sodium sulfosuccinate). This product has proved to be the most efficient in removing tar without damaging the underlying burn wound. Mineral oil and petroleum ointments, such as bacitracin or Neosporin, may also be used to remove tar. The tar should not be peeled off because of potential damage to the hair and skin incorporated in the tar.

Chemical burns should be irrigated with water for 20 min if this was not done at the scene of the accident. A neutralizing agent should not be applied, because the neutralizing reaction is always exothermic (heat-producing) and may result in a more severe injury.

Blisters may be managed in one of three ways:

- (1) The blister is left intact and the underlying wound allowed to heal in the blister fluid environment.
- (2) The blister fluid is evacuated and the overlying skin allowed to cover the underlying wound.
- (3) The blister is debrided.

Which technique is used depends not only on the location and size of the blister, but also on the compliance of the patient. If there are questions concerning a patient's reliability, or there is a potential for wound infection, it is more practical and safer to debride the blister.

After the wound is cleansed and debrided, the extent and depth of injury and areas involved are estimated. If the burn is minor, the patient can be treated as an outpatient. If there is any question regarding the feasibility of outpatient management, it is safer to admit the patient to the hospital for 24 to 48h, after which outpatient management is feasible if everything is favorable.

Follow-up Wound Care

The problems associated with major thermal injury—immunosuppression, hypermetabolism, and increased susceptibility to infection—are not associated with minor burn wounds, as it is frequently erroneously assumed. Basic principles include keeping the wound clean and in a moist environment while the wound heals. Powerful topical chemotherapeutic agents such as mafenide acetate (Sulfamylon), or povidone-iodine (Betadine) should not be applied to minor burn wounds, since they have been shown to delay wound healing. In addition, systemic antibiotics are rarely indicated for the treatment of small burns, and may predispose the wound to later opportunistic infection with bacteria, fungi, or viruses.

Follow-up care should include: (1) washing the wound with bland soap and water in a bathtub or shower, (2) patting the wound dry with a clean towel, (3) applying a bland ointment, such as bacitracin, and a nonstick porous gauze, such as Adaptic or Xeroform, then wrapping the wound lightly with gauze rolls (Kerlix). This regimen is performed twice daily (see Table 7-6).

Normally, follow-up is performed weekly in the burn clinic. If there is some question regarding the extent or depth of the wound, or the reliability of the patient or his/her family, initial follow-up may be performed on a daily basis.

The patient with a minor burn should be given instructions for a vigorous program of range-of-motion exercises. Prolonged edema, which retards wound healing, is minimized with adequate physical therapy. Rehabilitation time may also be decreased if an active exercise program is followed.

If the wounds are deep partial- or full-thickness burns, the patient can still be treated as a outpatient until primary excision and grafting can be performed. In children, the thickness of the epidermis and dermis is decreased, and thus the depth of injury is difficult to determine. If there is any question regarding depth of injury, the wounds are treated conservatively for 10 to 14 days before excision and grafting.

If the wound is healing properly, it will be totally epithelialized in 2 to 3 weeks. A prediction of wound healing can be made by 14 days postinjury. If the wounds are not healing within this time period, then primary excision and grafting should be performed. Waiting for eschar separation and formation of granulation tissue delays wound closure, prolongs wound healing, and results in increased scarring and prolonged rehabilitation time.

A superficial partial-thickness injury should be followed until epithelial coverage has occurred and then examined at 6 weeks for evidence of hypertrophic scarring. If hypertrophic scarring occurs, compression dressings should be fitted and worn until the wound becomes quiescent, which usually requires 12 to 18 months. A problem that occurs in a healed burn is the formation of very thin water blisters, produced as a result of minor trauma. These blisters, which rarely exceed 1 cm, occur 2 to 6 weeks after wound closure and leave small open areas that heal without incident in 3 to 5 days if they are kept clean with bland soap and water and covered with a bland ointment, such as bacitracin. Recently healed partial-thickness burn wounds become very dry. A mild lanolin cream or lotion should be used until the natural skin lubrication mechanisms return, usually within 6 to 8 weeks.

The patient should be instructed to avoid sun exposure until the wound is completely healed; exposure to the sun may cause hyperpigmentation of the wound, which is frequently permanent. The use of a sun block (sun protection factor 30 to 50) is recommended for healed areas that must be exposed to direct sunlight.

Pruritus is a common complaint in patients with healing burn wounds, and severe pruritis is extremely difficult to treat. Diphenhydramine hydrochloride or hydroxyzine antipruritics have proved to be beneficial. Using a moisturizing cream also helps alleviate itching.

Alternative Methods of Wound Management

There are multiple methods for managing outpatient burn wounds. Some advocate bulky dressing for as long as 10 to 14 days. Although a bulky dressing prevents pain, there is a potential hazard of bacterial overgrowth in the warm, moist dressing, and range-of-motion exercising cannot be performed, which prolongs edema and interferes with rehabilitation. Topical chemotherapeutic agents, such as silver sulfadiazine, are frequently used in the management of minor burn injury. If this form of management is used, it is important to remember that silver sulfadiazine is inactivated by tissue secretions and must be changed at least twice a day. A pseudomembrane forms over the partial-thickness injury, which is frequently painful and difficult to remove.

The use of prosthetic skin substitutes in the treatment of partial-thickness outpatient burns has recently become popular. Two substitutes that are useful on partial-thickness injury are Biobrane and OpSite. Biobrane can be used effectively in partial-thickness burns. It is well tolerated, is capable of compressing underlying fluid collection, and does not promote bacterial proliferation. It may be left on until epithelialization occurs. OpSite is a synthetic barrier. Its use requires some expertise, because fluid collections beneath the barrier frequently occur. The problem of accurate diagnosis of depth of injury is a drawback to the use of either of these prosthetic skin substitutes. If the wound is a medium to deep partial-thickness injury, an eschar does form, necessitating frequent dressing changes.

Biological dressings including allograft, xenograft, or amnion have been used to treat outpatient burns in the past. The cost of cadaveric allograft is prohibitive. Xenografts should not be used on superficial partial-thickness injury because of the incorporation of xenograft tissue into the healing donor sites, which occurs in as many as 35 percent of the patients so treated.

Management of Critical Areas

Face

Superficial burns of the face should be left exposed. The face is washed twice daily with a mild soap and water, and a thin layer of a bland ointment (bacitracin) is applied to the open wounds to prevent drying.

Ears

Superficial burns of the ear should be treated with a bland ointment. Deeper injuries must be treated with topical antibiotics; excessive pressure may cause chondritis, and should be avoided.

Eyes

Suspected corneal burns should be stained with fluorescein for confirmation of diagnoses. Superficial corneal burns should be treated similarly to corneal abrasions, with vigorous irrigation, the application of ophthalmologic antibiotic ointment, and eye patching. Superficial corneal burns not asymptomatic by 48 h and more serious injuries should be evaluated by an ophthalmologist.

Hands

Superficial burns of the hand should be elevated for 24 to 48 h to minimize swelling. Circumferential hand burns may require hospitalization for observation of adequate circulation. Range-of-motion exercises should begin as soon as possible after injury. Instructions for the exercise program should be part of the management protocol.

Feet

Although burns of the feet are painful, walking and range-of-motion exercises should be performed. Crutches should not be allowed. To prevent edema, burned feet should be elevated when the patient is not walking or exercising. An elastic bandage should be applied over the wound dressing when the patient is walking or sitting, but it should be removed at night when the feet are elevated.

Perineum

Perineal burns frequently require hospitalization for 24 to 48 h for observation of urinary obstruction secondary to edema. Minor perineal burns can be treated with a bland ointment. Extensive superficial perineal burns, e.g., pediatric bathtub scald injuries, are best treated with topical chemotherapy (silver sulfadiazine), utilizing a diaper as the wound dressing.

Complications

Most complications in small burn injuries result from overtreatment: too-vigorous dressing changes that pull off newly formed epithelium, or the use of a variety of topical and systemic antibiotics that can cause secondary infection or formation of a pseudomembrane that may require debridement, either of which may delay healing. Treatment that is appropriate for larger burns is overtreatment of small burns. In small burns, systemic antibiotics are rarely indicated and topical agents are usually not necessary.

Management of Moderate or Major Burns

Patients with superficial partial-thickness moderate and major burn injuries can also be successfully managed as outpatients. In these burns, as in minor burns, survival is not an issue. The cost of burn care in the United States is approaching \$2000 per day, and outpatient management of these types of burns can markedly decrease these costs. The presence of endemic, drug-resistant microorganisms within any hospital environment, particularly in burn centers, poses a potential threat to the moderately severe superficial partial-thickness burn patient, with frequent colonization of the burn wound by endemic staphylococcal and pseudomonas organisms.

Medical Criteria for Outpatient Management

The medical criteria for whether a patient is treated as an outpatient include: (1) no existing complications of thermal injury, such as inhalation injury, (2) fluid resuscitation completed, (3) stabilized hospital course, (4) adequate nutritional intake, (5) adequate pain tolerance, and (6) no anticipated septic complications. The patient's family must be willing to participate in the care of the patient and must be capable of doing so. Criteria for "capability" include personal cleanliness, no apparent aversion to burn care, an ability to perform the dressing changes and assist in range-of-motion exercises, and access to transportation to return the patient for burn care and physical therapy. Not all families initially meet these criteria, but with persistent education by the burn team, most families can learn how to conform to the guidelines.

Outpatient Treatment Program

Outpatient management of moderate and major thermal injuries involves two phases of treatment: a home-treatment program and physical therapy. At home, the patients are instructed to bathe or shower twice daily, wash the wound with mild soap and water, gently debride the wound with a washcloth, and apply prescribed topical ointment followed by a light dressing. Patients exercise every hour, following a program outlined by the physical therapist. Family members are instructed to encourage self-care and active range of motion at all times.

When the patient is discharged from the hospital, an appointment is made for outpatient physical therapy. The frequency of treatments decreases as the wound heals, but initially patients are seen daily. The treatments consist of hydrotherapy, debridement of burn wounds, and a supervised exercise regimen. The wounds are evaluated by the physical therapist and feedback given to the family members regarding wound status, dressing changes, and range of motion. All patients are seen weekly in the burn clinic, where the wounds are examined, treatment reviewed, and necessary changes made.

PAIN CONTROL

All burn injuries are painful, whether the injury is a simple sunburn or an extensive second- or third-degree burn covering a large portion of the body. Attempts to manage pain in a patient with a burn injury are frequently frustrating because of the often-changing physiological and psychological reactions to the injury. A superficial, partial-thickness injury (first-degree burn) damages the outer layers of skin, the epidermis, producing mild pain and discomfort. The pain associated with moderate to deep partial-thickness or second-degree burns varies depending on the extent of destruction to the dermis. Superficial dermal burns initially are the most painful, and even the slightest change in air current on the exposed superficial dermis causes the patient to experience excruciating pain. Without the protective covering of the epidermis, nerve endings are sensitized and exposed to stimulation. Areas of deep partial-thickness or full-thickness injury show little or no response to sharp stimuli, yet a patient may complain of a deep, aching pain, which is related to the inflammatory response. The physiological effects of pain include increase in heart rate, blood pressure, and respiration; a decrease in O₂ saturation; palmar sweating and facial flushing; and dilatation of the pupils. No single physiological change is an absolute indication of pain.

The burn patient may experience acute pain from dressing changes, operative procedures, and rehabilitation exercises. The patient may also have chronic background pain associated with the wound maturation process. Pain management involves pharmacologic and nonpharmacologic modalities. Pharmacological modalities include hypnotics and analgesics, such as morphine, methadone, codeine, acetaminophen, and nonsteroidal anti-inflammatory agents. Anesthetic agents, such as ketamine nitrous oxide and fentanyl, are useful for severely painful dressing changes, such as those after skin-grafting procedures. Psychotropic drugs, such as antianxiety drugs, major tranquilizers, and antidepressants, may be useful in the management of burn wounds. Nonpharmacologic methods include providing verbal and/or physical comfort, arranging activities that will distract the patient from concentrating on the pain, and relaxation therapy.

During burn shock resuscitation, pain medication, such as morphine sulfate, should be given sparingly; if administered, the preferred route is intravenous. Because of irregular absorption, pain medication should not be given intramuscularly. Morphine and/or methadone therapy is begun early on all patients with acute burns admitted to the hospital, unless the patient will be in the hospital only for a very short time. Total elimination of pain in burn patients is not possible, short of general anesthesia. The major neuroleptic modalities, including ketamine nitrous oxide and fentanyl, all compromise respiratory function and should be given only under direct supervision by personnel skilled in airway control and respiratory support. The concomitant use of benzodiazepam, hypnosis, and psychological support reduces narcotic requirements. During the convalescent phase of burn management, a regular dose of an oral analgesic such as methadone produces more effective pain relief than unscheduled "as-needed" doses.

REHABILITATION AND CHRONIC PROBLEMS

Rehabilitation

Inpatient Therapy

Maintaining function and preventing the complications of prolonged immobility are the specific goals of the rehabilitative treatment of burn patients. Daily assessment of the patient's range of motion, ambulation, and functional status is necessary to determine the effectiveness of ongoing treatment plans and to make modifications as needed for new problems. The location of the burn in relation to the joint axis determines what movement will be limited as the burn heals. In most burns of extremities, the position of maximal comfort promotes the formation of scar contractures. Because compliance is a major factor in a successful rehabilitation program, the burn therapist must work closely with the entire burn team in developing the patient's trust, understanding, and confidence. Burn teams should include physical and occupational therapists, and a play therapist who can engage children in physical activities in an environment in which they often are not aware of the therapeutic nature of the exercises they are performing. Such an approach commonly enables children to achieve full range of motion independently. Passive exercises must be carefully planned, since overzealous activity may lead to tendon disruption, muscle tears, heterotopic ossification, and traumatic release of scar contractures.

Physical therapy begins on the day of admission. Burned extremities are elevated and actively exercised to minimize edema and reduce the need for escharotomy. Stable patients are initially placed in chairs; ambulation begins when it can be tolerated. Excessive use of analgesics and antianxiety drugs impedes a successful mobilization program. When patients get out of bed, burned legs are wrapped with disposable compression bandages to prevent venous stasis and edema. Even when patients are in bed, nearly all burned extremities develop varying degrees of edema. Active exercises maintain muscle mass and strength. Passive exercises are most often used with debilitated patients and patients whose state of awareness is clouded. Continuous passive motion devices and dynamic splinting increase function in these patients and do not require the continual presence of rehabilitation therapists. Objective measurements of joint stiffness can be quantified by piston-drive displacement transducers, which provide useful documentation of patient progress.

Burn contractures are unlike other types of contractures. Burn scars commonly envelop the entire circumference of single joints, and the scar may involve multiple joints. Contraction may pull a joint in one direction and an adjacent joint in the opposite direction. A classic example is boutonnière deformity of the fifth finger. The entire burn scar must be uniformly stretched; complex splints or multiple joint pin fixation are often required. Positioning of body parts in the antideformity posture prevents wound contracture. Splints are also used for immobilizing freshly grafted extremities to prevent inadvertent shearing of the new autograft.

All second- and third-degree burns produce permanent scarring. Some scars in healed second-degree burns are barely noticeable, while deeper burns, even when grafted, may develop bulky hypertrophic scar tissue. Scar hypertrophy can be retarded by the use of custom-fitted pressure garments over healed scars. Specially fitted inserts can be used to apply pressure to concave areas of skin. Patients whose burns do not heal by the second postburn week should be treated with pressure devices, first with a generic tubular elastic stocking and later, after the surface of the burn scar stabilizes, with custom-made compression garments. Adults usually wear these garments for 3 to 6 months (Fig. 7-4), while small children require up to 4 years of compression therapy before scar maturation is complete.

Outpatient Therapy

Many functional deficits persist after burn patients have been discharged; therefore follow-up must be continuous for prolonged periods. For many patients, the burn center outpatient facility provides their only access to primary care. Pressure garments require regular refitting. Patients using complex physical training devices and special techniques for reducing contracture may need to visit the rehabilitation department daily. Outpatients are evaluated 1 week after discharge, and the interval between visits is gradually lengthened, depending on individual patient needs. As burn scars mature, permanent residual deficits or deformities may be amenable to reconstructive surgical correction. Usually, multiple small operations are used; corrective surgical procedures may take place over a period as long as 10 years.

Patients develop follicular infection in the burn wound several months after injury. These plugged follicles usually disappear once hair erupts through the overlying epithelium. Severe itching and vague, but intense, neuritic pain are long-lasting and are poorly responsive to antipruritic medications and analgesics. Providing detailed printed explanations of burn care and treatment procedures is an essential component of follow-up care.

Psychological Support

Burned patients display a variety of psychological responses to their injury, including anxiety, depression, denial, withdrawal, and regression. Withdrawal and regression are especially common in children, who may refuse to participate in treatment regimens. Play therapy provides a way for children who have similar cosmetic deformities or functional deficits to interact. Rapport between team members—play therapists, physical and occupational therapists, burn nurses, and physicians—promotes compliance among the patients and encourages the proper use of pressure garments and assistive devices.

Nearly half of older children and adults develop posttraumatic stress disorder after thermal injury, which is characterized by recurrent and intrusive recollections of the initial injury, avoidance of circumstances that invoke memories of the event, loss of interest in daily activities, feelings of isolation, hyperalertness, memory impairment, and sleep disturbances. Noncompliance with burn therapy is a serious outward manifestation of a patient's attempt to avoid recollections of the traumatic event. A patient's psychological state after the burn injury, and symptoms of posttraumatic stress disorder during acute-care hospitalization, are predictive of the likelihood of chronic posttraumatic stress disorder. The severity of injury does not correlate with the symptoms. Both short- and long-term psychotherapeutic support is frequently necessary in burn patients, and a full-time psychiatrist is an essential member of the burn team. Burn patients rarely seek treatment for psychological problems. The frequency of posttraumatic stress disorder increases with length of follow-up.

Family support groups convene on a weekly basis. The burn team presents an update on the patient's progress, and addresses specific short- and long-term goals to allay anxieties. Psychosocial support is critical for the burn patient for the duration of the course of treatment and follow-up.

Chronic Problems

Hypertrophic Scar and Keloid Formation

Burn scar hypertrophy typically develops in deeper partial-thickness injuries and third-degree burns that are allowed to heal by primary intention. Hypertrophy of grafted areas of excised burn wounds occurs less frequently, and is dependent, in part, on the time from injury to excision, the site of the wound, and the surgical technique used. With tangential excision, necrotic tissue of a partial-thickness burn is removed in successive layers until a base of partially viable dermis is reached; in most circumstances, the wound is then immediately grafted. Sequential excision extends to varying levels of the skin and subcutaneous tissue until all nonviable tissue is removed. Sequential excision goes beyond tangential excision to include the complete excision of necrotic full-thickness injuries. Delayed tangential excision is more likely to result in residual scar hypertrophy in grafted burn wounds.

Because only a few epithelial elements, sweat glands, and hair follicles remain viable in deep, partial-thickness burns, healing takes place from these remnants over a period of 3 to 6 weeks. The resulting scar epithelium is of poor quality, and is prone to hypertrophy. Hypertrophic scar should be distinguished from a keloid. Both exhibit excessive collagen formation, but a keloid overgrows the original dimensions of the initial injury and hypertrophic scar develops in the bed of the injured tissue and is confined to its original anatomic boundaries. Hypertrophic scars frequently flatten with time and pressure, whereas keloids do not. Long- term controlled trials have not clearly demonstrated permanent benefits from compression therapy, but compression garments quickly reduce the mass of hypertrophic immature scars and provide patients with tangible evidence of the benefits of conscientious follow-up. The mechanism by which constant compression reduces scar mass is not well defined. It may work by causing ischemia in the microvasculature of hypertrophic scars, with focal degeneration of cells because of hypoxia after microvascular occlusion. Another mechanism may involve the healing sequence; electron microscopy has demonstrated that fibroblasts in pressure-treated wounds are more linearly organized than fibroblasts in non-pressure-treated wounds, and collagen is manufactured in a more organized fashion.

Other forms of therapy for hypertrophic scarring include radiotherapy, cryotherapy, and reexcision and wound closure. Radiotherapy in doses of 1500 to 2000 Gy has been used with varying results. Radiation probably reduces fibroplasia and capillary budding. Cryotherapy is rarely used. It is associated with depigmentation and increased melanocyte sensitivity to subsequent cold exposure. The most successful approach to residual hypertrophic burn scars is initial pressure therapy until the wound matures, followed by subsequent excision and application of skin grafts. Tissue expansion techniques have been used to expand normal skin and replace the excised hypertrophic scar or keloid. Complication rates with tissue expansion are as high as 40 percent, and include infection, implant extrusion, and device rupture. These complications usually require removal of the original implant, treatment of any infection, and replacement.

Because of their marked propensity to recur, keloids are difficult to treat. Excision of the keloid and primary closure are effective for linearly oriented keloids with a narrow base, but excessive wound tension leads to recurrence. Broad-based keloids may be removed flush with the surrounding skin, and a split-thickness skin graft placed over the base of the keloid to prevent it from recurring. Excision alone has a recurrence rate of greater than 50 percent. Intralesional injection of corticosteroids may reduce the bulk of keloid and hypertrophic scar mass, and may be used in combination with excision or split-thickness skin grafting. It is believed that triamcinolone, the most commonly used steroid, acts by decreasing collagen synthesis and increasing collagen degradation through the collagen inhibitors α_2 -macroglobulin and α_1 -antitrypsin. When steroid injection is used in conjunction with surgery, keloids should be injected for at least 1 month prior to the operation. Some surgeons inject triamcinolone in the base of the wound and along its edges during the surgical procedure. Postoperatively, the patient receives injections monthly until the wound matures. The major side effects of intralesional injection of steroids are hypopigmentation and atrophy of the skin surrounding the keloid.

Burn scar hypopigmentation and surface irregularity can be significantly improved by dermabrasion and thin split-thickness grafting. Adequate pigmentation and flat surfaces are obtained in most patients. Tissue expanders are particularly effective for treating burn scar alopecia.

Approximately 20 percent of patients treated in burn facilities are readmitted for reconstructive procedures. The most common areas of reconstruction involve the hand and wrist (most common), arm and forearm, face, and neck. Improved inpatient burn treatment and scar management have reduced the need for subsequent reconstructive surgery.

Marjolin's Ulcer

Chronic ulceration of old burn scars was noted by Marjolin to lead frequently to malignant degeneration. Squamous cell carcinoma is most common, although basal cell carcinomas occasionally occur. Rare tumors, including malignant fibrous histiocytoma, sarcoma, and neurotropic malignant melanoma, have also been described. Chronic breakdown of a healed burn wound scar should lead to suspicion of malignant degeneration. These lesions typically appear decades after the original injury, but burn scar carcinoma can be encountered within the first year. In the absence of cancer, most unstable burn scars should be excised and resurfaced. Burn scar carcinomas may metastasize aggressively. Malignancy dictates wide excision, but prophylactic regional node dissection has not improved mortality. Approximately 30 percent of burn scar carcinomas occur in the head and neck. Adjuvant radiotherapy improves survival. Very deep dermal burns that have not healed by the third or fourth postburn week occasionally develop nodules in the burn wound bed. These lesions are sometimes erroneously interpreted as carcinoma or lymphoma, but excision and grafting of the nodules wounds usually reveals that they are pseudoepitheliomatous hyperplastic lesions or keratoacanthomas.

Heterotopic Ossification

Heterotopic ossification occurs in up to 13 percent of burn patients. This complication may develop in patients with partial-thickness burns and around extremities not involved with the injury, but it most commonly involves patients with full-thickness burns greater than 20 percent TBSA, and is found adjacent to the involved joint 1 to 3 months after injury. The elbow is the most commonly affected joint. The diagnosis is usually made by the physical or occupational therapist, who discovers increased pain and decreased range of motion of the involved joints. Limitation of physical activity usually precedes radiographic evidence of calcification, which is located in the muscle and surrounding soft tissue of the joint. Although the mechanism causing heterotopic ossification is not known, it has been suggested that bleeding into the soft tissue, due to aggressive physical therapy, is the culprit. Prolonged immobilization of a joint encompassed by a burn also appears to promote heterotopic ossification. Restricted activity promotes mobilization of body calcium stores and may lead to deposition of calcium in the soft tissues. Some suggest surgical removal of all ossified soft tissue, but others recommend modification of rehabilitation therapy regimens and allowing the reabsorption of ossified tissue.

Fractures

Up to 10 percent of burned patients have associated fractures. Fractures in patients with large burns are treated with splints or traction until resuscitation is complete. Operative repair is preferably performed within 48 to 72 h of burn injury. Burn wounds adjacent to fracture sites are usually excised and autografted at the time of internal fixation. If internal fixation is not possible, external fixation is used, which permits access to the burn wound and provides stability.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 8 - Wound Care and Wound Healing

I. Kelman Cohen
Robert F. Diegelmann
Dorne R. Yager
Isaac L. Wornum III
Martin F. Graham
Mary C. Crossland

INTRODUCTION

Wound healing has been described throughout recorded history. Empirically, the ancients recognized that foreign bodies and dead tissue must be removed from wounds. They knew that cleanliness prevented infection and that loculated pus required drainage. Wound elixirs such as honey decreased wound suppuration (hypertonic glucose is bactericidal), and acute open wounds could be sutured primarily with hairs, cloth, or insect jaws. In the sixteenth century, Paré discovered that surgical destruction of tissue, such as was effected by pouring boiling oil into acute open wounds, impeded healing and led to sepsis. His observations led to the maxim of all surgeons today—"Do not put anything in a wound you would not put in your own eye." The clinician's treatment of tissue must be as atraumatic as possible. Lister, Semmelweis, Ehrlich, Fleming, and Florey realized, with increasing sophistication, that bacteria were pathogens that prevented healing and led to sepsis and death. Control of bacteria by asepsis, antiseptics, and antimicrobials heralded a new era in wound management.

Control of pain, understanding fluid and blood replacement, and antibiotics, were the major contributions to wound management during the first half of the twentieth century. Objective assessment of abnormal tissue repair received little attention. During the past two decades, however, wound care has made more advances than it had over the past two thousand years. Four major factors account for the logarithmic advances in wound healing knowledge: (1) the biologic mechanisms of tissue repair are now being defined on a biochemical and a molecular level; (2) financial support for wound healing research has increased markedly because the social and financial devastation of wounds has come to be appreciated by health care providers and federal health care funding agencies; (3) the medical-industrial complex can envision profit in the discovery of more efficacious modalities for wound care and, hence, is supporting wound healing research; and (4) reconstructive surgical techniques have changed drastically in the past two decades with the advent of muscular and musculocutaneous flaps as well as microvascular free-tissue transfers. The recent advances are only a prelude to the changes in wound care and management that will occur in the near future.

Mammalian healing is primitive in comparison with that of the "lower" forms of life. It is paradoxical that in creatures considered low on the phylogenetic scale tissue heals by a process of *regeneration* that is much more sophisticated than the primitive *repair* process in human beings—"wound healing"—which involves mechanisms of inflammation, matrix deposition, and scarring. The ultimate goal of wound healing research is to enable human wound healing to proceed by a process of regeneration.

GENERAL CONSIDERATIONS

Wound healing is a vague term that often diverts the clinician from focusing on the specific mechanism involved in a particular healing process. Only after defining the specific biologic processes of a particular wound can the clinician formulate a rational approach to therapy. Therefore, the reader must first have a very clear definition of the *mechanisms* of wound healing and how they contribute to various *types of wound closure*.

Classification of Wounds

Wounds can be classified into two general categories: *acute* and *chronic*. Acute wounds "normally proceed through an orderly and timely reparative process that results in sustained restoration of anatomic and functional integrity." In contrast, "chronic wounds have failed to proceed through an orderly and timely process to produce anatomic and functional integrity, or proceeded through the repair process without establishing a sustained anatomic and functional result."

Types of Wound Closure

Primary closure approximates the acutely disrupted tissue with sutures, staples, or tape. With time, the synthesis, deposition, and cross-linking of collagen and other matrix proteins, which are of primary importance in this type of repair, provide the tissue with strength and integrity (Fig. 8-1).

In *delayed primary closure*, approximation of the wound margins is delayed for several days after the wound has been created (Fig. 8-2). Delay in closure is indicated to prevent infection in wounds in which there is significant bacterial contamination, foreign bodies, or extensive tissue trauma. During the period that the wound remains open, moist, sterile saline dressings should be changed at least twice a day to optimize the wound for closure. The use of peroxide and iodophors in these open wounds should be avoided because these solutions destroy the host's tissues as well as bacteria. The rationale for delayed primary closure is based on the normal biological event of healing. If the wound is open, there is less chance that bacterial colonization will occur. During the time the wound is open, events occur that will significantly decrease the chances of wound infection after closure. For example, leukocytes are attracted to the wound site, and angiogenesis proceeds to provide enhanced blood supply and needed oxygen. All of these phenomena lead to destruction and removal of bacteria. Simultaneously, the normal phases of healing progress even though the wound remains open. In fact, wound tensile strength after delayed primary closure eventually becomes the same as that of the primarily closed wound.

Spontaneous closure, or "*secondary*" *wound closure*, occurs when the margins of the open wound move together by the biologic process of contraction (Fig. 8-3). A striking example is the lower extremity amputation stump, which, if left open, will heal as the margins contract toward one another. Failure of spontaneous open wound closure is one of the phenomena that result in a chronic wound. *Partial-thickness wounds* heal by the process of epithelialization. This occurs first by migration and then mitosis of epithelial cells (Fig. 8-4).

Mechanisms Involved in Wound Healing

Three distinct biologic mechanisms are involved in all healing processes; however, there are significant differences in the contribution of each mechanism, depending on the type of wound. It is imperative that the clinical surgeon recognize each of these separate mechanisms and the contribution of each in the wound being treated.

Epithelialization is the process whereby keratinocytes migrate and then divide to resurface partial-thickness loss of skin or mucosa. Examples include partial-thickness skin graft donor sites, abrasions, blisters, and first- and second-degree burns.

Contraction is the mechanism whereby there is spontaneous closure of full-thickness skin wounds or constriction of tubular organs such as the common bile duct or esophagus after injury.

Connective tissue matrix deposition is the process whereby fibroblasts are recruited to the site of injury and produce a new connective tissue matrix. This process is of major importance in primary wound closure, be it of skin, tendon, or intestine at an anastomosis. The cross-linked collagen and its organization in the connective tissue formed in the process provide the strength and integrity to all tissues.

Phases of Healing

The healing of an acute, primarily closed wound or a wound created by any mechanism—be it surgical intervention, trauma, chemical, friction, heat, or cold—usually follows a predictable pattern under *normal circumstances*. Chronic dermal wounds (such as pressure, diabetic, and venous stasis ulcers) or chronic parenchymal injury (such as subacute chronic hepatic fibrosis) are discussed elsewhere in this text. Under normal conditions, the phases of healing are divided into four specific events. Although they are described in a sequential fashion, they actually comprise an overlapping symphony of complex interactions. There is not a "lag phase" in the healing process. This is a misnomer propagated by several generations of authors. From the moment of injury, healing is an active, dynamic process.

Coagulation

Injury causes hemorrhage from damaged vessels and lymphatics (Fig. 8- 5). Vasoconstriction occurs almost immediately as a result of release of catecholamines. Various other vasoactive compounds, such as bradykinin, serotonin, and histamine, are released from tissue mast cells. They initiate the process of diapedesis, a passage of intravascular cells through vessel walls and into the extravascular space of the wound. Platelets derived from the hemorrhage form a hemostatic clot.

The platelets release clotting factors to produce fibrin, which is hemostatic and which forms a mesh for the further migration of inflammatory cells and fibroblasts. Fibrin is produced from fibrinogen, which is formed by the action of thrombin in the presence of thromboplastin. If the fibrin mesh is eliminated, the wound's ultimate strength is diminished.

Platelets are also extremely important because they are the first cells to produce several essential cytokines, which modulate most of the subsequent wound healing events. The subject of cytokines is discussed in detail later in this chapter.

Inflammation

The inflammatory phase is characterized by the sequential migration of leukocytes into the wound (Fig. 8-6). Within 24 h the wound is predominated by polymorphonuclear leukocytes, and then by macrophages. Although it is well known that inflammatory cells regulate connective tissue matrix repair, the specific messengers of regulation are now defined. These are the various cytokines that in the past were termed "growth factors."

Fibroplasia

It is during the phase of fibroplasia that the healing events most important to the surgeon occur. In particular, the fibrous protein collagen is synthesized. It is not only the synthesis but also the cross-linking and deposition of collagen and other matrix proteins that provide the healed wound with strength and integrity (Fig. 8-7). Within 10 h after injury, there is evidence of increased wound collagen synthesis. After 5 to 7days, collagen synthesis peaks and then declines gradually. Collagen is unique because it is only after it has been secreted into the extracellular milieu that the tissue strength characteristics are achieved following cleavage of procollagen peptides and cross-linking steps. In addition, there is significant production of ground substance within the matrix and proliferation of blood vessels.

Remodeling

The wound is an "up-regulated" process until "remodeling." At that point, acute and chronic inflammatory cells diminish gradually, angiogenesis ceases, and fibroplasia ends. The

equilibrium between collagen synthesis and collagen degradation is gradually restored (Fig. 8-8). Normally a fibrous repair is imperfect, but functional and not excessive. The complex interactions and associations of the various processes that take place during normal dermal wound healing are summarized in Fig. 8-9.

Cytokines in Wound Healing

Cytokines provide all the communications for cell-to-cell interactions and are the most exciting recent wound healing breakthrough (Table 8-1). Their potential clinical use is just beginning to unfold. Cytokines may have important pharmacologic roles in the many areas of clinical management of wound healing. For example, cytokines appear to play a role in the regulation of fibrosis, in the healing of chronic wounds and skin grafts, in vascularization, in the enhancement of bone and tendon strength after repair, and perhaps even in the control of malignancy.

Cytokines are really “wound hormones” (Fig. 8-10). They may be *endocrine*, like somatomedin or insulin-like growth factor (IGF-1), when they are secreted by one cell and then circulate in the bloodstream to reach a distant target cell. Others are *paracrine*, produced by one cell and affecting an adjacent target cell; examples include *transforming growth factor beta (TGF- β)* and *platelet-derived growth factor (PDGF)*. *Autocrine* factors are secreted by a cell and then act on receptors on the same cell. Finally, *intracrine* factors are produced by a cell and remain active in the same cell.

Cytokines regulate cell proliferation as either *competence factors*—i.e., they get the cell into G₁ phase, such as PDGF—or *progression factors*, acting to promote the cell through the proliferation cycle, such as IGF-1. Cytokines are also chemotactic, stimulating cells to migrate to the wound site. In addition, cytokines direct the cell to produce specific components needed for matrix repair, including proteins, enzymes, proteoglycans, and attachment glycoproteins.

The cytokine nomenclature is complex and somewhat confusing. Some factors are named for their cell of origin—e.g., PDGF (platelet-derived growth factor); others are named for their action—e.g., EGF (epithelial growth factor); and others are named for their first reported action—e.g., TGF- β (transforming growth factor). Several of these cytokines are of major importance in wound healing processes (see Table 8-1).

PDGF, for example, which is one actually of several growth factors derived from platelets, initiates many wound healing events and stimulates production of several other wound cytokines. PDGF's functions include chemotaxis for fibroblasts, neutrophils, and macrophages as well as smooth muscle cells. PDGF also stimulates the production of fibronectin and hyaluronic acid and may stimulate wound contraction. During the remodeling phase of repair, PDGF is thought to stimulate the production and secretion of collagenase in fibroblasts. *TGF- β is produced by a host of cells, including platelets, fibroblasts, smooth muscle cells, endothelial cells, keratinocytes, lymphocytes, and macrophages. This important cytokine increases collagen synthesis by specifically enhancing matrix gene expression and by inhibiting collagenase production and activity (Fig. 8-11). As a result, there is a significant stimulation of collagen deposition mediated by TGF- β . This cytokine thus appears to be essential for normal healing, and an excessive amount of TGF- β or its receptor may be extremely important in the pathophysiology of fibrotic states such as keloid and hepatic fibrosis.*

There is a family of TGF- β 's, and the specific isoforms may have diverse activities during wound healing. Three isoforms of TGF- β are found in mammals, and they are designated TGF- β 1, β 2, and β 3. TGF- β 1 is the most abundant of the three isoforms. TGF- β 2 is found in amniotic fluid, saliva, breast milk, and aqueous and vitreous humor of the eye. TGF- β 3 has been isolated from human umbilical cord and is the least studied of the three isoforms. TGF- β 1 and β 2 are thought to be interchangeable, whereas TGF- β 3 appears to result in reduced scarring of incisional wounds.

The *fibroblast growth factor (FGF)* family is another group of cytokines. They bind to heparin and heparin-like glycosaminoglycans. *Basic FGF (bFGF)* is a potent angiogenic factor, causes increased epithelial cell migration, and hastens wound contraction.

Epithelial growth factor (EGF) stimulates epithelial migration and mitosis. Although it has been reported to hasten reepithelialization of burn wound donor sites, these data remain questionable; in another study, split-thickness donor sites in healthy, nonsmoking, volunteers treated with EGF did not heal any faster than control sites. There is still hope that EGF will provide acceleration of partial-thickness wound healing because EGF activity may be inhibited by wound proteases. Therefore, inhibition of these proteases should accelerate epithelialization in the presence of EGF. EGF and the closely related compound *transforming growth factor alpha (TGF- α)* continue to hold clinical promise as wound healing agents.

As summarized in Table 8-1, there are several other cytokines from a variety of cell and tissue sources that probably have key roles in the complex process of wound healing. Much of the basic information about these factors has been obtained by animal and cell culture studies in which single compounds were analyzed. In vivo, of course, they interact in complex associations.

Of equal importance are the mechanisms that regulate the expression of specific growth factor and cytokine receptors on the cell surface. This area of research is being explored in a number of laboratories and should provide additional insights into the critical role of these signals during tissue repair.

Extracellular Matrix Metabolism

The extracellular matrix is a complex structure in which a number of cell types and components interact (Fig. 8-12). Collagen is the major component of the extracellular matrix of all soft tissues, tendons, ligaments, and bone. At least 19 distinct forms of collagen have been characterized, the first five major forms of which are described in Table 8-2. In addition to collagen, the extracellular matrix contains glycosaminoglycans, proteoglycans, fibronectin, laminin, and elastin. The role of these components in normal connective tissue and during wound repair is summarized in Table 8-3.

Synthesis

There are several major steps in the biosynthesis of collagen (Fig. 8-13). First, at the transcription level, the amount of messenger RNA (mRNA) for the specific collagen is tightly controlled. Once synthesized, the mRNA undergoes extensive modifications before it is ready to be translated. The next point of regulation is at the translational step, where the actual synthesis occurs on ribosomes on the rough endoplasmic reticulum.

Collagen is composed of three polypeptide chains, and each chain is formed in a very orderly sequence with a high frequency of glycine-proline-x (Fig. 8-14). Collagen is unique because it contains the amino acid hydroxyproline. Hydroxyproline, however, is not incorporated as such into the collagen chain; hydroxylation occurs to specific prolines as synthesis occurs. This step is extremely important because failure to hydroxylate proline produces an unstable collagen molecule that is degraded rapidly in the intracellular or extracellular environment. This key hydroxylation step requires several hydroxyproline and cosubstrates. A lack of ascorbic acid or oxygen will compromise collagen production and result in insufficient wound strength.

As the collagen molecules are assembled, they are glycosylated by the addition of galactose and glucose to specific hydroxylysine residues, and then the three chains are stabilized to one another by hydrogen bonds. At the amino and carboxyl terminal ends of the molecule are extra pieces termed “extension peptides”; when they are present the molecule is termed “procollagen” (see Fig. 8-14). The carboxyl extension peptides “register” and hold the molecule in position as it is moved to the cell membrane for secretion into the extracellular space. Procollagen is approximately 1000 times more soluble than the final collagen product. The extension peptides are then removed by specific procollagen peptides as the molecule is secreted. Evidence suggests that the breakdown products of the procollagen extension peptides are taken up by the cell and down-regulate collagen expression.

Collagen is different from all other proteins because it undergoes several modifications once it has reached the extracellular environment. Here, collagen cross-linking occurs to form fibrils and fibers of collagen. The enzyme responsible for this step, lysyl oxidase, requires copper and can be inhibited by collagen cross-link inhibitors such as β -aminopropionitrile (BAPN) and α -aminocaproic acid. These collagen cross-link inhibitors have been used in animal studies and to a limited degree in human beings in an attempt to prevent excessive collagen deposition and subsequent adhesions and fibrosis. The results have been modest.

Degradation

For normal wound healing, collagen must be degraded as well as produced. The breakdown of extracellular components is initiated by very specific enzymes termed matrix metalloproteinases (MMPs), synthesized by a variety of cells including inflammatory cells, fibroblasts, and epithelial cells. MMP-1, 8, and 13 initiate degradation of the collagen molecule by splitting it into specific three-quarter and one-quarter fragments termed the TC_A and TC_B fragments. After this initial split, other nonspecific proteases can further degrade the pieces into peptides and eventually amino acids. The MMPs exist in an inactive or zymogen form that must be activated by other proteases such as plasmin. Once the MMPs are activated, they can be inhibited by complexing with the plasma and tissue protein α -2 macroglobulin. The MMPs can also be inhibited by forming a complex with tissue inhibitors of metalloproteinases (TIMPs). The pathway responsible for extracellular matrix degradation is thus complex and rigorously controlled.

Ground Substance

The precise role of ground substance, or proteoglycans and glycosaminoglycans, in wound healing remains unclear. Recent evidence suggests that ground substance has more importance in the healing process than recognized previously. Proteoglycans are composed of glycosaminoglycan subunits attached by covalent bonds via a serine residue to a protein core. They form a “bottle brush” structure and, as macromolecules, occupy a significant amount of space in the extracellular matrix. They function as molecular “shock absorbers” in combination with cartilage, provide for moisture storage, and also sequester cytokines. Some evidence suggests that after tissue injury, when ground substances are degraded, the bound cytokines are released to provide initial signals to facilitate the repair process. One specific glycosaminoglycan, hyaluronic acid, is most unusual because it is not sulfated and not bound to protein. It has a very high molecular weight and provides a fluid environment, thus facilitating rapid cell movement, cell differentiation, and extracellular matrix organization. Hyaluronic acid has an early and transient appearance soon after injury in the adult and has a much longer persistence in fetal skin and fetal wounds. Perhaps the application of hyaluronic acid to adult wounds could transform healing into the regeneration-like response seen in the fetus.

Wound Contraction

Contraction is one of the most powerful mechanical forces in the body. The precise biologic mechanisms of the process remain a controversial subject. Surgeons have always found the process of wound contraction both an ally and a foe. Even the ancients knew that open skin wounds healed if kept clean and protected with a dressing. During the healing process, the skin margins move in until they meet one another to provide a healed wound.

In the Civil War, more extremity amputations were performed than at any other time in human history. The wounds were left open because of the recognition that closure would usually lead to sepsis and death. Although healing often took months, very large, dinner-plate-sized, above-the-knee amputation sites would heal with a sturdy stump for a weight-bearing prosthesis. Similarly, today hidradenitis suppurativa in the groin region, which can be very difficult to heal, is often managed best by total excision of all involved skin in the groin and perianal area and then allowing the wound margins to come together spontaneously until the wound is healed (Fig. 8-15). In both instances, contraction is the ally of the surgeon, allowing closure without added morbidity.

However, in many instances, *contraction* (the normal, active biologic process) results in a *contracture*, a fixed deformity that is an aesthetic and functional disability to the patient. Most dramatic are contractures of skin and hollow organs. Loss of skin secondary to burn injury or trauma may result in a contracture as the skin edges are drawn together to heal a defect. In addition, whenever redundant skin is absent, a contracture will result. This is especially true over flexor joint surfaces such as the neck (Fig. 8-16) or the volar surfaces of digits. But the process is not limited to the skin. Any type of injury to hollow organs such as the esophagus or common bile duct may trigger the contractile process, resulting in a contracture that mechanically blocks the function of the hollow organ.

Mechanism

Alexis Carrel noted that open animal skin wounds healed by a process wherein the wound margins would remain open for a few days—a lag or plateau phase—after which followed a rapid rate of closure. The kinetics of the process of wound closure can be seen in the data shown in Fig. 8-17. There were and still remain questions as to whether the cells responsible for these powerful forces of tissue movement are located within the central granulation tissue of the wound or at the wound margins (picture-frame theory). Experimental studies by Gross have demonstrated that removal of the central granulation tissue did not alter the completeness of wound closure or the final pattern of the scar. The precise mechanisms responsible for wound contraction are not fully understood.

In the early 1970s researchers noted the presence of fibroblast-like cells in contracting open skin wounds that had smooth muscle components in the cytoplasm as well as fibroblast characteristics. They termed these cells *myofibroblasts*. They also found that when strips of open wound granulation tissue were placed in a water bath, they contracted in the presence of smooth-muscle agonists and relaxed in the presence of smooth-muscle antagonists. Furthermore, myofibroblasts have been identified in a number of contracted human tissues such as Dupuytren's contracture, burn contractures, and contractures of capsules around silicone breast implants. These cells are at their peak during and after the process of wound contraction (Fig. 8-18).

Some researchers postulate that the extracellular matrix may be as important as the cell type within the contracting tissue. The work that supports this hypothesis is based mainly on experiments done in vitro in a tissue culture system using a collagen gel lattice. Perhaps the varying evidence points to a contribution of both elements, i.e., contracting cells acting on a matrix susceptible to contraction.

All attempts to use pharmacologic agents to control contraction of wounds have failed. For example, some investigators attempted to inhibit open wound contraction with smooth-muscle inhibitors such as thiphenamil hydrochloride (Trocinate), which was successful only as long as the agent was present on the wound surface. Splinting a wound open will not prevent contracture. As soon as the splint is removed, the powerful biologic forces of the process place the wound margins in just the position they would have been had the splint not been placed.

Clinical Approaches

There are several helpful clinical principles for the surgical correction of contractures. If there are signs of inflammation remaining, the surgeon must be aware that various maneuvers, such as skin graft or Z-plasty, can result in recurrent contractures because of the various cytokines made by these cells and the myofibroblasts still present in the wounds. Before operation the surgeon should make a judgment as to the inflammatory status of the tissue. Mature scar is soft and pliable, whereas immature scar may be stiff, indurated, hypertrophic, and even tender. Remember that immature scar still has inflammatory cell components and residual myofibroblasts, which facilitate contracture of the bed under any skin graft that may be used in an attempt to correct the deformity. Often it is preferable to correct the defect by placing a flap that contains both skin and subcutaneous tissue and in some cases muscle. Because a flap is made of composite tissue and supplies the defect with all of the components of soft tissue, contraction is rare.

In correcting a mature contracture, a skin graft may be used to fill the defect. For reasons unknown, the open wound contracts less after the placement of a full-thickness graft than after the placement of a partial-thickness graft (Fig. 8-19). It is not a matter of graft thickness but of whether the graft is *full* or *partial*. In both instances, it is advisable to splint the repaired wound in a fully open position. It appears that splinting is required until all myofibroblasts and inflammation are gone from the wound, which may take as long as several months. Splinting time is highly variable, however, and must be determined by clinical judgment rather than science.

Epithelialization

All surfaces exposed to the external environment are covered by epithelial cells. Skin is an example of epithelialization, but mechanisms of epithelial repair are similar throughout the body. The outer layer of skin, the epidermis, is composed of a stratified squamous epithelium that protects the body from fluid loss, bacterial invasion, electromagnetic radiation, and general trauma. Normally the epidermal thickness is maintained at a constant level. Cells in the basal layer of the epidermis divide and migrate superficially and mature in the process. They become anuclear and die as they reach the surface and slough or desquamate, by which time they are really acellular masses of keratin. Although the term *keratinocyte* is used often interchangeably with the term *epithelial cell*, the epithelium is composed of several types of cells. The majority are keratinocytes, but the epithelium also contains other types, such as Langerhans' cells, which are involved in immunologic responses.

The collagen-rich dermis, and not the epidermis, provides all the strength attributed to skin. The epidermis, however, provides the barrier that protects the internal milieu of the host from the external environment. The basement membrane of the basal lamina provides structural support for the epidermis and attaches epidermis to dermis (see Fig. 8-12). It is a thin, glycoprotein-rich layer with complex layering and structures. A true adult-type wound healing response with classic inflammatory changes only occurs when the basal layer has been violated.

Mechanisms

Partial-thickness wounds heal by the process of epithelialization (see Fig. 8-4). There are two major phenomena in the process of epithelialization: *migration* and *mitosis*. After epithelial destruction, a blood clot is formed; it dries and forms a scab on the exposed dermis, thus protecting the dermis. Epithelial cell migration then begins the reparative process and is independent of epithelial mitosis. Migration of cells is the dominant process in epithelialization. Under experimental conditions, blocking cell mitosis has no significant effect on cell migration or closure of the wound. These migrating cells are derived from the margins of the wound and from hair follicles and sebaceous glands within the dermal base of the wound (see Fig. 8-4). The more superficial the wound, the more rapid the process of reepithelialization. Very superficial wounds such as minor burns and abrasions that do not penetrate the basement membrane heal by regeneration. Deeper dermal burn wounds or partial-thickness skin graft donor sites that penetrate the basement membrane also may heal by epithelialization, but the process requires a longer time. In addition, the result may be unsatisfactory because the inflammatory or healing response is initiated once the basement membrane is violated (Fig. 8-20).

Regardless of the type of injury, the basal layer of epithelium and the epithelium in the deeper hair follicles and sweat glands is where migration is initiated (see Fig. 8-4). One can observe these cells as they change morphologically. They flatten out and send out cytoplasmic projections into the surrounding tissue (Fig. 8-21). These cells also lose their attachments to their neighboring basal cells and begin to migrate, with the leading cells phagocytizing debris and blazing a path for the keratinocytes migrating behind them. The epithelial cells, which secrete proteases such as collagenase and plasminogen activator, do not migrate as single cells but in a sheet of cells. Some researchers have proposed that the keratinocyte sheets pile up at the leading edge and that cells tumble over the top in leapfrog fashion by a process called *epiboly*. When there are not enough cells for further migration, then mitosis begins.

Some of the biochemical mechanisms that control these processes have been defined. The blood and tissue fluids contain fibronectin and vitronectin—both of which support epithelial cell migration. Moreover, several growth factors stimulate keratinocyte migration and mitosis. These include bFGF, PDGF, TGF- α and EGF. By contrast, TGF- β inhibits epidermal cell proliferation but stimulates motility. Although the mechanical aspects of epidermal cell motility are not defined clearly, it is known that these cells contain a cytoskeleton and move by an actin-myosin contractile system. Once the surface is covered, the epithelial cells revert to their normal phenotypic behavior, with intercellular and basement membrane attachments. This reversion to a normal state may be a key to understanding epithelial cancer (Fig. 8-22).

The movement of epithelial sheets is most important in the healing of partial-thickness wounds. Clinically this process is enhanced by keeping the surface moist rather than dry. Nature's scab may be satisfactory, but appropriate nonadherent dressings that will keep the wound moist are of vital importance and can enhance significantly the process of epithelialization (see the section Wound Dressings later in this chapter). Growth factors also will be used in the near future to accelerate these processes. In addition, agents that enhance the detachment of cells and enhance the epibolic mechanisms will be added to the therapeutic regimen.

Nutrition

Inadequate nutrition is devastating to the healing process. If caloric protein intake stops for a mere 24 h, collagen synthesis ceases. Inadequate nutrition inhibits the immune response, and opsonization of bacteria is ineffective. Several dietary insufficiencies have been described.

Ascorbic acid is essential for human beings, and its lack is the most common cause of wound healing deficiency. The diseased state associated with a lack of vitamin C (ascorbic acid) is known as *scurvy*. Historians have reported that old scars in sailors would spontaneously open years after healing when the sailors were at sea for many months and without fresh fruits that contain large quantities of ascorbate. Long before biochemical analysis for ascorbate was available, the British admiralty recognized that scurvy could be avoided if all ships carried adequate supplies of ascorbate-containing fruits such as limes—hence the term “Limey” became affixed to British throughout the world.

The biochemical function of vitamin C is well known, and secondary functions relating to collagen gene expression have also been described. As mentioned in the discussion of collagen metabolism, ascorbate is a cofactor in the hydroxylation of proline to form the amino acid hydroxyproline during the synthesis of collagen. Ascorbate is essential for the addition of molecular oxygen to form the hydroxyl group of hydroxyproline. In human beings, thermally unstable collagen is produced if dietary ascorbate is insufficient. Old healed wounds thus tend to disrupt preferentially compared with the normal surrounding skin for two reasons. First, the scar is never as strong as surrounding skin. Second, there is more collagenase activity in normal scar tissue than in normal skin. Hence, in the scorbutic patient, scar tissue breaks down before there is breakdown of the normal skin. The problem is of more than mere historic and biochemical interest. Ascorbate deficiency is found commonly after major trauma and is more common than generally realized in malnourished people in the United States, particularly among the lower socioeconomic classes. Approximately 20 percent of patients seen in an inner-city emergency department may be ascorbate deficient. The mechanisms responsible for acute ascorbate deficiency after major trauma are not clear. There may be rapid sequestration of ascorbate in organs, or there may be renal loss, degradation, or lower absorption of this vitamin. Aggressive replacement of vitamin C should be undertaken immediately after major trauma to prevent wound healing complications. Although 60 mg/day is the current recommended daily allowance, up to 1 g of vitamin C per day has been recommended in trauma patients.

Iron in trace amounts is needed for prolyl hydroxylation. Calcium and magnesium are required for collagenase activity and protein synthesis in general. All the essential amino acids required for protein synthesis are needed for wound healing. Supplementation of the diet with increased levels of arginine appears to enhance the wound healing response, but the precise mechanism is not fully understood. An adequate supply of oxygen is essential for wound healing, not only for the hydroxylation of proline and lysine but also for all of the energy required for bacterial killing and cell viability. Some speculate that many chronic wound healing problems and perhaps impaired wound healing after trauma can be treated

effectively by increasing tissue oxygenation. There is evidence that if blood volume is low, then tissue oxygen will be low despite normal arterial blood gases. Some claim that hyperbaric oxygen enhances healing of chronic wounds and burn wounds, but there are few objective studies to support the concept that delivery of oxygen to the pathologic tissues enhances healing in a sustained fashion.

Immunosuppression

Only a small number of immunosuppressed patients actually manifest clinical wound healing problems. Acquired immunodeficiency syndrome (AIDS) is manifested by myriad signs and symptoms. A direct relationship between the leukocyte defect and healing, however, has not been reported. The wound complications found in AIDS patients are secondary to other manifestations of the disease—for example, various skin lesions such as Kaposi's sarcoma or infected traumatic wounds occur. Chronic wounds in these patients should undergo the same therapy as any other chronic wound (see the section Chronic Wounds later in this chapter).

Chemotherapeutic anticancer drugs inhibit healing. It is often unclear whether the wound healing problems are caused by the drugs or by the malignant tumor. Malignancies deplete nutrients and also inhibit wound healing directly. In animal models, chemotherapeutic agents have been demonstrated to be very harmful to healing. Various growth factors such as PDGF and TGF- β are able to prevent the harmful effects of these drugs on collagen metabolism.

Genetic Disorders of Connective Tissue Metabolism

Research in molecular biology is constantly providing new data on wound healing. Most exciting in recent years is information on a host of genetically controlled defects of collagen and glycosaminoglycan metabolism that result in both subtle and lethal diseases of connective tissue. There are probably hundreds of variants awaiting characterization and chromosomal identification. All of these genetic defects in connective tissue metabolism result in a poor wound healing response.

Osteogenesis Imperfecta

Osteogenesis imperfecta (OI) is a congenital form of osteopenia due mainly to mutations in the genes for Type I collagen. There are four types of OI, ranging from mild to lethal manifestations (Table 8-4). Patients with OI present a particular problem to the surgeon because of (1) the increased propensity of the bones to break under minimal stress, (2) dermal thinning, and (3) increased bruisability. Scarring is usually normal in these patients, and the skin has normal extensibility.

Patients affected severely with OI may have difficulty with diaphoresis, which is not only unpleasant but also, coupled with fasting before surgery, may lead to dehydration and fever. The administration of parenteral fluids preoperatively may circumvent these problems. Children with OI also have a higher incidence of hernias, especially umbilical and inguinal. These can be corrected successfully with surgery.

Ehlers-Danlos Syndrome

Ehlers-Danlos syndrome is a genetically, biochemically, and clinically distinct group of collagen disorders characterized mainly by joint laxity, skin hyperextensibility and fragility, poor wound healing, and vascular rupture. At least ten types of this disorder have been distinguished. The enzyme or biochemical defects are known only for a few of the types. Each type presents a distinct challenge to the surgeon. These challenges encompass vascular complications such as arteriovenous fistulas, true and false aneurysms, varicose veins, bleeding ulcers, arterial rupture, and defective platelet adhesion. Invasive procedures such as angiography and surgery carry a very high morbidity and mortality in these patients. Rupture or dissection of a major artery may occur spontaneously in Ehlers-Danlos patients after minor trauma or with angiographic manipulation. Due to connective tissue weakness, adolescent males are at an increased risk during their normal growth and development as a consequence of the increased stresses of physical activities common in this age group. Due to connective tissue weakness, adolescent females are at a higher risk with the hormonal changes of menstruation. There are probably other forms of the syndrome that will be identifiable with modern molecular biological techniques. The types of Ehlers-Danlos syndrome are summarized in Table 8-5.

Marfan's Syndrome

Marfan patients are characterized by tall stature, arachnodactyly, lax ligaments, myopia, scoliosis, pectus excavatum, and often dissecting aneurysms of the root and ascending portions of the aorta. Some of these patients have defects in collagen structure, and some have abnormal fibrillin in their elastin. Spontaneous rupture of the aorta can cause sudden death. All these diseases may make surgery more difficult and wound healing more complicated.

Epidermolysis Bullosa

Epidermolysis bullosa is characterized by blistering and ulcerations. Some forms of this defect are thought to be caused by excessive production of matrix metalloproteinases by fibroblasts. Other forms are caused by abnormal matrix adhesion to the epidermis and associated basement membranes, resulting in tissue separation and blistering of skin after minimal trauma. The majority of these ulcers heal spontaneously, but in the more severe forms the epithelium cannot regenerate adequately, and chronic inflammation and scarring ensue. This genetic disease creates a challenge for the surgeon in several realms. Alimentary tract surgery exhibits poor healing that is impaired by stenosis and strictures in many cases. Dermal incisions and tissue injury must have meticulous skin care to limit the amount and severity of blistering in the traumatized tissue. Phenytoin decreases collagenase activity in human skin fibroblasts and has been used to treat patients with recessive dystrophic epidermolysis bullosa.

Clinical Importance

Patients with these genetic disorders provide a challenge for the surgeon. Because of the relatively rare occurrences of these genetic defects, any one surgeon may see no more than one or two cases in a lifetime. Despite this low exposure rate, it behooves the surgeon to be aware of the physical and clinical signs of these diseases so that disaster during surgery can be avoided. A thorough history and physical examination, with emphasis on the family history, may alert the surgeon to a potential problem, which can then be successfully overcome. At times, preoperative diagnosis is impossible.

Systems and General Factors that Affect Wound Healing

Inadequate or poor-quality matrix production may be secondary to a variety of causes, such as malnourishment or various drugs that inhibit cell proliferation and protein synthesis. Table 8-6 is a list of some of the many factors that may alter wound healing. Infection is the number one cause of impaired wound healing, and many factors can contribute to infection (Table 8-7).

SPECIFIC WOUND HEALING PROBLEMS

Gastrointestinal Tract

There has been progress in our understanding of healing and repair in the gastrointestinal tract over the past 10 years. This is a result of efforts to determine the cellular and connective-tissue elements involved in normal intestinal architecture and in the fibrotic responses that characterize many chronic inflammatory processes in the intestine.

Anatomy

The gastrointestinal tract is divided into several distinct layers (Fig. 8-23). The inner mucosal layer is for absorption, and the outer muscularis mucosae layer functions for motility. They are wrapped in a strong serosal layer, which is an extension of the peritoneum.

Unlike the skin, the mucosal epithelium is only one cell thick and renews itself about every 8 days. Beneath the epithelium is a basement membrane composed of Type IV collagen similar to the basement membrane of skin. Beneath this is the lamina propria, which is made up of collagen Types I, III, and V and elastin with an array of specialized cell types, including various inflammatory cells. The outermost layer of the mucosa is a very thin layer of smooth muscle, the muscularis mucosae, which contributes to gut motility. The submucosa separates the mucosa from the muscularis propria and appears to attach these two important layers to one another. It is composed of several collagen types in a ratio similar to that found in human aorta.

The muscularis propria is densely packed smooth muscle with a predominance of Types I and III collagen and a small amount of Type V. This collagen serves as an intramuscular tendon, or the source of strength through which a force can be transmitted through the smooth muscle cells. This is supported by the fact that hypertrophy of the muscularis propria also increases the collagen content of the muscle.

Injury and Repair

The quality of the intestinal repair process and its end result are determined by two important factors: the depth of the injury and the chronicity of the injury. Inflammation in the mucosa without epithelial injury, as seen in gastritis, is not associated with a repair or healing process. Once the inflammation subsides, the erythematous appearance of the mucosa seen at endoscopy returns to its pale salmon-pink appearance with easily identified ramifying blood vessels. If the epithelium is injured, a rapid restitution process activates, and the epithelial lining is restored within hours (Fig. 8-24). These lesions, termed erosions, are often induced by acid/peptic digestion or aspirin or chemotherapeutic agents in the stomach. There is no mesenchymal cell response if the injury is confined to the mucosal layer and does not penetrate the muscularis mucosae, even if the injury is chronic.

When the injury does penetrate into the submucosa, a mesenchymal cell repair response is evoked. Smooth muscle cells from either the mucosal or deeper muscle layers are significant in this response, migrating into the area of injury and laying down reparative collagen. This lesion is called an ulcer and would characteristically have a plug of smooth muscle cells and collagen at its base.

If the injury in the submucosal layer is acute, the reparative process abates, the collagen laid down is resorbed, and the normal architecture of the intestine is preserved. If the submucosal injury is chronic, however, scar tissue accumulates, the mucosa loses its compliant movement over the muscle, the bowel wall thickens, and stricture and obstruction can ensue (Fig. 8-25). In this way, any chronic lesion involving the intestinal submucosa can lead to stricture.

Crohn's disease is a chronic idiopathic inflammatory bowel disease that affects the terminal ileum and cecum most frequently but can involve any part of the gastrointestinal tract. It is characterized by inflammation in the submucosa rather than the mucosa and often extends from mucosa to serosa (transmural). Inflammation leads to collagen deposition and contraction, which cause stricture and symptoms of intestinal obstruction. Biochemically, there is increased collagen Type I and Type V content in the involved bowel compared with normal. This is in contrast to ulcerative colitis, in which the bowel tends to have thinning of its matrix and to perforate rather than get thicker and develop strictures. The healing process of ulcerative colitis is very different from that of Crohn's disease. In ulcerative colitis the inflammation is confined to the mucosa and does not extend into the submucosa.

Human intestinal smooth muscle cells isolated from normal human jejunum have been used to clarify further the pathogenesis of bowel healing. Demonstrating their differences when compared with normal skin, the bowel cells in vitro produce twice the collagen as skin fibroblasts. Moreover, *TGF- β enhances collagen synthesis by intestinal smooth muscle cells. Although corticosteroids inhibit collagen production by dermal fibroblasts, they are not effective in down-regulating collagen synthesis by human intestinal smooth muscle cells. Radiation injury* is commonly seen in the alimentary tract and terminates in extensive and progressive submucosal, muscularis, and serosal fibrosis with striking hyalinization of the

accumulated collagen. Resection is required in some patients but is hazardous because bowel anastomosis in irradiated tissue is associated with anastomotic leak and fistula formation.

Lye ingestion is a common cause of esophageal tissue injury. The mucosa is penetrated, and hence fibrosis and stricture occur. Although parenteral corticosteroids have proved efficacious in animals, there has never been a well-controlled study in human beings. Repeated dilatation of the esophagus may abate symptoms of obstruction, but repeated contractions often make serial dilatations a lifelong necessity.

Healing in the Gastrointestinal Tract

The same basic process of repair occurs with anastomotic healing in the gastrointestinal tract as occurs in skin. The same factors that inhibit development of tensile strength in skin do the same in the gastrointestinal tract. But the gastrointestinal tract is a unique tubular structure. It is closed with sutures or staple devices and must then rely on the anastomosis to provide bowel integrity until the anastomosis has developed sufficient tensile strength to prevent disruption. The major complications of intestinal anastomoses are anastomotic leakage and actual bowel wall disruption, which are associated with a significant morbidity and mortality. Although clinical complications of anastomotic leak occur in 2 to 18 percent of patients, up to 50 percent of anastomotic sites leak early, as demonstrated by contrast radiographic studies. Early bowel leak cannot be associated with insufficient collagen metabolism but must relate to mechanical anastomotic failure, bowel ischemia, or failure to obtain a mucosal seal. However, most studies have focused on collagen metabolism. During the first few days after anastomosis, there is significant turnover of collagen not only at the anastomotic site but in the adjacent bowel wall. Reduction in collagen content extends a significant distance from the actual anastomotic site. The significance of these data versus mucosa barrier defects remains to be clarified.

Skin

More wound healing data come from skin than from any other organ system. Most of the information gained from the study of skin in animal models and in human beings can be applied to the process of tissue repair in all other organ systems.

Keloids and Hypertrophic Scars

Keloids and hypertrophic scars are both abnormal healing processes that occur after injury from trauma or surgery, but they differ clinically and biochemically. Hypertrophic scars remain within the boundaries of the original wound and almost always regress over a period of time (Fig. 8-26). By contrast, keloids extend beyond the boundaries of the original wound and usually do not regress (Fig. 8-27); they usually recur after excision unless additional therapy is provided.

Normally there is an equilibrium between collagen synthesis and collagen degradation in skin and normal scar tissue. By contrast, both keloids and hypertrophic scars are characterized histologically by an overabundant deposition of collagen. However, the rate of collagen synthesis in keloid tissue or keloid fibroblasts in culture is significantly greater than in normal skin and normal scar tissue. The keloid tissue contains more soluble collagen and has a greater water content. Fibroblasts isolated from keloids continue to produce more collagen, through up to 40 passages in cell culture, than hypertrophic scars or normal skin. The equilibrium may be disrupted further by the collagenase inhibitor α_2 -macroglobulin, which is abundant in keloid and hypertrophic scar tissue. Both increased collagen synthesis and decreased degradation seem responsible for the increased collagen deposition in keloid and even hypertrophic scar tissue. Molecular biologic studies applied to these earlier findings seem to confirm the routine biochemistry. There is increased messenger RNA for Type I collagen, which suggests that these lesions result from abnormal regulation of collagen production at the level of transcription.

Growth factors may be important in the regulation of these lesions. Some studies suggest that cultured keloid fibroblasts produce increased amounts of cytokines compared with normal skin fibroblasts and that keloids may contain increased concentrations of TGF- β . In addition, the keloid fibroblast produces significantly more collagen when exposed to TGF- β in cell culture. Neutralizing antibodies to TGF- β placed in primary guinea pig wounds at the time of closure resulted in scarless healing. These observations suggest that inhibition of TGF- β may make a significant contribution in the clinical control of keloids and hypertrophic scars. Such animal studies require further validation before this basic science information can be applied in the clinical setting.

At present the clinical treatment of keloids and hypertrophic scars is not consistently effective. Surgical excision should be done only after careful consideration. Hypertrophic scar tissue will regress usually without operation. Keloid tends to recur after excision. There may, however, be clear indications for surgery in both groups of patients. In some patients, keloid is very disfiguring, and this alone may be reason enough to excise the lesion. Excision may be used for debulking so that pharmacologic agents can be used to control the abnormal scar. In other patients excision is indicated to improve function. All patients must be made aware of the high risk of recurrence and the importance of careful follow-up.

As in any disease process that is not understood clearly, many treatments have been advocated for keloids and hypertrophic scars. Perhaps the most popular and effective is the intralesional injection of triamcinolone, which is a long-lasting synthetic glucocorticoid. This treatment makes the lesion softer and often smaller, and it may be the only treatment required for small lesions. It also often relieves the burning, itching, and pain associated with keloids and hypertrophic scars. If these somatic symptoms are the patient's complaints, then corticosteroids may be all that is required regardless of the size or appearance of the lesion. The authors recommend triamcinolone 40 mg/mL, never using more than 2 mL every 6 to 8 weeks in adults to avoid any systemic effects. Complications include local atrophy of skin and subcutaneous tissue, which may be severe. Telangiectasia may appear locally, and there may be depigmentation in dark-skinned patients. Pregnant women should never be given triamcinolone because of the remote possibility that a birth deformity may be related to the use of corticosteroids. We do not recommend the use of a "Dermajet" to deliver triamcinolone through the epidermis to the keloid in the underlying dermis. Triamcinolone may be used in conjunction with surgery. Some inject triamcinolone into the wound margins at the time of closure after excision, and others begin treatment at various times after operation. The newer "supersteroids" such as clobetasol propionate (Temovate) may be effective topically and can be used with intralesional therapy.

Keloids are known to contain a significant number of mast cells and have increased histamine content as compared with normal skin or scar. As a result, these patients often suffer severe itching. An oral antihistamine is often helpful to control this symptom. Several other topical devices have been used to treat keloids and hypertrophic scars, including a pressure dressing, steroid-impregnated tape, and silicone sheets. There are no valid data indicating that any of these methods are efficacious. Recent data suggest that pressure garments and silicone sheeting may increase skin temperature enough to increase collagenase activity, which may reduce the bulk of the scar. Silicone sheeting is expensive and until its efficacy is validated, clinicians should avoid the expense. Simple pressure or tape over the scar is adequate. Radiation therapy is mentioned only to be condemned. There have been no long-term follow-up studies reported, and the late development of skin cancer after radiation constitutes a potential hazard.

Factitious Wounds

Factitious lesions are much more common than one might expect. The surgeon should always be on the alert for factitious diseases of a surgical nature. Beware of any patient who has had multiple surgical procedures, especially patients with some relationship to the health care field. When examining patients, note the pattern of scars, accessibility to self-mutilation, "la belle indifférence." Confrontation will often lead to the patient's leaving—never to return. Both diagnosis and treatment are extremely difficult.

Marjolin's Ulcer

Any nonhealing wound in an area of previous trauma may represent squamous cell carcinoma, termed Marjolin's ulcer. These virulent malignancies arise from old areas of trauma. Although the causation may be multifactorial, they appear to arise because the dense scar of the lesion does not allow normal immunologic surveillance of the area by the host, and the host cannot destroy the malignant transformation within the scar.

Tendon

Tendons link muscle to bone and thereby allow muscular force to exert motion. They are composed mainly of Type I collagen, with a significant amount of proteoglycan, which arranges and regulates the size of the collagen fibrils. If fetal muscle development is prevented, tendons do not develop. When tendons cross the concavity of a joint such as finger or wrist and do not insert immediately distal to the joint, they pass under fibrous structures called pulleys. Specialized cells located in the pulleys and elsewhere lubricate the tendons. In the areas where tendons are compressed, they receive blood supply from segmental vessels in mesotendon structures called vincula (Fig. 8-28).

There are some unique factors in flexor tendon healing that are important clinically. The flexor tendons within fibrous flexor sheaths of the digits present the surgeon with a healing problem so challenging that some termed the area "no-man's-land." Surgical results in the past in this area were generally poor. This is because healing must occur in the sutured ends of the tendon as well as in the sheath itself. Both are collagenous structures. When they heal together as a unit rather than separately, tendon function is lost because gliding is impeded by the scar between tendon and sheath. Surgeons have learned to use meticulous technique and early mobilization in "no-man's-land" to obtain good functional results. A major contribution to the functional healing of tendon within a fibrous flexor sheath has been the revelation that the segmental blood supply to the tendons through the vincula is extremely important in promoting healing with less restricting scar. Destruction of the vincula leads to ischemia and scarring, just as would occur in any ischemic tissue. Tendon nutrition is also supplied by synovial diffusion. Experiments have shown that, deprived of a normal blood supply, tendons still heal, obtaining nutrition from extrinsic diffusion.

There is excellent experimental and clinical work to show that early motion provides the stress forces to lengthen and remodel scar. Motion cannot prevent collagen deposition, but it can lengthen the scar and allow enough motion for functional gliding of tendon. Motion may also enhance lubrication from synovial fluid. Developments in flexor tendon surgery have focused on designing stronger, multistranded tendon repair techniques to withstand more aggressive early motion. While early passive motion has been used for a decade or more, early active motion requires stronger surgical repairs and has shown significantly improved results in motivated patients.

Experiments to enhance tendon healing include electrical stimulation, nonsteroidal anti-inflammatory agents, and growth factors, but there is no evidence that these treatments are effective. Various pharmacologic agents have been used experimentally in an attempt to reduce tendon adhesions. Hyaluronic acid holds some promise. The use of proline analogues has not been successful. β -Aminopropionitrile was used successfully in human beings undergoing flexor tendon grafting, but side effects led to termination of the study.

Bone

Fracture repair involves physical, biomechanical, and biochemical factors that interact in a process that often leads to a healed bone that is normal in form and function and indistinguishable from its preinjured state. Grossly, a fracture consists of a gap or break in a bone with or without displacement. This leads to a loss of structural support with subsequent instability and deformity based on the amount of displacement. Inflammation occurs with hematoma formation, edema, and pain. This early inflammatory phase is associated with a relative tissue hypoxia at the site of the fracture, which contributes to a transformation of surrounding osteoprogenitor cells and migration of hematogenous cells into the fracture site. These events provide the fracture with the platelets, monocytes, neutrophils, fibroblasts, osteoblasts, and osteoclasts needed to accomplish fracture repair. At the same time, new vascularization occurs, with extensive blood vessel ingrowth into the fracture.

Over the course of several weeks these changes lead to the formation of the soft callus, a local fibrocartilaginous splint of granulation tissue and cartilage that gives the fracture some stability. In a process that takes a total of 6 to 8 weeks from the time of injury in an uncomplicated fracture, the soft callus is transformed into bone by endochondral ossification. In this process the cartilage of the soft callus is ossified as osteoblasts originating in the periosteum and the medullary canal move from the periphery of the fracture inward creating new bone. As this process is completed, pain diminishes and union occurs. Over several years remodeling of the bone occurs as its structure is altered until its

appearance approaches a normal uninjured bone (Fig. 8-29).

The process of fracture repair is different when rigid internal fixation with plates and screws are used to immobilize the injury. When the bone ends are in exact opposition and there is no motion at the fracture site, no soft callus forms. Instead, there is direct bone-to-bone healing across the injury without endochondral ossification.

Delayed union or nonunion are failures of fracture repair. Many local and systemic factors are associated with these complications, which occur in approximately 5 percent of long-bone fractures. The local factors that seem to be the most important are the site of the fracture, presence and degree of soft-tissue injury, bone loss, inadequate reduction, inadequate immobilization, infection, previous radiation, malignant growth at the fracture site, and poor blood supply to the fracture. When the diagnosis of delayed union or nonunion is made or when a patient has a fracture known to be associated with these complications, treatment can be altered to enhance bone repair. This involves treatment of one or more of the local factors listed above, e.g., performing a muscle flap procedure to cover an open tibial fracture with extensive tissue loss, or bone grafting a fracture with bone loss.

Bone grafts can be used to treat established nonunion or to replace missing pieces of the skeleton. The healing of bone grafts involves many of the mechanisms present in fracture repair, with important differences. Osteogenesis and osteoconduction are the primary mechanisms by which bone grafts heal. Osteogenesis is the formation of new bone by cells that survive in the graft. This mechanism occurs primarily in vascularized bone grafts, such as fibula flaps, in which the blood supply to the graft is maintained or reconstituted with a microvascular anastomosis. Osteoconduction is the process by which blood vessels and cells from the surrounding tissues grow into the bone graft, which acts as a scaffold for the laying down of new bone as the dead bone of the graft is resorbed. Also termed "creeping substitution," this is the primary mechanism by which corticocancellous bone grafts, such as those from the iliac crest, heal.

Osteoinduction is the transformation of local, undifferentiated cells into bone-forming cells. Although this mechanism may participate in bone graft healing, it is a topic of active research in fracture repair and dentistry. This area of research has grown out of the increased understanding of the biochemical basis of the many cellular processes discussed earlier and the desire to control these processes to improve healing in fractures prone to delayed union and nonunion and to improve alveolar ridge contour in patients who need dentures.

Many proteins have been found to influence bone healing in experimental models. Bone morphogenetic protein (BMP), first described by Urist, was the first to be studied extensively and has been found to induce perivascular connective tissue cells to become osteoprogenitor cells. BMP's smallest active subunit is a protein that has been termed osteogenic protein-1 (OP-1). OP-1, a member of the TGF- β family of cytokines, accelerates the healing process in bone in both fracture and skull-defect models. Other cytokines, from the PDGF and FGF families of growth factors, also have effects on bone healing. As in other areas of wound healing research, pharmacologic manipulation of the process of bone healing should become a reality in the near future.

Cartilage

Cartilage of joints, unlike the bone that surrounds it, has little propensity to heal. Superficial injuries to articular cartilage cause minimal inflammatory response and are totally dependent on the chondrocytes of the cartilage to heal. Because the chondrocytes usually do not respond with enough proliferation and synthesis of the necessary proteoglycans and collagen for repair to occur, a persistent structural defect in the joint surface usually remains. Whether or not traumatic arthritis develops depends on the ability of the remaining joint cartilage to survive under increased stress.

Deep injury to articular cartilage that goes into the subchondral bone produces an exuberant inflammatory and healing response. The bone of the subchondral area is a source of reparative cells. A fibrocartilage intermediate is formed that creates hyaline cartilage over time. These properties are the basis for abrasion of severely injured joint cartilage and passive motion while healing is under way.

Cartilage grafts are used extensively in reconstructive and cosmetic surgery of the nose and ear. These grafts maintain their structure with minimal resorption over time. Carving cartilage can alter the balance of forces that maintain its shape and can lead to bending. This is important in carved rib cartilage grafts to the nasal dorsum, which can warp late after operation, and it is probably the mechanism by which an injured nasal septum becomes progressively more deviated after a broken nose.

Chronic Wounds

Chronic wounds remain one of the most costly unsolved problems in health care today. It is estimated that some 15 percent of the approximately 14 million diabetics in the United States develop skin ulcers, which result in about 60,000 amputations per year. The range of medical costs for diabetic foot ulcers and amputations has been estimated to be \$10,000 to \$27,000 per patient. Thus the overall cost of diabetic foot problems, including loss of productivity, could be as high as \$20 billion per year. In addition, there are 175,000 to 250,000 spinal cord injury patients in the United States, with approximately 7,000 to 10,000 new patients per year. Approximately 60 percent of these patients develop pressure ulcers, and cost estimates range from \$14,000 to \$25,000 per patient for medical, surgical, and nursing care. If the elderly nursing home population with pressure ulcers is added to the spinal cord injury population, then the figure for the care of all pressure ulcers is enormous. The care of venous stasis ulcers is also very costly.

Acute Versus Chronic Wounds

There are significant differences between acute and chronic wounds. An *acute wound* usually occurs in a normal, healthy person and is either closed primarily or allowed to close by "secondary intention." Most injuries to whole organs or tissues can be considered acute wounds.

A *chronic wound* is one that fails to heal because of some underlying pathologic condition. For example, pressure ulcers, diabetic ulcers, and venous stasis ulcers are chronic wounds. These complex wounds will not heal until the underlying cause is corrected. Curiously, many of these chronic wounds heal to a point, and then the healing process arrests. The precise factors that cause this phenomenon are unknown. With proper clinical management, most chronic wound healing problems are resolved and healing occurs, but recurrence is common.

Pathophysiology

The pathophysiologies of chronic wounds, such as diabetic foot ulcers, pressure ulcers, venous stasis ulcers, and ischemic ulcers, are complex and diverse, but they appear to share one common feature. Unlike acute wounds, which typically result from transient, externally induced damage to tissue, chronic wounds arise from physical and biochemical insults of extended duration. This prolongs the inflammatory stage of wound repair, resulting in extensive tissue damage and impaired healing. Impaired wound healing in the guise of chronic dermal ulcers represents a major health problem that affects millions of Americans yearly.

During normal wound repair, *polymorphonuclear leukocytes (PMNs)* quickly respond to chemoattractants, including fragments of fibrinogen, fibrin, the complement degradation fragment C5a, formyl methionyl peptides released from bacteria proteins, and soluble mediators such as platelet-activating factor (PAF) and PDGF. In response to these chemoattractants, PMNs become activated and transmigrate through the endothelium into the wound site. Wound PMNs destroy and phagocytize contaminating bacteria as well as damaged or necrotic tissue. To accomplish this task, PMNs release a large number of enzymes and cationic proteins and express plasma membrane enzymes that generate reactive oxygen metabolites.

In normal repair, PMN infiltration lasts only a few days. However, the perseverance of severely damaged or necrotic tissues, intermittent or continuous ischemia, or microbial infection leads to a persistent and excessive response by activated PMNs. The products released by PMNs continue to degrade the extracellular matrix (generating additional fragments, which are PMN chemoattractants) and prevent or impair the migration of other reparative cells into the wound. The high levels of active proteases in many chronic wounds may partly explain the disappointing results obtained when these wounds have been treated with peptide growth factors.

Venous Stasis Ulcers

Venous stasis ulcers are the eventual result of deep venous obstruction or valvular incompetence. The resulting increase in venous pressure dilates veins and increases capillary permeability, which promotes extravasation of fluid and high-molecular-weight proteins. Ulceration can result after several previous manifestations (swelling, leg pain, dilated superficial veins, and erythema) of chronic venous insufficiency. Ulcers typically occur superior to or near the medial malleolus and tend to be large and irregularly shaped (Fig. 8-30). The extent of ulceration can range from just below the epidermis to as deep as the fascia. The ulcers commonly are rimmed by an area of hyperproliferative keratinocytes.

The appearance of perivascular "fibrin cuffs" is a particular characteristic of this disease. The abundant presence of extracellular matrix components, such as laminin, collagen, and fibronectin, suggests that an attempt to seal the gaps between the endothelial cells may be occurring. Fibrin cuffs were once believed to act as barriers to transportation of oxygen, nutrients, and growth factors. However, cuff formation around the capillaries is discontinuous, and there is little correlation between the extent of cuff formation and the clinical presence of lipodermatosclerosis, ulcer size or transcutaneous oxygen measurements taken at sites adjacent to the ulcer.

An alternative hypothesis suggests that venous hypertension leads to endothelial cell damage, and this injury leads to an aggravated and persistent inflammatory response that involves PMNs. This concept is supported by the findings of fibronectin degradation products and elevated levels of PMN-specific proteolytic enzymes in fluid exudates of stasis ulcers.

Pressure Ulcers

Pressure and shear forces over bony prominences have a key role in the formation of pressure ulcers (Fig. 8-31). Immobilized persons are at risk because pressure causes cell death in the least vascularized tissues. In porcine models, damage initially occurs in the fat and deep muscles and progresses outward toward the skin. Necrosis spreads outwardly, beginning with the subcutaneous fat and eventually affecting the epidermis. Histologic analysis of the course of ulcer formation in human beings reveals a sequence of capillary and venule dilation accompanied by a perivascular infiltrate. As with venous insufficiency, capillary permeability allows intravascular cells to enter the connective tissue with the release of degradative enzymes and reactive oxygen metabolites. In cases in which ulceration extends to the bone, osteomyelitis can be a complicating factor. Marjolin's ulcers rarely occur in these chronic ulcers.

Patients with spinal cord injury are more susceptible to the ravages of pressure than the average healthy person. Although some factors remain unclear, it is well documented that spinal cord injury patients do not have a normal leukocyte response to injury below the level of denervation. These patients often have low ascorbate levels despite supplementation, which can be deleterious to proper collagen formation. In animal studies, collagenase activity is greater in denervated tissues. In patients with spasticity, shear forces may be even more harmful than pressure.

Elderly bedridden patients have additional promoting factors besides pressure. Fecal and urinary incontinence leads to maceration and skin breakdown. Poor nutrition also is a major problem.

In both groups, good wound care is critical. Antibiotics should not be used to treat these wounds unless they are causing systemic toxicity. When these patients develop fever, the cause rarely is the wound except if there is loculated purulence. Pulmonary, renal, or intraabdominal problems are more common. If the wound colony count is 100,000/g of tissue or higher and the patient is not septic, topical antibiotics should be used to reduce the bacterial load before closure or in order to facilitate spontaneous healing. Bacterial counts sometimes can be lowered with good wound care alone, without the use of antibiotics. The best evidence to date is that systemic antibiotics will not penetrate the ulcer and therefore will not decrease the bacterial colony count within the wound granulation tissue. The use of systemic antibiotics is often abused in patients with chronic wounds, often simply altering the bacterial flora and making the host more susceptible to resistant organisms.

The National Pressure Ulcer Advisory Panel has developed a staging system (derived from other systems) for classifying pressure ulcers:

Stage I. Nonblanchable erythema of intact skin, the heralding lesion of skin ulceration.

Stage II. Partial-thickness skin loss involving epidermis and/or dermis. The ulcer is superficial and presents clinically as an abrasion, blister, or shallow crater.

Stage III. Full-thickness skin loss involving damage or necrosis of subcutaneous tissue that may extend down to, but not through, underlying fascia. The ulcer presents clinically as a deep crater with or without undermining of adjacent tissue.

Stage IV. Full-thickness skin loss with extensive destruction, tissue necrosis, or damage to muscle, bone, or supporting structures (e.g., tendon, joint capsule, etc.).

One must use this classification with caution because it can be misleading.

Diabetic Ulcers

Chronic ulcers in diabetics typically present as foot ulcers (Fig. 8-32). Pressure and tissue trauma can be considered the major promoting factors, but the neuropathy from the primary disease is the most important causative element. Lack of sensation results in increased mechanical stress under the metatarsal heads, heel, and callosities. This in turn leads to intermittent or continuous ischemia, resulting in pressure ulceration. In addition, severe neuropathy prevents the timely detection of skin punctures, or improper shoes, or the detection of foreign bodies in shoes.

There are other factors that may be involved in the formation of diabetic ulcers. In addition to neuropathy, diabetics are also prone to angiopathy; the capillaries of diabetics are thicker and more permeable, and the absolute number of capillaries is smaller. Granulocyte chemotaxis and phagocytic function is impaired in diabetics. Diabetic wounds are characterized by decreased levels of extracellular matrix components. The collagen of diabetics is subject to nonenzymatic glycosylation, which alters the filtration function of basement membranes, thus leading to many of the pathologies characteristic of diabetes. These include reduced vessel permeability, altered lens-capsule function, and renal problems secondary to abnormal glomerulus filtration. The breaking strength of wounds in diabetics is also reduced.

Mechanisms Involved in the Healing of Chronic Ulcers

Sometimes the basic biologic mechanism of healing chronic ulcers is contraction, which reduces the area of the wound. Usually minimal epithelialization is required to heal chronic ulcers, and the result is often cosmetically and functionally acceptable. Pressure ulcers and diabetic ulcers heal mainly by contraction, but there are exceptions. For example, venous stasis ulcers do not heal by contraction, but mainly by epithelialization. If the margins of these ulcers are tattooed and then the wound heals, the tattoo marks remain in their peripheral location. It would appear that epithelial cells cover the wound and that these cells lead to the induction of a neodermis. This is supported by the evidence that cultured epithelial cells applied to a granulating wound bed will induce new dermis formation.

Chronic Wound Care

Most chronic wounds, regardless of the cause, will heal by secondary intention only if the underlying biochemical and mechanical causative factors are corrected. For example, compression stockings or dressings must be used to relieve venous hypertension in order for venous stasis ulcers to heal. Conversely, pressure must be eliminated over the pressure sore. Special pressure-relieving beds and other devices have been designed for this purpose. Diabetes must be controlled to ensure healing of diabetic ulcers. Bacterial counts above 100,000/g of tissue must be reduced and nutritional deficiencies corrected.

Successful treatment of a chronic ulcer depends on an accurate assessment of the factors contributing to the ulcer pathology (Table 8-8). For example, immunocompromised patients may require systemic or topical antibiotics to prevent or control infection and sepsis. Nutrition deficiencies must be corrected. Adequate amounts of protein, fats, carbohydrates, and vitamins are necessary to provide the energy and biosynthetic requirements for repair. Albumin level is an important indicator of malnutrition; it is sacrificed to provide essential amino acids in the event of inadequate protein intake. Because the half-life of albumin is 20 days, a low albumin level is indicative of chronic malnourishment; sporadic deficiencies in intake do not result in perceptible changes in serum albumin level.

Tissue perfusion and subsequent cellular oxygenation are important factors for chronic wound repair. Oxygen is required for the production of bactericidal reactive oxygen metabolites and for the synthesis of collagen. Destruction of bacteria by neutrophil-mediated oxidative killing is significantly reduced in hypoxic tissues. Poor perfusion can result in an increase in infection and an increase in the impairment of healing. There is no evidence that hyperbaric oxygen can improve the healing of most chronic wounds with the exception of osteoradionecrosis. This mode of therapy is very expensive. Although used for decades, its efficacy has never been validated. It is a prime example of a therapy lucrative for the manufacturers and therapists that has been promulgated by marketing without substantiation. It is to be hoped that third party payors will refuse such bogus therapy unless users can prove efficacy.

Other factors that influence repair include age, metabolic disorders (uremia, jaundice, diabetes), pharmacologic agents that interfere with repair (glucocorticoid steroids), distant malignancies, and infection. The location and severity of the ulcer also have a role in determining treatment. Small pressure and diabetic ulcers are best managed by local wound care and removal of the pressure source. Larger defects may require flaps for closure. Management of venous stasis ulcers involves removal of venous pressure by the use of elastic stocking support and elevation of the limb. Skin grafts and sometimes flaps are required for severe larger ulcers.

Wound cleansing has a limited role in the treatment of chronic wounds. The objective of wound cleansing is not wound sterilization but rather reduction of the microbial load, to remove necrotic tissue and foreign bodies, and to diminish levels of autolytic enzymes in the wound. Debridement to remove damaged and necrotic tissue often helps to accelerate healing. Sharp surgical debridement is the least expensive, quickest, and most effective method. Recent studies have demonstrated that total excision of diabetic ulcers accelerates the healing process. Use of disinfectants or high-power sprays should be avoided. Disinfectants such as povidone-iodine, acetic acid, hydrogen peroxide, or Dakin's solution (sodium hypochlorite) are as likely to injure normal tissue as they are to destroy microorganisms. In some cases irrigation of the wound with normal sterile saline is helpful.

Power sprays pose the danger of forcing microorganisms deeper into healing tissues, and they should be eliminated from chronic wound care regimens. Irrigation pressures should not exceed 15 pounds per square inch (psi). Whirlpool treatment has been recommended for debridement of large areas of foreign debris or nonviable tissue. Care must be taken to prevent infection between patients by strict hygienic control of whirlpools. Given the debate about the efficacy of whirlpool treatment, the concerns about spreading wound contamination, and the expense of using whirlpools, this therapy may become obsolete in the future.

WOUND DRESSINGS

During the past decade, dressings for acute and chronic wounds have changed dramatically. As new material technologies have evolved, new products have inundated the market. Dressing products should be selected on the basis of the type and characteristics of the wound.

The acute wound that is closed primarily requires coverage with a dry sterile dressing for a few days. This helps to protect the wound from bacterial invasion and absorbs any wound fluid. The dressing also may provide some psychological benefit to the patient in keeping the surgical wound out of sight. When epithelialization is complete, the dressing can be removed.

Partial-thickness wounds—e.g., donor sites, abrasions, first- and second- degree burns—require semioclusive dressings. These products provide a moist environment that enhances the reepithelialization of partial-thickness wounds. If a semioclusive dressing is applied early to a partial-thickness wound, dry scab formation can be avoided. Scab formation can damage the underlying dermis, and removal of the scab may be painful, cause bleeding, damage new epithelium, and increase the deformity in the final scar. Semioclusive dressings may not have enough absorption capacity to handle the amount of wound exudate produced in the early phase of healing. Wound fluid can become trapped and cause the area around the wound to macerate or leak, necessitating frequent dressing changes. Combining an absorbent dressing material, such as an alginate, with the polyurethane film or hydrocolloid alleviates this problem.

The classification, composition, indications, and functions of dressings are summarized in Table 8-9. Occlusive/semioclusive dressings may exacerbate infection when covering areas in which the bacterial count is higher than 100,000/g of tissue. This risk may be reduced by first controlling the wound bacterial count with topical antimicrobial or systemic antibiotics. In addition, several of the hydrogels can be used effectively with antibiotic creams. Hydrogels are thought to accelerate the rate of penetration of the antibiotic into the granulation tissue.

Dressings are used to debride. If the wound is dry, it is important that a slower and more gentle process be used, especially when tendons are exposed. This can be accomplished with hydrogels that have little permeability and therefore tend to keep the wound moist. When eschar is present, a hydrocolloid may be more appropriate than a hydrogel dressing. This type of debridement occurs through autolysis. The hydrocolloid traps the wound exudate and creates a moist environment that softens and lifts or causes proteolytic digestion of the eschar. However, in most cases, sharp surgical debridement of eschar is the most efficacious treatment.

Wounds with a large volume of exudate require dressing with a greater capacity to absorb. When dressings do not adequately absorb, the healthy tissue around the wound may become macerated. The use of a skin protectant around the wound may alleviate this problem. There has been concern that hydrocolloids leave pectin base in wounds, which forms granulomas. There are, however, several non-pectin-based hydrocolloids that address this problem and also absorb wound secretions. There are no data to support the contention that one dressing promotes healing better than another. However, selection of dressings may be made on the basis of overall costs, efficacy, and ease of use. It is likely that over the next decade dressings will be designed to have “selective absorption” properties to remove only factors that impede the healing of chronic wounds or prevent scar hypertrophy in primarily closed wounds. Most of the “new” concepts in dressings are marketing gimmicks rather than improvements based on solid and objective scientific data.

MECHANICAL WOUND CLOSURE

The materials used for wound closure are much less important than the techniques of closure. Sutures may be generally classified as *absorbable* or *nonabsorbable*. The absorbable sutures may be synthetic, such as polyglycolic acid sutures, or biologic, such as “catgut” sutures, which may be plain or chromium-treated. Traditional teaching is that absorbable sutures are buried and, as they absorb, will not be a nidus for late infection. Nonabsorbable sutures are used on the skin because they are less reactive and allegedly provide a better-appearing scar. These dogma make little sense in the schema of healing. Although it is true that gut sutures are more reactive than polyglycolic acid, which is more reactive than nylon, the argument that tissue reactivity to particular sutures is of significance in the healing process has never been validated. Nonabsorbable sutures may be used in subcutaneous tissue as well as in fascia or for organ repair. By contrast, absorbable sutures often are used on the skin in infants and children; 6-0 plain gut sutures placed in the skin do not require removal in a child, thus avoiding the additional trauma of suture removal. Similarly, absorbable sutures are used commonly for closure of hand wounds—even in adults.

The notion that nonabsorbable sutures in fascia are better than long- lasting absorbable sutures is partially myth, because tension will gradually cause remodeling of the connective tissue around the suture. If fascia wound strength has not developed adequately in the repair process, the suture material will not have an influence on outcome.

The most important fact about sutures is that any woven suture is more likely to facilitate infection than a smooth suture because bacteria can become entrapped in the interstices of a woven suture and not be destroyed by the normal host responses, thus leading to bacterial propagation and infection. Therefore, woven suture material should not be used in the closure of contaminated wounds.

A few suture materials deserve special comment. The nonabsorbable polypropylene suture is extremely smooth and is therefore the best material to use when creating a subcuticular pullout suture. The absorbable polyglycolic dermal suture (PDS) is best for areas in which long-term tensile strength is required. Although stainless steel sutures are still used, they should be banned in the operating room because they often cut through gloves and create an environment where the surgical team may be more susceptible to pathogens such as hepatitis or HIV.

Staples

Although surgical staples have revolutionized the repair of organ parenchyma and the anastomosis of various hollow organs, they must be used with discretion on the skin. Well-designed staples, applied properly, are excellent for everting the skin margins appropriately. Staples must be removed within a few days if permanent skin marks are to be avoided.

Adhesive Tape Strips

In certain types of wounds, adhesive tape strips can be used. For example, once dermal sutures are in place, tape strips may be substituted for surface skin sutures as long as the wound edges are well approximated. Tape strips also can be helpful in supporting the wound margins after early removal of skin sutures. The tape strips are simple and clean, and they may be used to close many small lacerations painlessly without the need for local anesthetics.

FETAL WOUND HEALING

During the past decade, there has been a logarithmic increase in interest in fetal surgery and fetal wound healing. Basic research focused on animal fetal surgery has led to the repair of life-threatening conditions in human fetuses without maternal mortality, though the long-term fetal survival is still disappointing. Fetal surgery likely will become routine in future decades and can be expected to have better survival. It is imperative that the mechanisms of fetal tissue repair be understood if human fetal surgery is to be performed optimally. An even more important reason to study fetal healing is to discover the mechanisms that make it so dramatically different from adult healing. Understanding fetal healing should prompt major advances in our treatment of the adult healing problems of scarring, contraction, and wound disruption.

Fetal wound repair is characterized by a significantly reduced inflammatory response. Studies have shown that fetal platelets have different aggregative characteristics and reduced cytokine release compared to adult platelets. However, an inflammatory response can be induced by the addition of living or dead bacteria, removal of hyaluronic acid from the connective tissue matrix, or a variety of cytokines, including TGF-b and PDGF. Rather than collagen, the major component of the wound matrix is the glycosaminoglycan hyaluronic acid.

Primarily closed skin wounds in a third-trimester rabbit cannot be detected on routine histologic examination or by electron microscopy. There are no architectural abnormalities in the matrix that is formed in the wound site. Collagen synthesis measured daily after wounding is significantly greater in fetal wounds than in normal fetal skin. In addition, there is an astonishing gain in the tensile strength of the closed fetal skin wound by 5 days, in marked contrast to wounds in the adult, which gain little strength in this time period. These data indicate that there must be a very rapid turnover, remodeling, and reorganization of collagen during fetal repair.

Third-trimester open fetal rabbit wounds do not contract, in contrast to second-trimester fetal sheep wounds, which do contract. There is no evidence of myofibroblasts in the noncontracting fetal rabbit wounds. However, if TGF-b is placed in the open fetal wound, the wound does not expand as does the untreated open wound. If the open wound is sealed from the amniotic fluid environment by a silicone patch, it closes by epithelial cell migration. There is reasonable evidence that amniotic fluid may inhibit fetal wound contraction, but the mechanisms and component(s) remain unclear. The content of the matrix of the open fetal wound is mainly hyaluronic acid, and this may be an important factor in the contractile process in both adult and fetus wounds. There is promise that the many clinical pathologic wound healing responses seen the adult will some day be treated with new strategies based on the optimal responses observed in the fetus.

Acknowledgments

The authors wish to thank Cary Klett, M.D., for her helpful contribution to the section on genetic defects in connective tissue metabolism, to Benedict C. Nyomeh, M.D., for compiling information in Table 8-1, “Cytokines that Affect Wound Healing,” and to Stephen J. Leibovic, M.D., for his helpful comments on the section on tendon repair. Portions of the information presented in this chapter were the result of studies funded by National Institutes of Health grants GM-20298 and GM-47566.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 9 - Oncology

John M. Daly
Monica Bertagnolli
Jerome J. DeCosse
Donald L. Morton

INTRODUCTION

Approximately 90 percent of patients with malignancy undergo surgical therapy for diagnosis, primary treatment, or management of complications. Resection is the initial curative treatment for about 75 percent of patients, because cancer is assumed to be a localized disease for an interval, allowing cure after adequate surgical removal. Long-term survival statistics of patients with cancer treated surgically support this assumption. The lack of further improvement of survival rates with larger or more radical operations suggests that other forms of treatment should be implemented to maximize cure rates. Inclusion of radiation therapy or chemotherapy, or both, may improve the overall survival rate while permitting less extensive operative resection and enhancing cosmesis and function.

The surgeon is responsible for the initial diagnosis and management of many types of cancer. Knowledge of tumor staging and the natural history of neoplastic disease is essential to a multimodal approach for treatment of the patient in collaboration with the medical oncologist and radiotherapist. The surgeon's guiding principles should be the accurate diagnosis and staging with adequate operative removal of localized disease (Table 9-1) and palliation of symptoms when possible.

If the cancer has spread beyond local cure, the goal is to control the patient's symptoms and to maintain maximum activity as long as possible. The success of palliative therapy is measured in terms of useful life. The most common criterion of incurability is distant metastasis. However, patients with solitary pulmonary metastases may be curable by resection, and those with widespread metastases from choriocarcinoma may be curable with chemotherapy. Local extension also may be a criterion of incurability. If extensive studies fail to demonstrate metastatic disease or incurable local extension, the patient's treatment should be directed toward a cure.

Management of the cancer patient is a multidisciplinary effort requiring collaboration among surgical oncologists, radiation oncologists, medical oncologists, reconstructive surgeons, and other oncologist specialists. Evidence suggests that combinations of surgery, radiation therapy, chemotherapy, hormone therapy, and immunotherapy significantly improve cure rates above those achieved with any single therapeutic modality. Multimodal therapy, for example, is standard for most breast and colon cancer patients (Fig. 9-1). As the primary coordinator in cancer management, the surgeon must fully understand the indications, risks, and benefits of surgery, adjuvant chemotherapy, hormonal therapy, radiation therapy, and the importance of reconstructive surgery.

The optimal combination and sequence of treatments is determined by the patient's physical, emotional, psychological, and rehabilitative needs. The patient's general condition and the presence of any coexisting disease must be considered. Surgery may be contraindicated in a patient who has recently experienced a myocardial infarction. A patient with preexisting diabetes will be more susceptible to the toxic effects of hormonal therapy with corticosteroids. Renal disease can increase the toxicity of some of the chemotherapeutic agents, such as methotrexate. Any evidence of infection or bleeding may require vigorous treatment before definitive therapy is initiated. The patient's psychologic makeup and family, friend, and job support structure must be considered. A patient who is unable to accept the realities of a given treatment should be offered an alternative approach if possible. This is particularly true of surgical procedures that significantly alter appearance or involve change of organ function requiring the patient's daily care, such as colostomy. Experimental forms of therapy, such as intraarterial infusion of drugs, should also be avoided in some patients; a patient who is unwilling to tolerate the inconvenience of an intraarterial catheter might remove it without medical approval.

Determining a treatment plan requires the integration of information from four areas: (1) natural history of the disease by histologic type, (2) clinical staging, (3) goals of specific treatments, and (4) indications and risks for each treatment (or combination of treatments) based on results of experience and clinical trials.

EPIDEMIOLOGY

Cancer is overtaking heart disease as the most frequent cause of death in the United States, accounting for 24 percent (approximately 520,000) of deaths each year. The five leading causes of cancer death among males in the United States are: lung, 32 percent; prostate, 14 percent; colon and rectum, 9 percent; leukemia and lymphoma, 9 percent; and pancreas, 5 percent (Fig. 9-2). Among females the leading causes of cancer deaths are: lung, 25 percent; breast, 17 percent; colon and rectum, 10 percent; leukemia and lymphoma, 8 percent; and ovary, 6 percent (Fig. 9-3). Prostate cancer is the most frequent life-threatening cancer in men, and breast cancer the most frequent in women. The impact of cancer is no less on a worldwide basis; hepatocellular cancer is the cause of approximately 1 million deaths a year.

Although neoplasia is a disease of the genome with many common molecular pathways, human cancer may be envisioned as more than 100 distinct entities, each defined by the cell or tissue of origin and the appearance under the microscope. For some, the initial inductive event is inherited, but, as demonstrated by Knudson, general belief is that more than one genetic or epigenetic event is necessary for promotion of human carcinogenesis. Each site has a host-tumor interface and a subclinical growth phase, which might be measured in decades before emerging at a clinical threshold. At a threshold level, almost all common sites of life- threatening cancer have a definable benign precursor (Table 9-2) that shares the epidemiology of the life-threatening tumor and that may be reversible.

Screening policies for the prevention and early detection of cancer emerge from understanding the natural history of cancer at specific sites. The hallmark of an effective screening policy is demonstration of a reduced mortality rate. This goal has been achieved by the Pap smear for cervical cancer, mammography for breast cancer, and the fecal occult blood test and sigmoidoscopy for colorectal cancer.

To a large extent our present knowledge about the etiology and control of cancer has emerged from epidemiology. Observational ecological studies, international correlations among populations and migrants, and cross- sectional surveys, reinforced by case-control and cohort studies, have been the basis for conclusions about etiology. Nevertheless, when all known causes are accounted for, the odds of a person's acquiring cancer still depend substantially on chance.

The primary determinants of risk have been incidence rates and mortality rates, the former defined by the number of new events or cases that develop in a population of individuals at risk during a specific interval. These rates should be distinguished from prevalence (the number of affected persons within a population). Rates can be crude, category-specific (e.g., age, gender, or race), or adjusted (e.g., accounting for mortality from other causes). Herein resides the value of population-based registries and of high-risk registries in which persons at risk can be concentrated for study and treatment.

Long-term changes in cancer incidence and mortality in the United States (see Figs. 9-2 and 9-3) defy a satisfactory explanation, but an aging population and some interventions are partial explanations. In the United States, a notable decrease in the incidence and mortality of gastric cancer has occurred over the past half century, presumably as a result of general public health measures and, possibly, refrigeration, but without a specific understanding or intervention. For many decades breast cancer incidence has been increasing while mortality has remained stable, but now there is a decline in mortality, a result of widespread acceptance of mammography and detection of breast cancer in an earlier, curable stage. In recent years the mortality rate of lung cancer has stabilized in men, an effect attributable to reduction in cigarette smoking.

From global studies by Doll and Peto, vastly different incidence rates of site-specific cancer have been found (Table 9-3). Cancer of the stomach remains a leading cause of death in Asia and eastern Europe. When natives of Japan, where gastric cancer is frequent and colorectal cancer uncommon, emigrated to Hawaii, over one generation the frequency of these cancers was reversed, a change attributed to adoption of the Western diet. Given that diet and nutrition may be characterized as an environmental factor, these differences incriminate environment-induced molecular events in human carcinogenesis. The more frequent occurrence of cancer in older persons may reflect accumulation of environmentally based genetic events as well as molecular events associated with senescence. Acquired genetic or epigenetic events originating in physical, chemical, and viral etiologies are described in more detail below; initial insights were derived from observational studies.

The increased incidence of colorectal cancer in the United States can be attributed to the high fat and meat content of the Western diet. Metabolic epidemiology encompasses other effects of the internal environment on carcinogenesis as well. For example, breast cancer in women can be related to the lifetime duration of unprotected exposure to biologically active estrogens. Early menarche, late menopause, and delayed or absent child- bearing increase risk, whereas oophorectomy and estrogen antagonists reduce risk.

Recognition of internal and external environmental interactions in human carcinogenesis provides the means for risk reduction and the development of prevention strategies. Elimination of exposure to asbestos or radiation, avoidance of occupational carcinogens, and reduction of cigarette smoking remove certain carcinogens from the environment and reduce risk. Sunscreens block the carcinogenic wavelength of ultraviolet light. Vaccination for hepatitis B virus reduces the risk of hepatocellular cancer.

Other measures to control environmental carcinogens may be directed to the internal environment by nutritional modification and chemoprevention. A low-fat, high-fiber diet reduces risk of colorectal cancer and, possibly, breast and prostate cancer. Oral vitamin A analogues reduce leukoplakia and second cancers in the aerodigestive epithelium. Oral nonsteroidal anti- inflammatory drugs decrease rectal adenomas in patients with familial adenomatous polyposis. Antiestrogens reduce second cancers in the breast.

BIOLOGY OF MALIGNANT TRANSFORMATION

History

Cancer has been recognized as a disease for thousands of years. The first known mention of cancer is in the Ebers Papyrus, a medical treatise written in Egypt about 1600 b.c. The manuscript recommends surgical excision or cauterization for tumors. The scientific advances of the Greeks and Romans led to the understanding that cancer results from physiological processes. Galen (a.d. 129– ca. 199) defined cancer as “tumors against nature” and ascribed the cause of cancer to “humoral” imbalances. Paracelsus (a.d. 1493–1551) argued that cancers were caused by external agents—which is supported by current literature linking cancer development to environmental carcinogens. It was not until the 1850s that Rudolf Virchow, the father of cellular pathology, recognized cancer as an alteration of the native tissues. In the 1870s Julius Cohnheim argued that cancer was a failure of embryonic cells present in all tissues to mature properly, the first mention of carcinogenesis as “dedifferentiation.” Theodor Boveri, a professor of zoology at the University of Würzburg in Germany, reported in 1902 that his studies of mitosis in sea urchins led him to believe that cancer was due to abnormal chromosomes. Boveri's theory competed with the infectious disease theory of carcinogenesis, but the identification of the chicken leukemia virus by Ellerman and Bang in 1908 and Rous's identification of the chicken sarcoma virus in 1911 brought the infectious disease theory back to the forefront. Knudsen, in his 1971 description of familial retinoblastoma, suggested that a combination of inherited and acquired events govern cancer development. More recently Vogelstein and de la Chapelle, on the basis of their identification of an association between cancer and inherited defects in DNA repair, postulated that epithelial cancers result from a failure of the cell to maintain DNA fidelity during replication.

The common denominator in all theories of carcinogenesis is alteration of the cell's genome, either by direct damage from radiation or chemicals, by integration of viral genomic

sequences, or by an inherited defect in DNA repair capacity. The study of cancer etiology is focused on the molecular events required for malignant transformation, particularly the interplay between genetic and environmental influences in carcinogenesis.

Cellular Homeostasis

It is instructive to study normal regulatory processes to understand the derangements that occur with carcinogenesis. To achieve homeostasis in tissues, renewable cell populations must perform four related functions; they must (1) proliferate with proper timing and fidelity of DNA content, (2) differentiate in a pattern consistent with normal function of the tissue, (3) involute in a manner such that the proliferation and involution rates are balanced, and (4) repair any damages to their DNA resulting from exposure to mutagens such as radiation, toxins, and transforming viruses (Fig. 9-4). A defect in any of these functions can result in tumor formation.

Carcinogenesis

The term cancer refers to a group of diseases characterized by the autonomous growth of abnormal, “neoplastic” cells. Cancer results from a deregulation of critical aspects of cellular function, such as proliferation, differentiation, and apoptosis. Without the proper constraints on these processes, neoplastic cells reproduce in great numbers, invade adjacent structures, and develop metastatic colonies. The natural history of most cancers suggests that the development of these abnormal characteristics occurs in a progressive fashion. In describing this process, tumor *initiation* is defined as the exposure of cells to agents that induce an inheritable genetic change, i.e., agents that are genotoxic or induce critical mutations by binding of electrophilic carcinogenic metabolites to DNA. Tumor *promotion* is the exposure of initiated cells to agents that induce their proliferation. This proliferation may allow other spontaneous mutations to occur that culminate in expression of malignant phenotype (malignant transformation). Tumor *progression* describes successive development of increased local growth, invasion, and metastasis by transformed cells.

Cancer Phenotype

Progression of a tissue to malignancy disturbs host homeostatic mechanisms, as characterized by (1) unresponsiveness to normal growth regulators, (2) invasive phenotype, and (3) evasion of immune-mediated tumor destruction. Tumors are thought to be clonal in origin (i.e., all of the cells within a tumor arise from a single progenitor cell whose growth regulation has become deranged). Support for this clonal theory comes from the observation that tumors arising in women who are heterozygous for a gene encoding the enzyme glucose-6-phosphate dehydrogenase (G6PD) express only one of the two isoenzymes. In addition, all of the neoplastic cells in patients with chronic myelogenous leukemia whose tumor bears the Philadelphia chromosome (Ph1) contain Ph1. Some data from animal models, such as chemically induced murine fibrosarcoma, suggest that solid tumors may be multicellular in origin. Clonal dominance, however, has been demonstrated for both murine and human tumors. Despite their possible clonal origin, cancers, particularly the solid tumors, are heterogeneous in character. A cancer mass includes tumor cells and their supporting blood vessels and stroma. Within the tumor, abnormal regulation of cellular function leads to nuclear and cellular polymorphism, loss of cellular polarity, and variation in DNA content from cell to cell (aneuploidy). These processes are accelerated as cells lose control over DNA repair. As regions of tumor outstrip their blood supply, areas of inflammation and necrosis further contribute to tumor heterogeneity. As a result of loss of the fidelity of DNA replication, changes in the malignant cell population occur throughout the course of tumor progression. This is best demonstrated by a change in differentiation state or tumor antigen expression between primary tumors and their metastatic foci.

In an idealized model, the growth of a tumor proceeds exponentially in the early phases when the supporting tissues and nutrients are optimal, with a shortening of the doubling time as the tumor enlarges. This relationship between time and tumor size is referred to as *Gompertzian growth*, named for the eighteenth-century mathematician Benjamin Gompertz, who developed the model that describes it. The natural history of many human tumors is described by this model (Figs. 9-5 and 9-6).

Progression of a tissue to malignancy involves several stages. The earliest visible evidence of neoplastic transformation is dysplasia, a condition in which epithelial tissues exhibit altered size, shape, and organization. Dysplasia is a common reaction of tissue to chronic inflammation or exposure to environmental toxins or irritants. The degree of deviation from normal cellular and tissue architecture defines dysplasia as mild, moderate, or severe. In epithelial tumors, dysplastic cells are confined to the region above the basement membrane. Because dysplastic cells retain a measure of control over cellular proliferation, dysplasia is generally reversible once the inciting factor is removed. In most tissues, however, severe dysplasia is associated with progression to carcinoma if left without intervention.

The hallmark of a solid tumor carcinoma is the ability to invade the basement membrane and spread without regard to normal tissue boundaries. *Local* disease is the term used to refer to invasive tumor that is confined to the tissue of origin. Once the basement membrane has been breached, the next barrier to tumor dissemination is the network of draining lymph nodes. Tumor spread to the lymph nodes draining the tissue of origin is termed *regional* disease. The final stage of tumor progression is metastasis, whereby independent colonies of tumor are established in distant sites favorable to tumor growth. This type of tumor is commonly referred to as *distant* disease.

Carcinogens

In 1775 Sir Percival Potts, a surgeon at St. Bartholomew's Hospital in London, described the “soot wart,” a squamous cell carcinoma of the scrotum in chimney sweeps. This was the first recognition that cancer is a disease related to long-term exposure to associated environmental factors. Molecular studies link some known carcinogens to specific cancer-associated mutations, e.g., benzo[a]pyrene, a component of cigarette smoke, is a potent mutagen that preferentially induces damage of the p53 tumor suppressor gene at nucleotides that are common mutation sites in human lung cancer. Environmental factors implicated in carcinogenesis include radiation, chemical agents, and viruses. The common denominator among these factors is the ability to induce heritable change in cellular DNA.

Physical Carcinogens

Tumor induction by physical agents occurs by essentially two mechanisms: (1) induction of cell proliferation over an extended period of time, which increases the opportunity for events leading to transformation; and (2) exposure to physical agents that induce damage or changes in cellular DNA. Foreign bodies, particularly irritants, can induce tumors in animal models; however, it is difficult in humans to attribute the induction of a tumor solely to the presence of a foreign body. Foreign bodies may act as cocarcinogens or promoters, or may produce a chronic irritation that exposes tissue to carcinogenesis by other environmental agents. The higher risk of skin cancer in patients with burns or esophageal cancer after chemically induced irritations, such as that following lye ingestion, provides an example of the close link between tumorigenesis and the induction of cell proliferation. Cellular proliferation associated with chronic inflammatory conditions may contribute to the carcinogenicity of other agents or may increase the probability of genetic changes resulting in neoplastic transformation. An example is the increased colorectal cancer risk in inflammatory bowel disease.

The best known agent of physical carcinogenesis is radiation. There are two types of radiation: ionizing and non-ionizing. Ionizing radiation includes x-rays and gamma rays, and alpha and beta particles, while the most common form of non-ionizing radiation is ultraviolet (UV) radiation. Both types of radiation are associated with human cancers, but the two groups induce different types of tumor.

Ionizing Radiation

When administered in high doses, all forms of ionizing radiation induce tumors. The carcinogenic potential of x-rays was reported by Friebe in 1902, soon after they were discovered, when many radiologists had developed skin cancers. From 1913 to 1924 a radium-based paint was used to illuminate watch dials, and more than a thousand workers in a New Jersey plant were exposed to this alpha-particle emitter. By 1931 unusually high rates of aplastic anemia and osteosarcoma among radium workers were noted, alerting scientists to the carcinogenic potential of alpha particle exposure. Survivors of the nuclear bombs dropped on Japan in 1945 have an increased incidence of chronic myeloid leukemia and carcinomas of the lung, thyroid, and breast, with a 10- to 20-year latency before development of disease.

The results of radiation exposure are dose and time related. The carcinogenic effect of ionizing radiation is through induction of breaks in DNA strands. Cells most susceptible to this damage are those with a high mitotic activity, such as hematopoietic and epithelial tissues. A population with increased sensitivity to ionizing radiation and other DNA-damaging agents has been identified among individuals heterozygous for a gene mutation responsible for the recessive disorder ataxia telangiectasia (AT). Approximately 2 percent of the population are heterozygous carriers of a defective AT gene allele and demonstrate a four-fold increase in cancer incidence, including tumors of the stomach, breast, skin, ovary, brain, and hematopoietic cells. It is postulated that these individuals lack the ability to repair DNA, although the specific defect associated with AT gene mutation is unknown.

An understanding of the damage caused by ionizing radiation, the tissue penetration of different agents, and the advent of modern dosimetry have reduced the cancer risk associated with therapeutic radiation. Radiation therapy for conditions such as Hodgkin's lymphoma and breast cancer still carries a risk of treatment-related, second malignancies such as leukemia or soft tissue sarcoma.

Non-ionizing Radiation

The association between skin cancer and ultraviolet light was first observed in murine experiments by Hyde in 1906. UV light may be responsible for the induction or promotion of a number of human skin tumors, including basal cell carcinoma, squamous carcinoma, and malignant melanoma. UV radiation is absorbed by DNA and induces DNA damage, particularly the formation of thymidine dimers. DNA repair enzymes that are functioning normally are able to repair these defects. Individuals with xeroderma pigmentosa have a defective endonuclease and are unable to incise the strands damaged by UV exposure. Because UV-induced errors go unrepaired in these individuals, they develop multiple skin cancers. Melanin has a protective effect in absorbing UV radiation, and therefore fair-skinned individuals are at greatest cancer risk.

Cancer-associated DNA damage may also result from physical factors such as the foreign-body reactions induced by asbestos or through reactive oxygen species liberated at sites of chronic inflammation. Exposure to radiation may induce immunosuppression, contributing to carcinogenesis by a lack of immune surveillance. Clinical and experimental models suggest that physical irritants, such as asbestos, may act as cocarcinogens; thus, for example, a combination of asbestos exposure and cigarette use may be synergistic in producing mesotheliomas, lung cancer, and associated epithelial tumors. Observations such as these suggest that in the multistage process of carcinogenesis, physical agents may fulfill one or more of the critical steps necessary to induce a tumor.

Chemical Carcinogens

The first report of a chemical agent inducing cancer was Hill's description in 1761 of nasal carcinoma associated with the use of snuff. The correlation between cigarette smoking and lung cancer was demonstrated by Doll and Hill in 1950, and tobacco products remain the source of most chemically induced human cancers. Epidemiologic studies reveal that a substantial number of compounds are associated with chemical carcinogenesis. Like tobacco smoke, most presumed carcinogenic substances are a complex mixture of chemicals, rather than a single, pure agent. In addition, a number of known chemical carcinogens require metabolism from a procarcinogen into carcinogenic reactive electrophilic intermediates that alter cellular DNA. DNA damage by carcinogen metabolism depends on the balance between the rate of oxidation to carcinogenic intermediates and the rate of detoxification of these oxidative products via conjugation with glutathione and glucuronic acid.

There are several main classes of chemical carcinogens, including organic and inorganic substances. Polycyclic hydrocarbons are organic carcinogens that undergo metabolism in the host to an active form. The most extensively studied polycyclic hydrocarbon is benzo[a]pyrene, a component of cigarette smoke, fossil fuel combustion products, and smoked foods. The metabolic products of benzo[a]pyrene and related compounds are electrophilic epoxides that bind to intracellular proteins, such as DNA. Benzo[a]pyrene is converted to

its toxic metabolites through the action of the hepatic enzymes cyclooxygenase and cytochrome P-450 (Fig. 9-7). Aromatic amines, such as b-naphthylamine, benzidine, aminofluorenes, and dimethylaminoazobenzene (DAB), also undergo metabolic conversion to active forms and are associated with tumors of the urinary tract and liver. Inorganic carcinogens include heavy metal products of fossil fuel combustion such as nickel, cadmium, chromium, arsenic, and lead. Metal carcinogens are mutagenic, inducing DNA adducts and DNA cross-linking. Chronic occupational exposure to these carcinogens is associated with a number of tumors, e.g., nasopharyngeal carcinomas subsequent to chronic exposure to nickel-smelting products.

Viral Carcinogens

Viruses are packages of genetic information, in the form of either DNA or RNA, protected by a structural protein coat. Viruses may insert their genetic material into host cells and induce changes morphologically consistent with neoplastic transformation (Fig. 9-8). Much of our present understanding of carcinogenesis is derived from the study of the effects of viral modification of cellular function. RNA retroviruses isolated from chicken sarcomas demonstrated that sarcomas could be transmitted between chickens by a cell-free infiltrate. It is now recognized that both RNA and DNA viruses are capable of inducing tumors in humans, although only a small proportion of individuals infected with a cancer-associated virus develop tumors. Virus-associated cancers arise after an incubation period of years to decades, suggesting that other genetic or environmental factors contribute to virally induced carcinogenesis.

Observation of the effect of oncogenic viruses on cellular homeostasis suggests that the different oncogenic viruses use similar strategies to deregulate cell growth. Many of the oncogenic viruses encode proteins that interact with the same intracellular targets, particularly Rb and p53 proteins. The contribution of oncogenic viruses to carcinogenesis may lie in the inactivation of these proteins that are essential for regulation of the cell cycle. The most common tumor viruses are listed in Table 9-4.

DNA Tumor Viruses

Most of the human virus-associated tumors are the result of infection with DNA tumor viruses. Because of the worldwide prevalence of hepatitis B- associated hepatocellular carcinoma, DNA tumor viruses may be the most common inciting factor for malignancy.

Epstein-Barr Virus (EBV)

The prevalence of EBV infection worldwide is extremely high, with seropositivity for EBV antibodies approximately 90 percent. EBV infects epithelial cells and some B cells and produces a spectrum of disease from active, highly symptomatic forms such as acute infectious mononucleosis, to a chronic, indolent manifestation. Cancers develop in a small subset of patients infected with EBV, and the epidemiology of these cancers suggests that cofactors in the environment or genetic makeup of affected individuals are important for EBV-induced carcinogenesis. For example, Burkitt's lymphoma is almost entirely restricted to equatorial Africa, where malaria may be a cofactor. In the United States, extranodal EBV-associated B cell lymphomas occur in the setting of immunosuppression after transplantation or in individuals with acquired immunodeficiency syndrome (AIDS).

Hepatitis B Virus (HBV)

Infection with hepatitis B virus is common worldwide. One of the many manifestations of hepatitis B is a chronic hepatic infection which ranges in severity from a mild, essentially asymptomatic form, to severe, chronic, active disease resulting in cirrhosis. Cohort and case control studies in Southeast Asia and Africa suggest that HBV infection is associated with the development of hepatocellular carcinoma. The HBV transforming agent may be the HBV X gene, a segment of the HBV genome that is consistently found in HBV-associated hepatocellular cancers.

Human Papillomavirus (HPV)

The indication that squamous cell carcinoma of the cervix was associated with an infectious agent first came from epidemiologic data showing an increased incidence of this cancer in women with multiple sexual partners. Over 90 percent of cervical carcinomas contain HPV DNA, predominantly subtypes 16 and 18, which is integrated into the cellular DNA. HPV virus infects epithelial cells and induces their proliferation. In the setting of epidermodysplasia verruciformis, an autosomal recessive condition associated with impaired cellular immunity, HPV infection results in squamous cell carcinoma. The significant incidence of squamous cell carcinoma of the anus observed in homosexual males is also ascribed to the combination of HIV-induced immunosuppression and HPV infection.

RNA Tumor Viruses

In 1936 Bittner reported that an agent transmissible through milk was responsible for increasing the incidence of breast cancer in mice, resulting in the identification of the mouse mammary tumor virus. Since then, many other RNA viruses have been associated with tumors in animals. RNA tumor viruses are retroviruses, a class of viruses defined by the presence of an enzyme called reverse transcriptase. Using reverse transcriptase, the retrovirus produces a DNA copy of the viral RNA genome that can be integrated into the host cell DNA, leading to transformation. In human beings, the T cell leukemia viruses HTLV-I and HTLV-II are cancer- associated RNA viruses. Like HIV, these viruses specifically target CD4+ T lymphocytes, contribute to immunosuppression, and induce tumors of hematogenous origin.

Immunodeficiency and Cancer

The theory of immunological surveillance against cancer is that immune effector cells can eliminate cells that undergo malignant transformation. According to this theory, the development of a tumor is a failure of immune surveillance in maintaining tissue homeostasis. Despite advances in the understanding of carcinogenesis, the nature of immune surveillance and the role of the immune response cells in the progression of malignancy is unclear.

States of immunosuppression are associated with an increased risk of cancer. Patients receiving long-term immunosuppressive medication for prevention of allograft rejection have an increased incidence of skin cancers and lymphoid malignancies. The carcinogenic effect of ionizing and non- ionizing radiation may be due in part to radiation-related immunosuppression. Gamma rays and x-rays directly and indirectly affect the viability and function of lymphocytes. UV light induces immunosuppression indirectly, possibly through release of biologically active substances from cells in the skin. The coincidence of virally induced tumor formation and states of immunosuppression provides strong evidence that a normally functioning immune system acts to suppress carcinogenesis. For example, AIDS is associated with Kaposi's sarcoma, non-Hodgkin's lymphomas, and squamous cell carcinomas. AIDS- associated Kaposi's sarcoma and non-Hodgkin's lymphomas commonly carry fragments of the Kaposi's sarcoma-associated herpes-like virus (KSHV) genome, while AIDS-related squamous cell cancers are HPV- induced.

In experiments with carcinogen-induced murine tumors, loss of T lymphocytes does not significantly affect carcinogenesis but does diminish host response to transplantable tumors. Analysis of the immune cell participants in the surveillance against neoplastic progression reveals that in the normal host, several lymphoid cell compartments, including T, B, and NK lymphocytes, as well as macrophages, participate in the prevention and destruction of aberrant cells.

Genetic Alterations

The Multistep Hypothesis Cancer is fundamentally an alteration in the genes that control cellular function. It is clear that the development of a malignant phenotype is multifactorial. Some abnormal genes (termed cancer susceptibility genes) can be inherited at conception as germline defects. The known cancer susceptibility genes affect the cell's ability to detect and repair genetic damage, alter immune surveillance for tumors, modify cellular metabolism of carcinogens, or regulate the growth of specific cell types. Other genetic changes, known as somatic mutations, are acquired through interaction with agents that alter the cellular genome, such as radiation, mutagenic chemicals, or viruses. Although the exact combination of genetic changes required for carcinogenesis is not fully understood, evidence suggests that carcinogenesis requires the successive accumulation of genetic defects that result in the altered cellular growth and differentiation characteristic of a malignant phenotype. The genetic changes implicated in this "multistep hypothesis" of cancer development are best understood for colorectal cancer (Fig. 9-9). In 1988 Vogelstein and associates published a report describing specific mutations in colorectal cancer and defined their relationship to the adenoma-carcinoma sequence. This report led to wide acceptance of the multi-step hypothesis as the basis for malignant transformation and gave a genetic perspective to the processes of tumor initiation, promotion, and progression. Data suggest that not only are specific mutations essential for carcinogenesis, but also the order in which these defects are acquired is important. Although the sequence of events is not completely understood, the multistep hypothesis gives us a useful framework for defining the genetic nature of carcinogenesis.

Oncogenes

Oncogenes are genes that promote the transformation of normal cells into tumor cells, generally by activating growth-enhancing intracellular signaling pathways. Oncogenes can be found within mammalian cellular DNA (cellular oncogenes) and in viruses (viral oncogenes). Oncogenes are designated by three-letter names, which are usually derived from the tumors or the cell line in which the oncogene was first identified. For example, the first identified RNA virus, the Rous sarcoma virus, contains the *src* oncogene. Oncogenes derived from viral genomes are labeled with the prefix "v" (e.g., v- *src*); cellular oncogenes are labeled with "c" (e.g. c- *src*). Proto-oncogenes are normal components of the genome that function as oncogenes when altered or inappropriately expressed through processes such as chromosome rearrangement or viral transformation. Oncogenes encode proteins, sometimes termed oncoproteins, that alter cell cycle regulation, resulting in tumor formation.

The discovery of RNA retroviruses in the early twentieth century provided the foundation for our understanding of oncogenes and their precursors, proto-oncogenes. After infection, retroviral RNA is transcribed into double- stranded DNA by the viral enzyme reverse transcriptase. The retroviral DNA then is integrated into the chromosomal DNA of the host, forming a DNA provirus. The viral DNA is transcribed and translated by host cell machinery into viral RNA and viral proteins, which are assembled within the host cell. The new virus particles are released from the host cell by budding from the plasma membrane. Oncogenic viruses contain genetic segments that lead, on integration into the host genome, to neoplastic transformation of the host cell. Retroviruses also can induce tumors by activating host cellular proto-oncogenes. For example, the *src* gene encodes a membrane-associated enzyme responsible for phosphorylation of cellular proteins, and viral transformation through alteration of this gene results in abnormal intracellular signaling that contributes to tumor development.

Early genetic studies showed that tumors contain many chromosomal abnormalities, including gain or loss of chromosomes (aneuploidy), or alterations in individual chromosomes, such as translocations or amplifications. The role of the Philadelphia chromosome in human chronic myelogenous leukemia is an example of oncogene activation through chromosomal rearrangement. In the Philadelphia chromosome, a segment of chromosome 9, carrying the cellular oncogene c- *abl*, is attached to chromosome 22, within a gene known as *bcr*. The expression of the resulting *bcr-abl* fusion protein is important in the development of leukemia, an observation confirmed in animal studies showing that introduction of endogenous *bcr-abl* sequences causes leukemia (Fig. 9-10).

Oncogenes are divided into categories depending upon the role their proteins play in cellular function. These include growth factors, growth factor receptors, cytoplasmic protein kinases, *guanosine triphosphate (GTP)* -binding proteins, nuclear transcription factors, and cell cycle regulators (Table 9-5).

Tumor Suppressor Genes and the Inherited Cancer Predisposition Syndromes

The normal effect of tumor suppressor genes is to keep cellular growth in check. The loss of function of these genes leads to tumor formation (Table 9-6). The classic example of a tumor suppressor gene is the retinoblastoma gene (RB1). Each normal cell has two copies of the RB1 gene, and loss of RB1 gene function requires mutation of both copies. In

familial retinoblastoma, affected individuals inherit one defunctionalizing germline mutation in RB1 (Fig. 9-11). Expression of the RB1 mutation phenotype requires loss of the second allele by somatic mutation, a concept known as Knudson's "two-hit" hypothesis after Knudson's description of familial retinoblastoma in 1971. Most of the inherited cancer predisposition syndromes described to date involve inheritance of one mutant and one normal allele of a tumor suppressor gene. Analysis of tumors arising in individuals with germline mutations in tumor suppressor genes suggests that additional mutations, such as in the oncogenes *ras*, *c-myc*, etc., are required for expression of the cancer phenotype, an observation in keeping with the multi-step hypothesis of carcinogenesis. This theory is further supported by variability in phenotypic expression and in time to development of cancer among the members of these families. It is probable that modifier genes or environmental influences contribute to this variability.

Hereditary Retinoblastoma and the RB1 Gene

Retinoblastoma, a pediatric retinal tumor, occurs in both sporadic and familial forms. The sporadic retinoblastomas develop within the first 7 years of life and are generally unilateral. In contrast, familial retinoblastomas usually present within the first year of life and are characterized by multiple tumors involving both eyes. The penetrance of the RB1 mutation in these families is over 90 percent. The retinoblastoma gene product is a nuclear protein important for cell cycle regulation. The RB1 gene product (Rb protein) binds to, and sequesters, cyclin D as well as E2F, a transcription factor leading to the G₁-S cell cycle transition. Rb inhibits the ability of E2F to bind DNA and initiate transcription of the genes for DNA synthesis, such as *c-myb* and *c-myc*. A mutation in RB1 promotes unregulated cell growth through increased E2F activity.

Li-Fraumeni Syndrome and p53

In 1979, p53 was first identified as a cellular nuclear protein associated with the SV40 tumor virus. It was not until 1989 that p53 was recognized as a tumor suppressor gene and identified as a germline mutation associated with Li-Fraumeni syndrome, a familial clustering of breast cancer, soft tissue sarcomas, brain tumors, osteosarcoma, leukemia, and adrenocortical carcinoma. Affected individuals develop cancer by age 70 through somatic loss of the wild-type p53 allele.

Inactivation of the p53 gene is one of the most detectable genetic defects in tumors. The p53 protein plays a crucial role in preserving the integrity of the cell's genome by temporarily halting the cell cycle in response to damage, allowing adequate time for DNA repair prior to replication. In instances of severe damage, the p53 protein is capable of triggering programmed cell death, consequently eliminating damaged cells before replication can occur. By stimulating synthesis of nuclear proteins, such as Gadd45, p53 may indirectly facilitate DNA repair (Fig. 9-12). Given its importance in preventing replication of a damaged genome, p53 has been christened the "guardian of the genome"; intact p53 function is crucial for tumor prevention.

Familial Adenomatous Polyposis and the APC Gene

The *adenomatous polyposis coli (APC)* gene mutation was first localized in 1991 in patients with a rare autosomal dominant form of inherited colorectal cancer known as familial adenomatous polyposis (FAP). FAP is characterized by the development of multiple intestinal adenomas and nearly 100 percent progression to colorectal carcinoma within the life span of untreated individuals. FAP is a systemic disease with a phenotype that may include duodenal adenomas or carcinoma, desmoid tumors, mandibular osteomas, congenital hypertrophy of the pigmented retinal epithelium, and cutaneous epidermoid tumors. FAP results from a germline mutation in the *APC* gene, located on chromosome 5q21, leading to production of a truncated APC protein. Although the exact function of the APC protein is unknown, this intracellular molecule binds to, and induces, the degradation of b-catenin, a glycoprotein associated with both cell-cell adhesion and intracellular signaling. Mutations in the APC gene have been found in over 80 percent of sporadic colorectal cancers, suggesting that this mutation, like that of p53, is important in the pathogenesis of sporadic tumors.

Hereditary Malignant Melanoma

Studies of melanoma prone families (familial malignant melanoma) led to the localization of the responsible gene. This gene, identified on 9p21, encodes a protein known as p16, belonging to a family of proteins that inhibit the activity of *cyclin-dependent kinases (CDKs)*. Cyclin-dependent kinase inhibitors, like p16, modulate the mechanism that drives cells through mitosis and DNA replication. CDK inhibitor genes act as tumor suppressors, and defunctionalizing mutations in CDK inhibitor genes are associated with tumor development.

Genetic analysis of melanoma-prone Dutch kindreds revealed intermarriage between two families with germline p16 mutations. This provided the unique opportunity to study two individuals homozygous for p16 mutations at birth. One of these individuals survived to the age of fifty-five with no evidence of melanoma, and is thought to have died of an intraabdominal adenocarcinoma. The other individual developed three malignant melanomas by age fifteen. This unusual human experiment demonstrates that the penetration and expressivity of the p16 mutation is highly variable. The overall contribution of p16 mutations to melanoma incidence is low; studies of weakly familial and sporadic melanomas demonstrate a very low rate of p16 mutation.

Multiple Endocrine Neoplasia

The *multiple endocrine neoplasia (MEN)* syndromes are characterized by autosomal dominant inheritance of altered proliferation in specific endocrine glands. There are three distinct MEN syndromes: (1) MEN I, characterized by pituitary, parathyroid, and pancreatic endocrine tumors; (2) MEN IIA, characterized by pheochromocytoma, medullary thyroid cancer, and parathyroid tumors; and (3) MEN IIB, characterized by pheochromocytoma, medullary thyroid cancer, ocular and oral neuromas, gastrointestinal ganglioneuromatosis, and marfanoid body habitus. The gene for MEN I has been mapped to chromosome 11q13, although its specific location has not been identified. Pedigree and tumor tissue studies of MEN I suggest that this gene belongs to the tumor suppressor gene family. MEN IIA and IIB constitute genetically separate syndromes, which are governed by an oncogene rather than a tumor suppressor gene. MEN II families have dominant germline mutations in the *ret* proto-oncogene, which encodes a growth factor receptor.

Familial Breast and Ovarian Cancer

Cancer susceptibility genes associated with breast and ovarian cancer include BRCA-1 and BRCA-2, presumed tumor suppressor genes on chromosomes 17q21 and 13q12, respectively. In families with a high rate of breast cancer, i.e., at least four cases per family, approximately 50 percent of affected individuals have mutations in BRCA-1 and 30 percent in BRCA-2. In families with high rates of both breast and ovarian cancers, 75 percent are attributable to BRCA-1 and 23 percent to BRCA-2. BRCA-2 is most often associated with male breast cancer, and inherited mutations in BRCA-2 may be involved in 15 percent of all male breast cancer. Families with germline mutations in BRCA-2 also are at increased risk for prostatic malignancy.

In the high-risk families comprising the linkage studies of BRCA-1, inheritance of a BRCA-1 mutation was associated with a greater than 80 percent lifetime risk of breast cancer in the female family members. Studies of Ashkenazi Jewish families bearing a particular BRCA-1 mutation, 185delAG, demonstrate overall penetrance of this mutation at 38 percent by age forty, 69 percent by age fifty, and 85 percent by age sixty-five. The functions of the BRCA-1 and BRCA-2 gene products are unknown.

Epidemiologic studies suggest that mutations in the gene associated with ataxia telangiectasia (AT) are also associated with an increased risk of breast cancer. Approximately 2 percent of the population are heterozygous carriers of a defective AT gene allele, and the proportion of breast cancer attributable to AT gene carriers is approximately 4 percent.

Hereditary Nonpolyposis Colorectal Cancer

Hereditary nonpolyposis colorectal cancer (HNPCC) was first described by Aldred Warthin, a pathologist at the University of Michigan, in 1895. His seamstress died at an early age of cancer of the "female organs" and was a member of a family with a high incidence of colon, endometrial, and gastric cancer. In 1966 Henry Lynch described what he called the cancer family syndrome in a study of two large Midwestern families with a high incidence of colon cancer. In 1995 the International Collaborative Group on HNPCC proposed the Amsterdam criteria as a definition for HNPCC (Table 9-7).

The molecular basis for *HNPCC* was characterized by linkage studies of known colon cancer susceptibility genes, such as p53 and *APC*, in families that met the Amsterdam criteria. Although no linkage to these genes was found, researchers discovered a high rate of mutation in DNA microsatellites, which are short repetitive sequences of mono-, di-, or tri-nucleotides interspersed throughout the normal genome. They postulated that the HNPCC-associated mutations producing this "microsatellite instability" encoded factors required for DNA replication and repair. A 1993 study used microsatellite markers to map HNPCC to 2p15-16, and a subsequent study linked HNPCC to mutations on 3p. It is thought that over 70 percent of cases of HNPCC involve mutations in one of four identified mismatch repair genes. These genes encode proteins that recognize, excise, and repair mismatched nucleotides, preserving the integrity of the genome. A defect in mismatch repair theoretically sets the stage for accumulation of the multiple uncorrected mutations required for cancer development.

An inherited tendency to develop cancer is thought to be associated with 1 to 5 percent of all colorectal cancers, 5 to 10 percent of breast cancers, and 5 to 10 percent of ovarian cancers. Many of the inherited cancer predisposition syndromes involve germline mutations in genes that also contribute to the development of sporadic cancers, such as p53, *APC*, and MLH-1. Families with inherited cancer predisposition syndromes provide the most striking examples of tumor suppressor genes in action, but most tumors associated with tumor suppressor gene dysfunction are sporadic. These syndromes, however, present the opportunity to study carcinogenesis in models that are relevant to more common, nonfamilial malignancies.

Functional Alterations in Carcinogenesis

Tumor progression involves acquisition of several abilities by the malignant colony. These cells must be able to (1) invade the basement membrane and surrounding tissues through the production of proteases, (2) recruit blood vessels to support the growth of the tumor mass, (3) avoid destruction by effector cells of immune surveillance, such as NK cells, (4) move through tissues, a process that requires production and recruitment of cell adhesion molecules and chemotactic cytokines, and (5) travel to distant sites, adhere, and establish a new tumor colony. There are many parallels between fetal development and malignant transformation. During development, the behavior of individual cells is determined by regulated developmental programs that result in the formation of cohesive cell colonies. The process of tumorigenesis involves the disruption of these normal developmental programs. This principle is clearly illustrated in the development of a teratoma, where the resulting tumor attempts to recapitulate an intact organism.

Normal embryonic development and tissue homeostasis depend on specifically timed intracellular communication. Much of this communication results from the production and release of growth factors, such as *epidermal growth factor (EGF)*, as well as signals delivered by adhesion molecules through cell-cell contact. Tumorigenesis is characterized by the breakdown of these growth regulatory interactions, a process that involves both the activation and expression of proto-oncogenes, and the inactivation of suppressor genes. Although tumors have long been thought of as autonomous tissues, it is clear that a tumor could not exist without its surrounding stroma. Certain tumors require particular environments in which to grow, as evidenced by the characteristic metastatic locations of many solid tissue tumors. An obvious example of a tumor-host dependency is angiogenesis, but there are many other possibilities for tumor growth regulation by the microenvironment of a tumor. The result of carcinogenesis is the disruption of the normal homeostatic relationship between the transformed cell and its surrounding normal tissues.

Relationships Between Tumors and Normal Host Tissues Local Invasion

During the transition from in situ to invasive carcinoma, tumor cells must cross the basement membrane and enter the surrounding stromal tissue. The intact basement membrane, a dense matrix of collagen, glycoproteins, and proteoglycans, does not contain pores large enough for intact cells to penetrate, and hence tumor invasion must involve partial destruction of this barrier. On the other side of the basement membrane is the tissue stroma, a collection of fibroblasts, myofibroblasts, and other stromal cells. These cells actively

participate in tumor invasion. The process of local tumor invasion can be divided into three steps: tumor cell adhesion, matrix dissolution, and migration.

Tumor Cell Adhesion

The growth of normal tissues is characterized by cell-cell adhesion, a process linked to cellular proliferation and differentiation, such that a cell losing this contact undergoes involution. During carcinogenesis, the requirement for cell-cell adhesion is lost, and the single cell infiltrates local tissues in a relatively independent manner. Cell-cell adhesion and the growth regulation provided by this contact are mediated by cell adhesion molecules (CAMs), which are complex transmembrane glycoproteins present on the surface of both epithelial and stromal cells. These adhesion molecules also interact with the surface of the basement membrane, which contains receptor glycoproteins such as laminin, type IV collagen, and fibronectin.

CAMs are divided into four main classes according to their structure and general function. These include the cadherins, the integrins, the selectins, and the immunoglobulin superfamily receptors. In addition, other cell surface molecules, such as the *major histocompatibility (MHC)* molecules and CD44, contribute to tumor cell adhesion (Fig. [9-13](#)).

Cadherins

Cadherins are transmembrane glycoproteins that mediate calcium- dependent homophilic cell-cell adhesion. There are three major classes of cadherins: E-cadherins, mediating adhesion between epithelial cells, N- cadherins, controlling adhesion in nerve, cardiac, and pulmonary tissues, and P-cadherin, providing adhesion in intestinal, cardiac, and pulmonary tissues. The intracellular portion of the cadherin is associated with the actin cytoskeleton as well as with molecules involved in transmitting growth regulatory signals to the nucleus. Analysis of human epithelial tumors reveals a consistent loss of E-cadherin expression, and cell line transfection studies suggest that reintroduction of E-cadherin into tumor cells inhibits their invasive ability.

Integrins

Integrins are heterodimeric transmembrane glycoproteins that mediate adhesion between epithelial cells and extracellular matrix proteins. More than 20 different types of integrins have been identified, many of them with tissue specificity. Integrins commonly recognize extracellular matrix proteins by a tripeptide sequence of Arg-Gly-Asp present in fibronectin, vitronectin, and a variety of other adhesion proteins. The surface expression of integrins and their interaction with extracellular proteins can be modulated by cytokine factors such as *tumor-necrosis factor (TNF)*, interferon-gamma (IFN-g, or *transforming growth factor- b*1 (TGF-b1). The binding of integrins to their ligands results in transmission of signals for cellular proliferation and differentiation. The process of tumor invasion or metastasis may involve loss of appropriate integrin-mediated adhesion and signaling, resulting in cellular activation and establishment of tissue-specific metastatic colonies.

Selectins

Selectins are adhesion molecules present on endothelial cells and hematopoietic cells. These cell surface molecules govern the processes of leukocyte adhesion and homing. Three types of selectins have been described: (1) E-selectin, or endothelial-leukocyte adhesion molecule-1, (2) P-selectin, or granule membrane protein 140, and (3) L-selectin, or leukocyte-specific selectin. E-selectin and P-selectin are expressed on activated endothelial cells and mediate the binding of leukocytes to endothelial cells. L-selectin, found on the surface of leukocytes, controls the homing of leukocytes to lymphoid organs. E-selectin has demonstrated an ability to recognize Le ^x (SLX) antigen, a complex carbohydrate that is expressed at high levels in metastatic colorectal cancers. Altered expression of cell surface carbohydrates may be one of the mechanisms of tumor progression.

Immunoglobulin Superfamily

The Ig superfamily is characterized by an immunoglobulin-like extracellular domain, and fibronectin type III repeats. These molecules mediate homotypic and heterotypic cell-cell interactions. Neural cell adhesion molecule (N-CAM) is one example of the Ig superfamily. The protein product of the DCC gene, located on chromosome 18q, shows significant homology to N-CAM. This gene is mutated in metastatic colorectal cancers, suggesting that normally functioning DCC protein is important in preventing cell invasion and/or metastasis.

Several other cell surface molecules contribute to the cell-cell interactions governing normal tissue behavior. These include major histocompatibility (MHC) molecules, particularly MHC class I. MHC class I molecules complex with endogenously processed peptides, such as viral proteins, and target abnormal or infected cells for destruction by immune effector cells. One of the ways that abnormal tumor cells hide from immune surveillance is by altered expression of MHC molecules. Another cell surface molecule with a special role in tumorigenesis is CD44, a protein present in various isoforms created by alternative mRNA splicing. The CD44 splice variants are thought to facilitate lymphocyte recirculation and homing to lymphoid tissues. In addition, CD44 is present on the surface of epithelial cells, and particular isoforms are overexpressed on both primary and metastatic tumors.

Matrix Dissolution

After adhesion of tumor cells in experimental models of tumor infiltration, a localized zone of lysis is produced in the basement membrane near the point of tumor cell contact. This matrix lysis occurs from 2 to 8 hours after tumor cell attachment as a result of enzymes secreted directly by the tumor cell and by stromal fibroblasts. These enzymes belong to a family known as the metalloproteinases (MMPs). Examples of these enzymes include interstitial collagenases, type IV collagenases, and stromelysins. Natural protease inhibitors, known as tissue inhibitors of metalloproteinases (TIMPs), produced either by the host or by the tumor itself, can counteract this process. Once the basement membrane barrier is lysed, tumor cells are free to migrate into the surrounding stromal tissue (Fig. [9-14](#)).

Migration

In addition to adhesion and proteolysis, active tumor motility is required for the invasion and metastasis of tumors. In a dramatic in vivo study by Wood in 1958, microcinematography was used to observe directly the migration of rabbit carcinoma cells from the vasculature of the rabbit's ear. Like most cellular functions, tumor cell motility involves both fixed cell surface interactions and soluble factors. Morphologic studies show that, like leukocytes, tumor movement is characterized by ameba-like pseudopod extension. This movement requires coordination of multiple steps, including cellular protrusion and new adhesion formation at the leading edge and release of old adhesive interactions at the trailing edge. A variety of cytokines stimulate motile responses in tumor cells. These include tumor cell-derived cytokines, such as autocrine motility factor, autotaxin, and scatter factor (Table [9-8](#)). Many adhesion molecules, particularly those found in the extracellular matrix, such as laminin, collagen, fibronectin, and thrombospondin, serve as tumor cell attractants in motility assays.

Angiogenesis

Angiogenesis, or the formation of new blood vessels, is important for all phases of tumor progression (Fig. [9-15](#)). Without new vessel growth, tumors would quickly outstrip their local nutrient supply and would be unable to form new colonies after distant metastasis. Like all aspects of tumor growth, the process of angiogenesis involves complex signals delivered between the tumor cell and the host environment. At the initiation of angiogenesis, endothelial cells in vessels near the tumor site are stimulated to degrade the extracellular matrix. This is followed by migration of endothelial cells into the perivascular stroma in the direction of the angiogenic stimulus, initiating a capillary sprout that eventually forms a tubular vessel. Growth of a tumor colony beyond 1 cm ³ in size requires vascularization of the tumor through angiogenesis. Angiogenesis is partially mediated by soluble locally secreted factors. These factors include: basic fibroblast growth factor (bFGF), acidic FGF (aFGF), vascular endothelial growth factor (VEGF), platelet-derived endothelial cell growth factor (PD- ECGF), transforming growth factors a and b (TGF-a and -b), angiogenin, *TNF*-a, and interleukin-8 (IL-8) (Table [9-9](#)).

Tumor angiogenesis is a complex process, only partially explained by the presence of angiogenic factors secreted by tumor cells. Tumor cells themselves do release angiogenic factors such as VEGF and FGF as well as several that have not yet been completely purified or characterized. Tumor angiogenesis may occur through down-regulation of normally secreted angiogenesis inhibitory substances, such as thrombospondin. Tumors may be able to recruit macrophages and then activate them to secrete angiogenic factors such as *TNF* and IL-8.

The development of angiogenic potential in a tumor may be an important indicator of its biologic behavior. For example, the presence of increased vascularity in early (Stage I or II) breast cancer specimens indicates a significantly higher chance of tumor recurrence. Angiogenesis in early stage breast cancer is also correlated with the presence of lymph node metastasis during initial operative staging, indicating that angiogenesis is closely linked with metastatic potential. It has also been suggested that development of metastatic potential is acquired at a critical density of blood vessels in a primary breast cancer lesion. These observations raise the possibility that antiangiogenesis agents, such as analogues of the fungus-derived angiogenesis inhibitor fumagillin, might prove beneficial in the treatment of malignancy.

Metastasis

Metastatic tumors develop as clones arising from a heterogeneous primary tumor. It is assumed that when a cell within a tumor reaches a particular stage of virulence, it acquires the ability to metastasize. Recent evidence suggests that tumor cell metastasis requires multiple host-tumor interactions, and the process of metastasis probably begins early in the growth of the primary tumor. The metastatic cell must be able to do more than simply grow and divide in a new location; it must also break away from the original tumor population, invade through the basement membrane into a blood vessel, travel and adhere to a distant site, and induce angiogenesis. These activities require coordinating the processes of proteolysis, motility, adhesion, growth factor responsiveness, and angiogenic activity. Since all of these processes are naturally occurring functions of growth and development, the basic defect of metastasis must lie in the aberrant regulation of these processes.

Adhesion is an obvious component of metastatic behavior, both in the initial loss of attachment to the primary tumor cell mass and in the preference of certain tumors for metastatic sites. Despite advances in our understanding of the mechanisms of cellular adhesion, it is still difficult to define the adhesion requirements for metastasis. Many of the adhesion events previously described are implicated in the development of metastatic lesions. An example of this is the interaction of fibronectin, collagen, and laminin in the extracellular matrix with tumor cell receptors of the integrin and nonintegrin varieties. Some organ-specific adhesion events have been identified, supporting the hypothesis that specific adhesion is the reason for the preferential metastatic sites of certain tumors. Both in vitro and in vivo studies show that interfering with adhesion molecules can inhibit metastasis.

The development of metastatic potential in a tumor cell clone is correlated with the presence of proteolysis, a process already described for its role in tumor cell invasion and migration through tissue. Blockade of certain proteases, such as urokinase-type plasminogen activator, prevents establishment of metastasis. Compared with nonmetastatic cells, metastatic human melanoma cells express higher levels of tissue factor, the major cellular initiator of the plasma coagulation protease cascades. Tissue factor may promote metastasis by allowing increased tumor-endothelial cell adhesion as well as facilitating transmigration of tumor cells across the endothelium.

Metastasizing tumors have a predilection for selected organ sites (Table [9- 10](#)). In human beings, colon tumors frequently metastasize to the liver, renal cell carcinoma to the lung, melanoma to the lung and brain, prostate cancer to the bone, breast cancer to the lung, liver, brain, and bone, and ocular melanoma to the liver. Selection of a metastatic site by a particular tumor is probably governed by the adhesion and growth factor characteristics of the metastatic site and the requirements of the metastatic tumor. For example, metastatic melanoma cells express more of the integrin VLA-4 than do nonmetastatic cells, suggesting that this adhesion molecule and its ligand, V-CAM, may have some role in determining the specific endothelium for locating the metastatic deposit (Fig. [9-16](#)).

The role of the regional lymphatics in neoplasia in general and in metastasis in particular is as controversial as it is important. Early theories of cancer progression postulated that

regional lymph nodes provided a filter that was an effective, but temporary, barrier to the spread of tumor cells. It is now thought that the properties of the tumor cells themselves, rather than the filtration capacity of the lymph nodes, determine whether neoplastic cells are trapped within nodes or allowed to disseminate. The regional nodes may also be involved in the initiation of systemic immunity to tumors. For example, adoptive transfer of lymphoid cell populations derived from regional lymph nodes has been shown to induce tumor allograft immunity in normal animals. Among the many tumor metastasis models, however, there are substantial differences in the quality and quantity of immune responses to tumors, and the functional contribution of regional nodes to systemic tumor immunity is unclear.

Tumor cell entrapment in the capillary bed of distant organs is a necessary prelude to secondary tumor growth. Although the morphologic aspects of tumor cell entrapment are studied extensively, little is known about the dynamics of the process. Exposure of the capillary basement membrane is a result of the normal and continuous physiologic process of endothelial cell- shredding and may allow adhesion of tumor emboli. Platelet adherence to damaged regions of the basement membrane, followed by degranulation, causes further retraction of endothelial cells and augments attachment of tumor emboli or platelet-tumor cell emboli. Fibrin deposits around a tumor embolus are frequently observed, but the role of fibrin in tumor cell entrapment and metastasis is uncertain. Theoretically, a protective coat of fibrin surrounding the tumor embolus shields the neoplastic cells from the host immune response and blood turbulence. Increased coagulation is commonly observed in patients with cancer and may be related to the increase in thromboplastin found in tumors. Some neoplasms produce large quantities of procoagulant-A, which directly activates factor X of the clotting cascade. A reduced rate of blood flow in the vicinity of the tumor leads to increased trapping of circulating tumor cells and contributes to the survival of already trapped cells. The use of anticoagulants in the control of metastasis is based on these observations.

Studies using oncogene transfection provide a model for activation of the cellular responses that are required for tumor metastasis. Many oncogenes, when transfected into appropriate recipient cells, such as fibroblasts and epithelial cells, are capable of inducing metastatic ability. Some of these “metastatic” oncogenes include mutated *ras*, *v-src*, *v-fes*, *v-fms*, and p53. Lack of expression of the “metastasis suppressor” gene, *nm23*, may also be responsible for the development of metastatic disease. This gene was first identified for its decreased expression in murine melanoma tumor lines. Decreased expression of *nm23* is also associated with metastasis in breast, colon, renal cell, hepatocellular, lung, and squamous cell carcinomas. Although there is no definite proof that lack of *nm23* expression results in metastasis, this conclusion is supported by research showing that transfection of *nm23* into melanoma cell lines results in decreased tumor incidence, metastatic potential, and cytokine responsiveness. The exact mechanism of metastasis suppression by *nm23* is unknown, but evidence suggests that this gene is important for signal transduction during development.

Metastasis is a multistage process in which tumor cells acquire more and more autonomy regarding growth factor and adhesion requirements. A complex process with no universally applicable mechanism, metastasis is dependent on the characteristics of the tumor (the “seed”), and the microenvironment of its implantation site (the “soil”).

Autocrine Mechanisms Production of Tumor Growth Factors

The autocrine model of tumorigenesis holds that tumor formation, like the process of embryogenesis, occurs through a coordinated series of cellular signaling events, regulated by growth factors. Transformation of normal cells to malignant behavior may result from either increased production of stimulatory growth factors or decreased production of inhibitory growth factors. In 1896 Sir George Beatson in Glasgow demonstrated that oophorectomy could produce tumor regression in patients with breast carcinoma, suggesting that secreted steroid hormones support tumor growth. The molecular characterization of tumor growth factors began with the observation that polypeptides secreted by a variety of retrovirally, chemically, or oncogene-transformed cell lines could induce tumor growth or cellular transformation. The first correlation between an oncogene and a normal cellular protein was demonstrated when the amino acid sequence of the *sis* oncogene product displayed significant homology to the b chain of platelet-derived growth factor (PDGF).

One example of the interplay between stimulatory and inhibitory growth factors is that of the transforming growth factors TGF- α and TGF- β . TGF- α and TGF- β were originally named for their ability to reversibly induce anchorage-independent growth in fibroblasts. TGF- α belongs to a family of growth-stimulatory cytokines, all of which share a common receptor with epidermal growth factor (EGF). As expected, these cytokines have similar activities, which include regulation of the growth and differentiation of normal and neoplastic epithelial cells. TGF- β , on the other hand, is an inhibitor for cells responding to growth factors of the TGF- α family. This protein normally provides autocrine growth inhibition of epithelial cells and also inhibits the in vitro growth of some tumor cell lines. Recent data suggest that the development of malignancy can be correlated with increased autocrine growth stimulation by TGF- α , or decreased autocrine growth inhibition by TGF- β .

Other growth factors reported to be involved in tumorigenesis include insulin-like growth factor-1 (IGF-1), platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), hepatocyte growth factor (HGF), fibronectin, transferrin, amphiregulin (a member of the *EGF* family), and interleukin-6 (IL-6) (Table [9-11](#)). IL-6 may be an inhibitor of tumor growth in early tumor stages, switching to a stimulatory factor for metastatic lesions. In one study, melanoma cells obtained from primary tumors exhibited growth inhibition upon in vitro culture with IL-6. Growth of metastatic melanoma cells, however, was stimulated significantly by culture in IL-6. Acquisition of the ability to proliferate in response to cytokines such as IL-6 may be one of the cellular events leading to metastatic disease.

Constitutive Activation of Growth Factor Receptors

In some tumors, loss of normal growth regulation occurs as a result of altered expression of growth factor receptors (Fig. [9-17](#)). Either constitutive activation of a growth factor receptor or an increase in growth factor receptor number would have the same effect as an increase in levels of the corresponding growth factor, e.g., an increase in receptors for epidermal growth factor (EGFR) may contribute to tumor growth in skin, colon, and breast cancers. The *erbB* oncogene product is a transmembrane receptor similar to the EGF receptor except that it is constitutively activated, i.e., able to generate a mitogenic signal in the absence of EGF binding (Fig. [9-18](#)). Expression of this oncogene is associated with poor prognosis in breast cancer. Several other receptors encoded by oncogenes, such as *kit*, *ros*, *met*, *ret*, and *trk*, demonstrate deletions of ligand-binding domains. The *MEN* IIA and IIB syndromes are associated with dominant activating germline mutations of different regions of the *ret* proto-oncogene. Murine embryo studies demonstrate that this gene encodes a growth factor receptor found on subsets of cells in the central and peripheral nervous system. In addition, *ret* is consistently expressed in both familial and nonfamilial neuroblastomas, pheochromocytomas, and medullary thyroid carcinomas as well as in normal thyroid and adrenal glands.

An interesting demonstration of the cumulative effect of mutations on tumorigenesis is provided by the relationship between mismatch repair defects and *TGF- β* activity. As described previously, a characteristic of deficient mismatch repair is alteration of microsatellite DNA. The gene encoding the epithelial cell receptor for TGF- β , the TGF- β receptor type 2 (TGF- β RII), contains a microsatellite site consisting of a string of ten adenosine repeats. In individuals with *HNPCC* and mismatch repair defects this gene is commonly mutated and the resulting TGF- β RII is dysfunctional. This results in loss of TGF- β suppression of epithelial cell growth, and contributes to tumor formation.

Intracellular Signal Transduction

Binding of a ligand to a cell surface receptor results in an intricate cascade of intracellular reactions, ultimately inducing transcription of appropriate cellular genes. This complex process is known as intracellular signal transduction. Any alteration in the network of signals responsible for ordered cell growth can contribute to carcinogenesis, and most oncogenes or tumor suppressor genes encode proteins essential for intracellular signal transduction. Oncogenes can be roughly divided into groups according to their cellular function. The small subset of oncogenes comprising growth factors or growth factor receptors has already been discussed. The other oncogene categories include signal transducers such as cytoplasmic protein kinases and *GTP*-binding proteins, transcriptional regulators, and regulators of programmed cell death (apoptosis).

Signaling Through Growth Factor Receptors

Engagement of a receptor by an appropriate ligand induces the assembly of the primary second messenger proteins at the inner surface of the plasma membrane near molecules such as *ras* that are involved in subsequent branching pathways of signal transduction (Fig. [9-19](#)). Growth factor receptors, also known as receptor tyrosine kinases, consist of three domains: (1) a large amino-terminal extracellular ligand binding domain; (2) a short transmembrane helix; and (3) an intracytoplasmic carboxyl terminus that is a ligand-activated tyrosine kinase. Binding of a ligand growth factor to the extracellular domain results in receptor dimerization and phosphorylation of the intracellular portion of the receptor. This triggers recruitment of intracellular substrates characterized by *src* homology 2 (SH2) domains to the receptor phosphotyrosine residues. An example of such a substrate is the protein Grb2, which, when associated with the appropriate phosphotyrosine of an activated growth factor receptor, recruits the *Sos* protein, a nucleotide exchange factor. *Sos* then engages the *ras* protein and assists in its activation by exchanging *guanosine diphosphate (GDP)* for *GTP*. This activation is assisted by yet another factor, a GTPase-activating protein (GAP).

Once *ras* has been activated, it triggers a cascade of serine-threonine kinases, such as members of the *raf* gene family, and a group of cytoplasmic enzymes known as *mitogen-activated kinases (MAP kinases)*. These kinases mediate the reactions that bridge the gap between the cell membrane and the nucleus. Activation of the MAP kinases results in phosphorylation of nuclear transcription factors, such as the c- *fos* and c- *jun* products, which are components of a transcriptional activator named AP-1. c- *fos* and c-*jun* induce transcription of the proto-oncogene c- *myc*, a direct regulator of the cell cycle, and stimulate the cell to progress from G₁ into S phase, thus committing the cell to a round of DNA replication.

The signaling pathway of growth factor receptors is one of several different mechanisms leading to cellular activation. Receptors that lack intrinsic kinase activity transmit their signal through association with intracellular protein kinases, such as the product of a gene called *lck*. *ras* belongs to a family of proteins known as G proteins that transmit signals from cell surface ligands to effectors, such as adenylate cyclase or phospholipase A2, through conversion of *GTP* to *GDP*. Regulation of the activation states of *ras* and related proteins is a crucial point in signal transduction. To participate in signaling, *ras* protein must attach to the inner surface of the plasma membrane by linking to an isoprenyl group, a process known as farnesylation. Strategies to inhibit the *ras* farnesylation reaction, such as through inhibition of the enzyme farnesyl protein transferase (FPTase), are under investigation as new chemotherapy agents.

A *ras*-independent signaling pathway has been proposed whereby receptor engagement activates a phospholipase that hydrolyzes inositol phospholipids in the plasma membrane and activates two additional second messengers: (1) inositol triphosphate, which triggers a signaling cascade through intracellular calcium flux, and (2) diacylglycerol, a membrane-associated lipid that activates another protein kinase called protein kinase C (PKC). Activated PKC is then able to activate *raf*, producing *MAP* kinase activation and leading to transcription of c-*fos* and c-*jun*.

A recurring theme in studies of signal transduction is the involvement of protein kinases. Intracellular second messenger proteins must be phosphorylated by a tyrosine kinase in order to amplify and distribute signals for cellular growth and differentiation. Although the roles of these kinases in signal transduction are not well understood, studies of cellular activation associate both receptor and nonreceptor kinase activity to multiple cellular processes, such as proliferation, metabolism, and cytoskeletal function. Many more pieces of the puzzle are needed to fully understand the complex regulatory mechanisms governing signal transduction.

Cell Cycle Control

Cells take their cues for proliferation and differentiation not only from external sources such as growth factor receptors but also according to an internal program. The cell cycle encompasses the progression from G₁ phase through mitosis and is coordinated by nuclear proteins called cyclins. The passage of a cell through the cell cycle is tightly regulated by a network of controls that act on the transcription of cyclin genes, the degradation of cyclin proteins, and the modification of cyclin-dependent kinases by phosphorylation. It is

now recognized that the cell cycle is a dynamic process that includes periods of arrest of cell proliferation when DNA damage occurs, presumably to allow time for DNA repair to occur. Provisions are made within the cell cycle program for apoptosis in circumstances in which the cell's genome has undergone irreparable damage.

Coordination of the cell cycle is achieved through a network of cyclin- dependent kinases (CDKs), which undergo programmed changes in activation state (Fig. 9-20). When in an activated state, each CDK is a combination of a cyclin and a kinase as well as a variety of other associated proteins. After activation of mammalian cells, a succession of kinases, known as CDK4, CDK2, and CDC2, are expressed in conjunction with a series of cyclins (cyclins D, E, A, and B) as the cell progresses from G₁ to mitosis. Each of the cyclin genes is transiently transcribed during a particular phase of the cell cycle, its mRNA translated, and then the protein is rapidly degraded such that its residence time during the cell cycle is brief. The elements of control over the cell cycle include redundant systems for regulating each step of this process. In addition to mechanisms allowing passage of cells through the cell cycle in response to mitogenic stimuli, negative controls on cell cycle progression, known as cell cycle checkpoints, are important for cell differentiation and senescence (Fig. 9-21). Defects in these negative controls are associated with carcinogenesis. At least two checkpoints detect DNA damage: one at the G₁-S transition, and one at the G₂-M transition (Table 9-12).

Alterations at the Cell Cycle Checkpoints

At the transition from G₁ to S phase, the main target of the activated kinase is the Rb protein (Fig. 9-22). The Rb protein is bound to the transcription factor E2F during G₁. Upon phosphorylation, Rb releases E2F, which in turn activates the transcription of genes required for S phase. Loss of function of the Rb protein leads to unregulated entry of a cell into S phase.

Upstream of Rb-E2F, several proteins control the G₁-S transition. One of these is the p53 gene product. Damage to the cellular DNA induces production of p53 protein, which in turn activates p21, a protein that inhibits all of the cyclin- CDK complexes (Table 9-13). This results in a delay in the transition from G₁ to S that allows time for DNA replication to occur. A cell with deficient p53 will enter S phase without sufficient DNA repair and replicate uncorrected mutations. An additional cell cycle check is provided by p27, a protein induced by binding of TGF- β to the cell. p27 is also an inhibitor of cyclin-CDK complexes. Finally, p16 is a cyclin-dependent kinase inhibitor and regulator of G₁-S transition whose activation stimulus is not yet known; germline mutations in the gene encoding p16 are associated with familial malignant melanoma.

The G₂-M cell cycle transition is also incompletely understood, but it appears to involve the product of the ataxia telangiectasia (AT) gene. Individuals with AT exhibit a high frequency of chromosomal breaks after irradiation, and their cells do not undergo proper arrest in G₂ after radiation. Individuals carrying a germline mutation in the AT gene are at an increased risk for cancers, presumably through loss of the tumor suppressor function of the AT gene product.

Evidence for another type of cell cycle regulator comes from examination of specialized nucleoprotein structures known as telomeres. Telomeres are located at the ends of chromosomes and are essential for maintenance of chromosomal stability. In immortalized cell cultures, telomere length is maintained by the enzyme telomerase. Telomerase is absent, however, in almost all normal human cells. With repeated replication of the chromosome, the telomere ends are lost, causing telomeres to undergo progressive shortening in normal human cells. This process leads to a limited life span for normal cells, as cells lacking telomeric sequences, unable to replicate, become senescent. Telomeres are thought to represent an internal clock by which the life span of somatic cells is measured. Studies of human tumors show no loss of telomere size with repeated replication and an increase in telomerase activity, suggesting that escape from the regulation caused by telomeric shortening may be important in tumorigenesis.

Regulation of Apoptosis

In order for tissues to maintain a normal state, cells subject to renewal must involute so that the proliferation and involution rates are balanced. This “programmed cell death” is known as *apoptosis* and, in contrast to cell necrosis, is characterized by cytoplasmic shrinking, nuclear chromatin condensation, and DNA fragmentation. Defects resulting in loss of normal apoptosis are associated with tumor formation. For example, the *bcl-2* oncogene was identified in follicular lymphomas and found to promote cell survival rather than proliferation. A family of genes whose products interact with one another to regulate cell death by apoptosis includes *bcl-2* (Fig. 9-23). BAX, a protein identified by its association with the *bcl-2* gene product, counteracts the survival-promoting effects of *bcl-2*. Although many different stimuli are able to induce apoptosis, most result in involvement of the BAX/*bcl-2* interchange; for instance, p53 may serve as a cellular surveillance factor by inducing either growth arrest or cell death, depending upon the cellular circumstances. Transcription of BAX protein is inducible by p53. Therefore, p53 causes growth arrest through activation of the p21 gene, or apoptosis through induction of BAX. Other regulators of apoptosis include TNF-R1, a cell surface receptor that promotes apoptosis when activated by its ligand, TNF- β , and the interleukin-1 β - converting enzyme (ICE) family, a group of cysteine proteases capable of inducing apoptosis. Finally, various growth factors and cytokines such as IGF-1 may act on the apoptotic pathway to promote cell survival.

SURGICAL MANAGEMENT OF PRIMARY TUMORS

General Considerations

The surgeon's role in the treatment of tumors varies with the type of cancer but includes diagnosis, clinical staging (Table 9-14), operative resection for cure and local disease control, pathologic staging or palliation, and the management of medical conditions common to the cancer patient (Table 9-15). The major goal of treatment is to provide the best chance for cure with the best functional and cosmetic result. Primary curative operations may be performed on patients with breast, lung, uterine, and large-bowel cancer. In other forms of cancer, such as lymphoma, the surgeon's role may be diagnostic, with a possible contribution to staging; in patients with leukemia, the only surgical role may be provision of vascular access for long-term administration of chemotherapy or the management of associated complications. Knowledge of clinical and pathologic staging is important; accurate diagnosis and assessment of the extent of tumors are essential to appropriate treatment. Pathologic staging gives insight into the natural course of the neoplasm, which is necessary to determine further treatment plans and to evaluate their efficacy (Tables 9-16 and 9-17).

The surgeon uses physical examination, roentgenography, ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI) techniques to determine if the neoplasm is potentially curable by surgical removal (Table 9-18). Studies to preclude metastases are indicated when they are cost-effective and would substantially alter surgical treatment. The asymptomatic patient with a 1 cm breast mass and clinically negative regional lymph nodes should not require liver or bone radionuclide scan or abdominothoracic CT scans, as these tests would not be cost- effective. However, the use of CT scans of the chest in a patient with lung cancer can effectively help to stage the disease and to ascertain whether surgery with operative resection is appropriate.

Surgical resectability is also determined by the tumor's relation to, and degree of invasion into and around, vital structures. Invasive and noninvasive radiologic studies are extremely helpful in outlining the goals of operative management and defining the operative approach in specific patients e.g., portal vein combined with hepatic/splenic artery occlusion denotes nonresectability in patients with carcinoma of the pancreas, although surgical bypass may be necessary for biliary or gastric outlet obstruction.

The surgeon has a major role in disease prevention and patient/family counseling. The extent of an operation may be related to the presence of additional precancerous lesions or to a strong family history of site-specific cancers. For example, a patient with cancer of the colon and numerous colonic adenomas may be treated by a total abdominal colectomy rather than simple excision of the cancer-affected segment of large bowel.

In some circumstances, surgical removal of the contralateral breast may be used as a form of cancer prevention in a breast cancer patient who is at high risk for subsequently developing cancer in her other breast. The presence of lobular carcinoma in situ and breast cancer family syndrome are two circumstances in which contralateral mastectomy may be justified as a form of cancer prevention.

Operability and treatment decisions must take into account the patient's medical status and ability to tolerate the proposed operation. Assessment of cardiopulmonary status, hepatic and renal function, and nutritional status is vital to determining operative risk and to assessing needs for hospital and home rehabilitation.

An important but often underemphasized goal of cancer management is restoring the patient's physical, emotional, social, and employment status. The rehabilitation for a woman with breast cancer might be directed toward minimizing scarring and swelling of the tissues of the chest and arm, regaining strength and mobility in the shoulder after axillary lymphadenectomy, and restoring contour and symmetry of the breast. For some women an external prosthesis is satisfactory; other women significantly benefit from breast reconstructive surgery. Similarly, patients with extremity sarcomas may be considered candidates for limb salvage surgery or prosthesis use to maximize function (Fig. 9-24).

Most patients expect to understand and participate in the decision-making process, but many are confused by conflicting input from family, friends, and even other physicians. Anxiety and uncertainty about the life-threatening nature of cancer and the prospects of physical disfigurement may make it difficult for them to understand treatment issues and to contribute to the decision-making process. The physician should combine empathetic listening with a clear and comprehensive 20- to 30-minute discussion of treatment options. The use of patient education materials (written and videotaped) and counseling by nurses or physician assistants can facilitate the decision- making process and increase the patient's level of comfort and information about treatment planning.

Clinical Diagnosis

A complete history and physical examination is indispensable before further judgments can be made regarding laboratory testing and treatment. Common symptoms in the cancer patient may include dysphagia, nausea, anorexia, vomiting, hematemesis, abdominal pain, melena, and hematochezia in patients with gastrointestinal cancer; productive cough and hemoptysis in lung cancer; or an enlarging tender mass in breast and soft-part tumors. Symptoms generally correspond to the sites involved, but nonspecific symptoms, such as night sweats and weight loss, may be the initial manifestations of an underlying neoplastic tumor. The duration of symptoms may indicate the aggressiveness of the cancer or early case finding. The degree of physical impairment also influences treatment decisions regarding palliation.

The patient's past medical history often lends clues to the diagnosis. Diethylstilbestrol hormone use by the patient's mother during pregnancy, thymic irradiation for asthma, skin irradiation for acne in childhood, or a history of chronic inflammatory bowel disease are historical factors known to be associated with the later development of cancer. The medical history can also reveal environmental factors such as smoking, alcohol ingestion, or exposure to asbestos or aniline dyes, which can be related to organ-specific sites of tumor development. A thorough medical history provides an important index of operative risk for the patient and medications that may require adjustment in their dosage and frequency.

Inquiry into family history may reveal findings that support an initial diagnosis, suggest additional associated lesions, or influence the extent of surgical treatment. A patient with a thyroid nodule may have relatives with episodic hypertension, renal stones, or an “adrenal tumor” suggesting a multiple endocrine neoplasia syndrome. A patient with relatives who developed large-bowel cancer before the age of forty should be screened for polyposis syndromes. Without a thorough history, these and other genetically influenced diseases will be missed.

Asymptomatic patients who are at high risk for development of certain types of cancer by virtue of age, family history, or personal history of disease should be systematically screened. They should have appropriate examinations as indicated, using mammography, fecal occult blood testing, flexible sigmoidoscopy, cervical cytology examination, etc., as indicated. Simple screening examinations, such as uterine cervical Pap smear and fecal occult blood testing, should be performed routinely. Often the surgeon is asked to evaluate

a patient with a mass in a regional lymph node area, such as the neck, axilla, or groin. Usually this mass represents one or more lymph nodes enlarged secondary to an inflammatory process, originating in the area drained by the regional nodes. Lymphadenitis is often multiple and tender, and the primary area of inflammation is usually obvious. If the mass represents a solitary lymph node or several distinct enlarged painless nodes with no obvious cause, the question of biopsy to diagnose cancer is raised. Open excisional biopsy of lymph nodes should be deferred until a thorough search for the source of a primary tumor has been made, but fine-needle aspiration and cytology examination can be done without causing regional tissue contamination with cancer cells.

Excisional biopsy of lymph nodes containing metastatic cancer can release viable cancer cells into the biopsy wound because efferent and afferent lymphatics containing such cells are divided in the process of excising the lymph node. The incision of a node or group of nodes containing metastatic tumor may increase the risk of seeding the wound with cancer cells and violates the surgical principle of en bloc resection of the primary cancer, regional lymphatic channels, and regional lymph nodes. Once neoplastic disease is growing free in connective tissue and is no longer contained within the lymphatics or lymph nodes, operative eradication of the tumor becomes difficult because the anatomic limits of the area containing cancerous tissue are broadened. If adenopathy is present in the head and neck region, then the oral cavity, larynx, and pharynx should be thoroughly investigated. If the mass is in the axilla or groin, the breast and the skin of the corresponding extremity and portion of the trunk should be examined. The patient should be queried about previous removal of skin lesions, particularly moles, that may represent melanoma.

Laboratory and Radiologic Studies

In addition to the routine laboratory tests such as complete blood count, coagulation profile, multichannel serum biochemistry profile, and chest x- ray, other studies are useful in determining the prognosis in patients with malignant disease (Table [9-19](#)).

Evaluation of the patient with a suspicious breast lesion should include bilateral mammography. Roentgenographic characteristics may support the clinical diagnosis and reveal occult lesions in the contralateral breast. Clinically occult lesions may undergo mammographically directed stereotaxic needle biopsy or needle localization for open biopsy. The patient with breast cancer who has abnormal liver function tests or palpable hepatomegaly should undergo a CT scan; those with skeletal symptoms should receive a radionuclide technetium 99m bone scan. Metastatic lesions 1 cm or more in diameter can be detected in the liver using unenhanced and enhanced CT scans with 3–5 mm slices.

Patients who have symptoms or signs attributable to the gastrointestinal tract should receive an air contrast upper or lower barium study. Air contrast barium studies show greater mucosal detail than barium studies without air contrast and are more sensitive to small lesions. Endoscopy can also be used for both diagnosis and tissue examination. CT scans of the chest are useful in evaluating suspicious lesions noted on chest x-ray and skeletal and soft tissue lesions. Ultrasound and CT scanning can be particularly useful in the detection of hepatobiliary cancer. Invasion of vascular structures and dilated extrahepatic and intrahepatic ducts secondary to extrahepatic biliary obstruction by a neoplasm can be demonstrated. Coexisting ascites also may be noted. Magnetic resonance imaging is valuable in determining the location and extent of soft tissue and bone tumors and their relationship to adjacent structures. By quantitative imaging, the physical characteristics of tumors often can be determined.

Serum markers, such as *carcinoembryonic antigen (CEA)*, *CA19-9*, *beta- human chorionic gonadotropin (beta-HCG)* , and *alpha-fetoprotein (AFP)* are useful in the management of patients with specific tumors (Table [9-20](#)). CEA is a tumor-associated glycoprotein found in several different solid tumors such as breast, lung, gynecologic, and gastrointestinal cancers and sarcomas. Plasma CEA levels are not sensitive or specific enough to be of great value in screening for cancer, particularly for early lesions. Elevated plasma levels are correlated with increasing tumor size, stage, and the extent of positive lymph node metastases in patients with large-bowel cancer. When surgical removal of the primary tumor is complete, measuring circulating CEA levels may be useful during the follow-up period as an early marker of recurrent systemic disease.

Surgical Pathology

The importance of accurate pathologic diagnosis in the proper surgical treatment of cancer patients cannot be overstated. Determinations of the presence of cancer, the histologic grade, the site of the primary or metastatic foci, and surgical resection margins provide critical information.

Fine-needle aspiration cytology of a solid lesion can be performed with a 22-25-gauge needle to assess the cytology of palpable tumors. Diagnosis using fine-needle aspiration techniques concurs with surgical pathologic diagnosis in 97 percent of tested lymph nodes and 77 percent of breast tumors examined. This technique is rapid, minimally traumatic, and highly accurate for diagnosis of a clearly palpable mass or a radiographically visible lesion (Fig. [9-25](#)). False-positive results are rare; false-negative results may occur because of small sample size, site of the lesion, degree of tissue necrosis, and tumor type. With the use of fine-needle aspiration techniques in combination with roentgenography and ultrasonography, deep, nonpalpable lesions are amenable to diagnosis with minimal morbidity, e.g., transthoracic aspiration of lung nodules using CT scan and CT-directed aspiration of pancreatic masses. Fine-needle aspiration cytology is particularly useful in the diagnosis of palpable masses in the breast and thyroid, as well as palpable suspicious nodes in the neck, axilla, or groin.

Aspiration cytology cannot be completely depended on for grading of solid tumors, for subdividing types of lymphoma, or for accurate diagnosis after radiation treatment, but a positive diagnosis greatly facilitates diagnostic and treatment planning. Use of immunohistochemical markers may be of additional diagnostic value (Table [9-21](#)).

When an accurate diagnosis of tumor type and grade is necessary, an incisional or excisional biopsy is required. Care should be taken in the planning of a surgical biopsy so as not to jeopardize later surgical extirpation or the use of skin flaps. In general, large soft-tissue lesions (5 to 7 cm) that are deeper than the superficial fascia are best sampled by incisional biopsy (Fig. [9-26](#)). Small (<2 cm) superficial lesions can be managed by excisional biopsy with a view toward further treatment depending on tumor type, grade, and depth of invasion. Frozen section diagnosis should not be relied on to provide histologic grade of the tumor and information about depth of invasion. Surgical margins of resection can and should be evaluated by frozen section examination. In some instances, such as malignant rectal polyps, local excisional therapy is all that is required if permanent histopathology determines only superficial invasion, adequate resection margins, low-grade differentiation, and absence of vascular or lymphatic invasion.

Decision for Operation

A decision for curative operation presupposes that the tumor is localized or confined regionally, that the area of the tumor can be encompassed by regional excision, that evidence of distant metastases cannot be found, and that the tumor is appropriately treated by operation.

Given a decision for a curative operation, the extent of the surgical procedure must be defined. In principle, an en bloc resection should be performed, encompassing the primary tumor, regional lymph nodes, and intervening lymphatic channels. This principle is perhaps best illustrated by operations for large-bowel cancer, in which the regional lymphatics of the colon (but not the rectum) course in one direction with the major arteries and veins. This principle is less applicable in the rectum, where lateral spread and the limiting confines of the lateral pelvic wall may preclude a wide margin. Similarly, the principle is less applicable for breast cancer, for which multiple pathways of lymphatic spread are well recognized (e.g., internal mammary and retropectoral nodes, axillary and supraclavicular nodes).

As a result of these variables, and partly due to emergence of earlier stage disease, the extent of various operations for cancer is undergoing change. This “remodeling” is best illustrated in breast cancer, for which the trend is toward local excision of the primary tumor with radiation therapy and adjuvant chemotherapy. The therapeutic value of regional node dissection has been questioned by some. Performed properly, it is of clear prognostic value and could establish the database for precise staging for other adjuvant treatments.

Surgical procedures may be performed solely for staging purposes. An example is staging laparotomy in Hodgkin's disease, in which multiple lymph node biopsies, liver and bone marrow biopsies, and splenectomy are performed. In a variety of settings, operative intervention is used for palliative treatment. Bypasses are performed around obstructed viscera, and gastrojejunostomy is done for obstructing carcinoma of the stomach. Choledochoduodenostomy or jejunostomy procedures are used as bypass procedures for periampullary carcinoma of the pancreas obstructing the common bile duct. Intraluminal stent tubes may be inserted through an obstructing esophageal carcinoma to allow the patient to take oral nourishment, and colon diversion (bypass or colostomy) may be performed for obstructing large-bowel cancer.

The use of cytoreductive surgery is controversial. This approach may be most relevant in ovarian cancer, for which deliberate removal of as much ovarian tumor as possible may enhance the patient's response to subsequent chemotherapy or radiation therapy. Cytoreductive operations also may be valid in the treatment of some childhood tumors, such as neuroblastoma. Rarely is cytoreductive surgery (“debulking”) applicable in other circumstances.

The extent of an operation should be dictated by precise staging. With precise staging, and with the presence of effective chemotherapy, operative intervention may be more limited. For example, the traditional management of non-seminoma germ cell tumor of the testis has been radical orchiectomy and retroperitoneal lymph node dissection extending to the renal vasculature. The predictability of normal levels of b- *HCG* and *AFP* with normal CT scans and ultrasound, backed by effective chemotherapy should dissemination occur, now may permit only radical orchiectomy without node dissection in those patients who are normal by these determinants. This approach results in fewer patients suffering adverse consequences from retroperitoneal node dissection and resultant nerve damage with impotence and retrograde ejaculation. More precise pretreatment planning of tumors traditionally treated by operation is a main objective of current research. Research in areas such as lymphoscintigraphy and the use of radiolabeled monoclonal antibodies is being explored to achieve greater precision of staging.

Preoperative Preparation

Appropriate preoperative preparation includes knowledge of the natural history of the patient's disease, evaluation of operative risk, decisions relevant to the need for and timing of operative intervention, estimate of the physiologic stress potentially imposed by the operation, and the quantitative assessment of the patient's physiologic status. Comprehensive preparation of a patient for operative therapy requires both physiologic and psychologic support. In the preoperative preparation of the surgical patient, the physician should consider potential abnormalities such as acid-base disorders, malnutrition, infection, respiratory insufficiency, hepatic and renal dysfunction, anemia, and clotting abnormalities.

The major aims of preoperative therapy should be to prepare the patient to withstand the stresses of operative therapy and to minimize the risks of the surgical procedure. The appropriate duration of the preoperative period depends on the urgency of the operative procedure. The length of time feasible for correction of preexisting deficits may be *short* (hours), as in the case of massive gastrointestinal hemorrhage or free perforation of an obstructed colon; *intermediate*, as in complete but nonstrangulated bowel obstruction; or *prolonged* (days), as in jaundice due to pancreatic carcinoma. Determination of the urgency of a particular operative procedure requires knowledge of the operative risk and the natural history of the disease without immediate operative intervention. Included in this evaluation are factors such as the age of the patient (chronologic and physiologic), degree of physiologic derangements and nutritional deficits, presence of organ system failure or insufficiency, and stage of the primary disease. Although the urgency of operative intervention may limit the length of preoperative preparation and the methods available for correcting preexisting abnormalities, partial repair of deficiencies should be initiated promptly with plans made for more complete correction during and after operative therapy.

Months of chronic undernutrition cannot be corrected in a matter of hours, but anemia, dehydration, and electrolyte abnormalities can be ameliorated by early initiation of intensive intravenous support, guided by appropriate laboratory monitoring.

Anemia is common among patients with cancer and should be corrected with packed red cell transfusions to a hematocrit level of 30 percent or more before major operative intervention when much bleeding is anticipated. When possible, patients should autodonate one or two units of blood within 4 weeks of surgery to decrease the risk of viral transmission from nonautologous blood transfusions.

Clotting abnormalities can usually be determined by an adequate patient history, physical examination for evidence of hemostatic problems, and routine laboratory investigation (prothrombin time, partial thromboplastin time, and platelet count). Ascertainment of drug ingestion (especially aspirin, phenylbutazone, or indomethacin) is important because some medications result in qualitative platelet deficiencies, leading to serious clotting abnormalities. Aspirin ingestion should be avoided for at least 1 week before elective hospital admission. Some medical conditions are also associated with platelet dysfunction, such as uremia, leukemia, and hepatic failure. The template bleeding time is a useful screening test for patients with suspected platelet dysfunction. It is generally agreed that a count of 50,000/mm³ functionally active platelets is sufficient for major operative procedures.

After major elective operations, wound infection rates have varied from 0.5 percent to 20 percent depending on whether the procedure was “clean” or “contaminated” (Table 9-22). Besides the added morbidity and potential mortality, surgical infection increases the average duration of hospitalization by at least 10 days, resulting in a large increase in hospital costs. Resistance to infection (immune function), dose and virulence of the invading bacteria, and the presence of foreign body, dead tissue, or anaerobic conditions are all factors that determine the probability of wound infection.

Antibiotic prophylaxis is used to reduce the incidence of postoperative wound infection in patients undergoing contaminated operations. Chemoprophylaxis is useful to cancer patients who are undergoing head and neck surgery, chest surgery, abdominal gastrointestinal surgery, and pelvic surgery, such as hysterectomy.

The use of antibiotics in these situations should be: (1) timed to allow adequate wound antibiotic levels before contamination; (2) administered only for a short time during the perioperative period; (3) specific for the most likely infecting organism(s); and (4) safe, i.e., it should present minimal additional hazard to the patient.

Cancer Operations

Local Resection

Wide local resection that removes an adequate margin of normal tissue with the tumor mass may be adequate for certain low-grade neoplasms that do not metastasize to regional nodes or widely infiltrate adjacent tissues. Basal cell carcinomas, thin melanomas (i.e., less than 1.0 mm), and mixed tumors of the parotid gland are examples of such neoplasms. Some normal tissue surrounding the tumor should be excised to prevent local recurrence.

Radical Local Resection

Neoplasms, such as soft tissue sarcomas and esophageal and gastric carcinomas, may spread widely by infiltration into adjacent tissues. In such cases it is necessary to remove a wide margin of normal tissue with the neoplasm. The wide margin of normal tissue between the line of excision and the tumor mass also acts as a protective barrier against tumor cell spill into the severed lymph and blood vessels. The greater the width of normal tissues between the plane of dissection and the tumor, the greater the likelihood of a complete local excision.

If the tumor was previously explored, but not removed, or if an incisional biopsy procedure was performed, a segment of skin and the underlying muscles, fat, and fascia must be removed far beyond the limits of the original incision because tumor cells may have been implanted in the incision during the initial operation.

Malignant neoplasms are not well encapsulated. A pseudocapsule composed of a compression zone of neoplastic cells may surround the tumor. This apparent encapsulation offers a great temptation for simple enucleation, because the tumor may be easily dislodged from its bed. The surgeon must cut through normal tissue at all times and should never disrupt the neoplasm during its removal. Dissection should proceed with meticulous care to avoid tumor cell spill. The surgeon should resect as far as possible from the gross extent of the tumor on all sides, including the deep aspect. Skin, subcutaneous fat, and some muscles may have to be sacrificed, but usually this causes little functional loss. Sacrifice of tumor- involved major vessels, nerves, joints, or bones may be necessary to obtain a curative result. The extent of operation is also determined by functional integrity.

Radical Resection with En Bloc Excision of Lymphatics

Since many neoplasms commonly metastasize by way of the lymphatics, operations have been designed to remove the primary neoplasm and the regional lymph nodes draining that area in continuity with all the intervening tissues. Conditions are best for this type of operation when the collecting nodes of the lymphatic channels draining the neoplasm lie adjacent to the primary site or when there is a single avenue of lymphatic drainage that can be removed without sacrificing vital structures. It is important to avoid cutting across involved lymphatic channels because such action can increase the possibility of local disease recurrence.

En bloc regional lymph node dissections such as those undertaken in modified radical mastectomy and radical total gastrectomy should be performed when there is clinical involvement of lymph nodes by metastatic tumor. In many cases the tumor has already spread beyond the regional nodes, and the likelihood of cure after such procedures may be quite low. En bloc removal of the involved nodes offers the only chance for cure and provides significant palliation and local control.

Lymphadenectomy

There are some general principles common to lymph node dissection (LND) at various anatomic sites, particularly the neck, axilla, and groin (Fig. 9-27).

- (1) The surgeon must thoroughly understand the anatomy of the lymph nodes in each area of the body and should incorporate all the draining nodes at risk into the surgical specimen.
- (2) The goals of LND must be clearly defined as cure, local disease control (i.e., palliation), and/or staging. LND may be curative in some patients with bulky nodal metastases, or palliative in certain patients with large, symptomatic nodal metastases in the presence of distant metastases. Staging to determine whether metastatic disease is present in the regional lymph nodes will assume increased importance with the availability of effective adjuvant therapies.
- (3) An incomplete LND is generally not acceptable except when the goals of surgery are strictly palliative. When the intent is curative, as in a patient with regional metastases from a primary cutaneous melanoma, complete LND is vital. But a partial axillary dissection (levels I and II) is satisfactory for breast cancer, because the goals are staging and local disease control.
- (4) The incision providing access to the underlying regional nodes should be placed to minimize the risk of dividing lymphatic vessels that could contain malignant cells.
- (5) For dissections of lymph nodes in the neck, axilla, or groin, closed- suction catheter drains are important to evacuate blood and serum and keep the tissues in apposition, minimizing the risk of seroma formation. One or two large catheters may be inserted percutaneously through the lower flap and placed in a dependent part of the wound. Although there are no objective data on the best time to remove these devices, the incidence of seroma formation increases if they are removed too early, and the incidence of wound infection increases if they are kept in too long (especially longer than 10 days). Catheters probably should be removed when the drainage is less than 40 mL per 24 h, or 8 days after surgery, whichever comes first.
- (6) The use of prophylactic antibiotics is controversial. The incidence of wound infection is highest after a groin dissection and lowest after a neck dissection. Staphylococci and streptococci are the most common organisms isolated from these wound infections. The use of prophylactic antibiotics is probably justified on an empirical basis for groin and axillary lymphadenectomies, but this must be confirmed with controlled clinical trials.
- (7) In patients with bulky nodal metastases, especially with invasion and fixation into surrounding tissues, such as in the axilla for metastatic breast cancer and in the neck for metastatic melanoma, the risk of regional recurrence is greater than 20 percent. Data from prospective clinical trials indicate that adjuvant radiotherapy might be considered to improve regional disease control.

Elective Lymph Node Dissection (LND)

The high rate of local cancer recurrence after surgical resection when lymph nodes are grossly involved and the high error rate when palpation is used to assess the extent of the involvement have led to elective or prophylactic dissection of clinically normal regional lymph nodes close to the primary tumor. Microscopic examination of these excised lymph nodes reveals evidence of tumor spread in 20 to 40 percent of carcinomas and melanomas. A higher 5-year survival rate was reported for patients undergoing excision of lymph nodes with microscopic rather than clinically evident tumor involvement.

This concept of elective LND has been challenged because it is not clear whether cure rates are improved if the nodes are removed before they are palpable. Controlled clinical trials directed toward this question in many types of neoplasms are currently under way. Regardless of direct therapeutic benefit, foreknowledge of tumor in regional nodes can affect staging, treatment, and prognosis. For example, patients with breast cancer who have metastases to regional nodes may benefit considerably from adjuvant chemotherapy or hormonal therapy. Also some patients with deep melanomas (> 4 mm) may become candidates for investigational adjuvant trials only if lymph node metastases are present. Furthermore, a comparison of experimental results from different institutions depends on accurate staging when therapy is initiated.

Selective Lymph Node Dissection

Morton and colleagues recently described a promising technique for detection of the regional draining lymph nodes most likely to contain metastatic tumor cells spreading from a primary cutaneous melanoma. Their technique of intraoperative lymphatic mapping and selective sentinel LND is currently under investigation in a phase III multicenter trial for melanoma and is also being applied to breast carcinoma and other neoplasms. Initially the technique relied on injection of a vital blue dye at the tumor site and visual tracking of this dye along the lymphatics to the nodal basin. Sentinel node mapping has been facilitated by adding a radiolabeled isotope to the dye and monitoring its path by a handheld gamma probe.

Extensive Surgical Procedures

Some slow-growing primary tumors can reach enormous size and may locally infiltrate widely without the development of distant metastases. Radical operative procedures can be undertaken for these extensive and nearly inoperable tumors, with occasional cure. Advances in surgical techniques, anesthesia, physiologic monitoring, and supportive care (blood transfusions, antibiotics, and fluid and electrolyte management) have permitted the development of more radical and extensive operative procedures that have significantly improved the local control rates for certain neoplasms. These extensive surgical procedures sometimes offer a chance for a cure that is not possible by other means and are

justified in selected situations when extensive laboratory workup shows no evidence of distant metastases. These operations should be undertaken only by experienced surgeons who can select those patients most likely to benefit. The surgeon considering extensive procedures such as pelvic exenteration, hemipelvectomy, forequarter amputation, or radical operations for head and neck carcinomas must be willing to accept the responsibility for the patient's postoperative emotional rehabilitation.

Pelvic exenteration is a well-conceived operation capable of curing patients with radiation-treated recurrent cancer of the cervix and certain well- differentiated and locally extensive adenocarcinomas of the rectum. This operation removes the pelvic organs (bladder, uterus, and rectum) and all soft tissues within the pelvis. Bowel function is restored with colostomy. Urinary tract drainage is established by anastomosis of ureters into a segment of bowel (ileum or sigmoid colon). The 5-year relapse-free survival after pelvic exenteration is 25 percent.

Hemipelvectomy (resection of the lower extremity and iliac bone) can sometimes be curative for skeletal sarcomas limited to the head of the femur or acetabulum or to one-half of the pelvic structures and for some slow-growing soft tissue sarcomas of the upper thigh and buttock that recur locally but metastasize slowly. Forequarter amputation (resection of the upper extremity and scapula) can offer similar results when the neoplasm is limited to the bones of the scapula and upper humerus or to the soft tissues of the shoulder girdle.

Surgery of Recurrent Cancer

Surgical resection of localized, recurrent, low-grade, slow-growing malignancies may produce a long period of remission. Surgical procedures are frequently successful in controlling recurrent soft tissue sarcomas, anastomotic recurrences of colon cancer, certain basal and squamous cell carcinomas of skin, and breast cancer recurring after lumpectomy. However, surgical resection of the recurrent neoplasm in the patient with metastatic disease is usually unsuccessful and rarely indicated, unless the entire tumor mass can be completely removed.

The results of routine second-look operations to detect early recurrence of colon cancer have not been impressive, but tumor markers, such as *CEA*, have been extremely useful for selecting patients most likely to benefit from reoperation. In general, a local recurrence can be treated surgically or with radiation. The surgeon must decide which form of treatment will achieve local control with the lowest morbidity.

Specific Organs Breast

In the United States breast cancer is the most common malignancy in women and second only to lung cancer in cancer-related mortality in women. In 1995 more than 180,000 women were diagnosed with invasive breast cancer, and the probability of a woman's developing breast cancer at some point during her life was estimated at 1 in 9.

The increased incidence of breast cancer has increased awareness of the disease. Screening mammography is used more commonly, because clear benefit has been shown in large prospective studies of women over the age of 50 years. The benefit has occurred in diagnosing more malignancies at in situ or early stages compared with women who are self-diagnosed and those outside of screening programs. Women should be screened for breast cancer with a baseline mammogram between the ages of 30 and 40 years, a mammogram every two years between ages of 40 and 50 years, and yearly thereafter. Education regarding breast self-examination and timed physical examinations by a physician should also be done.

In “early” cases, a woman presents to the physician with an “abnormal” mammogram showing microcalcifications, asymmetry, a mass with or without skin thickening, or other signs of potential malignancy. Alternatively, the patient may present with a palpable breast mass, nipple discharge, or signs of more advanced disease such as skin retraction or inflammatory skin changes. Physical examination should be carried out to look for local regional metastases.

A diagnostic algorithm may be developed, determined by the primary site and size of the suspicious lesion. Palpable lesions initially may be diagnosed by fine-needle aspiration in the outpatient setting. If this cytologic examination is positive, there should be a thorough discussion with the patient of treatment options and definitive therapy initiated. In cases in which a breast-conserving procedure can be done, the primary lesion is excised, with pathologic confirmation of the adequacy of tumor margins, and tumor tissue is prepared for estrogen and progesterone receptors. A limited axillary dissection is performed at that time, with the entire procedure done under general anesthesia. In the case of an equivocal or nondiagnostic fine-needle aspiration, the breast mass can be excised under local anesthesia with a margin of normal breast tissue (frozen section is done for diagnosis and tissue is processed for hormone receptors). The axillary dissection can be done at a subsequent date. If the lesion is not palpable but is mammographically defined, a needle localized biopsy or stereotactic core needle biopsy is done for diagnosis.

The responsibility of the surgeon is to provide the patient with the best chance for cure, for local chest wall control of tumor, and for minimal morbidity. If axillary lymph nodes are not clinically palpable, dissection of the lateral axilla (level I and II lymph nodes) is performed, with the lateral border of the pectoralis minor muscle as the medial limit of the dissection.

The management of these patients has evolved toward breast conservation surgery (lumpectomy, axillary dissection, and breast irradiation). This coincides with the trend toward smaller cancers associated with greater public awareness and increased use of breast screening using physical examination and mammography. In patients with larger primary tumors (>4–5 cm) or those who have high-risk features that might compromise local control, a modified radical mastectomy with or without primary reconstruction can be done. In the past, there was greater emphasis on irradiation of lymph node regions (internal mammary nodes, supraclavicular nodes), but the long-term survival benefit of adjuvant radiation therapy has not been shown.

Other approaches have been used to treat patients with small (<2 cm) cancers of the breast. Veronesi and associates (1981) compared radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in 701 patients with breast cancers measuring less than 2 cm in diameter. Local chest wall or ipsilateral breast recurrences and patient actuarial survival were similar in both groups.

In the National Surgical Adjuvant Breast and Bowel Project (NSABP) trials, Fisher and colleagues (1989) reported that among patients with negative nodes, 12 percent of those who underwent irradiation and 37 percent of those who did not undergo irradiation have a recurrence of breast tumor during the 8 years of follow-up ($p < .001$). The probability of a recurrence in those with positive nodes (all of whom received chemotherapy) was only 6 percent with irradiation and 43 percent with no irradiation. Irradiation of the breast obviously is important to decrease local recurrence and the need for subsequent mastectomy.

In situ intraductal or lobular breast cancer is seldom associated with lymph node metastases. In situ ductal breast cancer may be treated by local excision with or without radiation therapy or by total mastectomy with immediate reconstruction. The choice of therapy is often dictated by the extent of the in situ disease, the presence or absence of multifocal lesions, the histologic type, the patient's desires regarding treatment, cosmesis, and future risk.

Sixty percent of recurrent breast cancer occurs within two years after mastectomy. If the initial local treatment for breast cancer is adequate, local chest wall recurrence rates should be less than 5 percent in stage I disease and less than 10 percent in stage II disease. Breast reconstruction can be performed simultaneously with mastectomy in women with invasive breast cancer. The major criticism of breast reconstruction is the potential for delay in diagnosing local chest wall recurrence, but this risk is small. The best defined role for immediate breast reconstruction may be in the treatment of in situ breast cancer.

Esophagus and Stomach

Operative therapy for carcinomas of the esophagus has been associated with substantial morbidity attributed to patient factors, such as preexisting cardiopulmonary disorders and malnutrition, and to anatomic factors, such as the location of the primary tumor and the local, regional, and lymphatic spread.

In addition to intensive preoperative nutritional and pulmonary support to reduce postoperative complications, adjunctive radiotherapy and/or chemotherapy have been administered in an attempt to reduce local recurrence and increase patient survival. Randomized trials with preoperative irradiation alone have not shown survival benefit.

Preoperative combination chemotherapy, consisting of cisplatin, bleomycin, and vindesine, results in tumor shrinkage in nearly half of all patients. In addition, the surgical resectability rate is increased in patients given preoperative chemotherapy and radiation treatment.

In the event that resection for cure is impossible, resection of the tumor for palliation should be performed. If a patient can ingest adequate nutrients by mouth, the quality of remaining life is improved and the opportunity for further palliation from chemotherapy and radiation therapy is increased. In patients with large tumors, surgical palliation may include bypass by colon interposition. The risk of producing a major postoperative complication by doing a palliative resection must be weighed against the potential benefits. The operating surgeon must consider life expectancy and quality of remaining life in deciding between palliative resection or bypass and nonoperative treatment.

The surgical treatment of gastric cancer is based mainly on operative and pathologic observations. Carcinoma of the distal portion of the stomach (restricted to pylorus and antrum) should be treated by subtotal gastrectomy, omentectomy, and removal of subpyloric, hepatic arterial, and preaortic lymph nodes. Tumors located in the body of the stomach have been treated by subtotal or total gastrectomy with the choice based on tumor location; little difference in overall survival rate is noted between these two procedures.

Pancreas

If evidence of distant or regional metastases from carcinoma of the pancreas or periampullary region cannot be identified, the patient in good medical condition is a candidate for pancreaticoduodenectomy (see Chap. 30). The malnourished patient should receive either enteral or parenteral nutritional support during preoperative preparation, unless deteriorating liver function secondary to biliary obstruction necessitates immediate biliary decompression. Elevated levels of serum bilirubin and alkaline phosphatase in a well-nourished patient with normal hepatocellular function are not by themselves contraindications to surgery.

Liver

Advances in anesthesia, blood replacement, and surgical technique have made major hepatic resection feasible with low mortality. Patients with primary hepatoma and certain solitary metastatic cancers, e.g., from carcinoma of the colon, should be evaluated for possible curative resection.

Hepatic ultrasound or abdominal CT scanning combined with plasma *AFP* and *CEA* levels are the most effective tests for screening patients for liver disease, but their sensitivity is limited. Arteriography outlines the arterial blood supply to the liver. CT and intraoperative ultrasonography are sensitive tests that identify small metastatic deposits and can detect tumor in the lateral segment of the left hepatic lobe.

At exploration, metastases limited to one hepatic lobe should be considered for resection either by wedge excision or lobectomy. If the cancer is unresectable, hepatic arterial devascularization and/or chemotherapy infusion may be performed.

Large Bowel

Preoperative preparation of the patient with colorectal carcinoma requires thorough assessment by physical examination, radiographic methods, and endoscopic techniques to determine the size, mobility, and histology of the primary tumor. Synchronous cancers of the large bowel are noted in approximately 4 percent of patients. In the absence of obstruction, all patients should undergo a thorough mechanical bowel preparation with laxatives and enemas preoperatively. Oral neomycin and erythromycin is one antibiotic prophylactic regimen that is used along with mechanical bowel preparation on the day before operation to reduce intraluminal bacterial content.

In colonic resection, the mesentery of the cancer-bearing bowel should be removed as completely as possible to avoid local recurrence and obtain adequate lymph node removal for staging and attempted surgical cure.

Good surgical judgment is necessary when operating on any part of the colon, but it is required most when dealing with carcinoma of the rectum and rectosigmoid. The surgeon should provide a combination of the best opportunity for cure and local control of disease with the least chance of postoperative morbidity, mortality, and loss of function. Sphincter-saving procedures should be performed if the chance for local control is not jeopardized. Cancerous lesions within 2 cm of the anal verge that invade through the muscularis propria (T3) usually should be removed by abdominoperineal resection. A tumor cephalad to this level often can be managed with a low anterior resection or pull-through procedure. Preoperative radiation therapy with chemotherapy may be helpful to permit tumor shrinkage, destroy peripheral tumor micrometastases, and allow a sphincter-sparing procedure.

Local excision can be done for *early* cancer of the rectum in lesions that undergo complete excision with low-grade histology and with tumor confined to the submucosal plane of the bowel wall without lymphatic or vascular invasion. This approach is preferable to electrosurgical diathermy, which destroys histologic markings and makes staging impossible. Electrocoagulation is useful as palliation for selected patients. Adjunctive radiation and chemotherapy treatment may be indicated because surgical therapy alone yields high recurrence rates in Duke's C lesions with high- grade histology, size greater than 4 cm, and distance within 5 cm of the dentate line compared with other rectal tumors.

Preoperative radiotherapy and chemotherapy combinations may increase resectability, prevent seeding, and destroy cancer cells outside the operative field. Postoperative radiotherapy “spares” those patients with more favorable pathologic lesions and can be given to a surgically “defined” area, but complications such as small bowel radiation enteritis are more frequent.

Melanoma

Treatment of malignant melanoma is based on depth of skin invasion, location on the body, and areas of potential metastatic spread. Punch biopsy may be satisfactory for histologic diagnosis, but excisional biopsy, including a small amount of subcutaneous fat, is preferable in order for the pathologist to define accurately the depth of invasion and tumor thickness. Pathologic staging should be done on permanent sections, because interpretation of frozen-section preparations can be misleading.

Thin, radial growth phase melanoma that has not penetrated through the papillary dermis (Clark's levels I and II) or is less than 0.76 mm thick seldom metastasizes or recurs locally and has a 95 percent chance of initial cure by adequate local excision. Adequate local excision usually can be achieved by wide excision of the biopsy scar with 1 cm margins sufficient to allow primary closure of the wound. Some lesions, particularly the superficial spreading type, may be too large for excision and primary closure, and flap rotation or skin grafting may be necessary. Similarly, skin grafts may be needed in areas of the body where skin is not easily mobilized, e.g., the sole of the foot.

Thick vertical growth phase lesions that have penetrated to the reticular dermis (level III), into the reticular dermis (level IV), or through it and into the subcutaneous fat (level V) have a much higher potential for metastasis. This risk is directly proportional to the depth of invasion (especially when greater than 1.5 mm in thickness) and to the location on the body (the risk in head and neck tumors is greater than that in trunk tumors, which is greater than that in tumors of the extremities). All regional node-bearing areas and the area between the primary melanoma and the regional lymph nodes should be carefully evaluated for clinical regional lymph node metastases or in-transit metastases coursing to the regional lymph nodes through the intradermal or subdermal lymphatics.

Surgical treatment of all thick (1.5 to 2 mm) melanomas must be directed to eradication of the primary tumor, areas of potential in-transit metastases, and the regional lymph node-bearing areas. Treatment of the primary lesion should include a wide excision with 1 cm margins. Treatment of regional lymph nodes also must be practical. A thick melanoma in the middle of the trunk may potentially metastasize to the axillary, inguinal, or supraclavicular lymph node areas on either side; prophylactic lymph node dissection of each of these locations would be impractical.

Adult Soft Tissue

Soft tissue sarcomas in adults may grow to a large size without detection, as they are often deeply situated. Prognosis has been shown to be correlated with tumor size, site, histologic type, and degree of differentiation. Tumors that are small (less than 5 cm), superficial (not extending beyond the superficial fascia), and low grade are grouped together as tumors with a favorable prognosis. Tumors that are large (more than 5 cm), deep (extending beyond the superficial fascia), and high grade have a poorer prognosis.

Traditionally, soft tissue sarcomas have been treated surgically (Fig. 9-28). Experience has shown that excision of only the tumor results in a local recurrence rate of 90 percent because of the “pseudocapsule,” an outer sheath of viable tumor cells that is stripped away and left behind after simple excision. Local recurrence rates are around 40 percent after more extensive resection and 10 to 25 percent after radical muscle group soft- part resection or amputation.

In a prospective randomized trial comparing amputation with wide local excision plus radiotherapy, local recurrences occurred in none of 16 patients treated with amputation, and 4 of 26 undergoing combined treatment. Eilber and associates reported a 3 percent local recurrence rate for extremity sarcomas treated with preoperative chemotherapy and radiotherapy, followed by en bloc resection of primary tumor with limb salvage.

Modifications of the protocol (e.g., reduction of radiation dose to a total of 2800–3000 cGy) are associated with excellent tumor control and reduced wound toxicity. Adjunctive radiation therapy, with or without chemotherapy, decreases local recurrence after a conservative surgical approach that *completely* removes the tumor with adequate surgical margins. The surgeon and radiation therapist ideally should plan their approach together. All areas of the operative site should be irradiated to minimize local recurrence.

RADIATION THERAPY

The effectiveness of radiation therapy, like surgical therapy, must be assessed by comparing local and regional tumor control with treatment- induced morbidity. Ionizing radiation is effective in the management of a wide variety of malignant tumors and is part of the treatment for 50 to 60 percent of patients with cancer. The radiation oncologist should be involved in the selection of patients and their evaluation before, during, and after treatment.

With radiation, tumors can be destroyed while anatomy is preserved. Often function and cosmesis can be preserved if the anatomy is intact before treatment. Concurrent medical problems have less influence on radiation therapy than on surgical or chemotherapy, although treatment-related sequelae may be more frequent or more severe in patients with certain systemic illnesses, such as diabetes or collagen vascular disease.

The differential effect of radiation on tumors and normal tissues results in a favorable therapeutic ratio in most clinical situations. Radiation can, however, have immediate and delayed side effects on normal tissues. The incidence and severity of late sequelae, which may progress over many years, are highly dependent on treatment technique. The type of equipment used, field arrangement, accuracy of tumor localization and field placement, treatment schedule, and expertise of the treating physician all influence outcome. The appearance of late sequelae may be the unfortunate consequence of treatment techniques long abandoned.

Physical Basis

Ionizing radiations are characterized by their capacity to ionize (and excite) atoms and molecules in an absorber such as tissue. Electromagnetic radiations can be produced artificially in kilovoltage radiation therapy units and linear accelerators by impinging energetic electrons on a target (e.g., tungsten). The energy of the resulting x-rays is related to the energy of the accelerated electrons as they reach the target material. Electromagnetic radiations of characteristic energy (gamma rays) are also produced by the radioactive decay of naturally occurring radioisotopes (e.g., radium 226) or artificially produced isotopes (e.g., cobalt 60). According to quantum physics, x-rays and gamma rays can also be represented as particles called *photons*. Removal of the target in a linear accelerator results in an emitted beam of high-energy electrons that have different absorption characteristics from photons and are very useful in the treatment of relatively superficial malignancies. Other types of particulate radiations (e.g., protons, neutrons, pi mesons, and helium ions) are produced by very powerful linear accelerators, or *cyclotrons*, and have been used therapeutically, primarily in investigative settings. Because the basic physical mechanisms of action of all ionizing radiations are the same, the different effects observed with equal physical doses result from differences in spatial or temporal distributions.

For decades, doses at the point of interest, such as the tumor or spinal cord, were grossly extrapolated from skin reactions or doses measured in air (in roentgens). Today clinical specification of radiation doses is derived from direct measurements of absorbed doses within the patient (using thermoluminescent dosimeters) or from doses calculated within a tissue phantom that simulates the human being. Phantom measurements are adapted for precise clinical application through the use of computer programs. Recent technological advances have made it possible to correct for tissue inhomogeneities (air cavities and bone) within the treatment volume using CT-based treatment planning.

Until 1980, absorbed radiation doses were quantified in rads, with 1 rad equal to 0.01 joule per kilogram of the absorber. According to the 1980 recommendations of the International Commission on Radiological Units, doses should be quantified in gray (Gy) units, with 1 Gy = 100 rad = 1 joule per kilogram of the absorber, and 1 cGy = 1 rad.

The availability of modern radiation therapy was initially facilitated by widespread distribution of cobalt 60 teletherapy units, dating to 1949, and the later versatile 4 to 6 MeV linear accelerators. Linear accelerators that produce a range of photon and electron energies between 4 and 25 MeV are available and are useful in a wide variety of clinical situations. Modern equipment provides deeply penetrating beams, short treatment times, isocentric patient setups, improved skin sparing, decreased bone absorption (compared with kilovoltage units), and sharp beam margins with less side scatter than the cobalt 60 units.

In some clinical situations, brachytherapy, or the direct placement of radioactive sources within tissue, may permit delivery of tumor doses higher than those achievable with external beam radiation therapy. Because the dose delivered falls off in a manner proportional to the square of the distance from the source, very high doses can be delivered to tissues immediately adjacent to the implant with relative sparing of surrounding normal tissues. A variety of artificially produced radioisotopes, including iridium 192, cesium 137, and iodine 125, are used in the treatment of cancers of the breast, prostate, brain, lung, head and neck, gynecologic organs, soft tissues, and eye. High-dose-rate (HDR) remote afterloaded brachytherapy, a new delivery technique that is gaining greater clinical acceptance, involves the delivery of several grays in minutes. Low-dose- rate (LDR) brachytherapy may require several days of hospitalization, but HDR can be delivered in a fractionated manner as an outpatient procedure. In HDR remote afterloading, a high-activity source is driven to a predetermined series of positions for specific periods. Because of the small source size, smaller-diameter catheters can be applied to interstitial and intraluminal sites, such as

the bronchus, esophagus, and bile duct, which previously could not be easily treated with LDR techniques. This computer- operated remote afterloading technique optimizes the dose distribution. These brachytherapy techniques have led the radiation oncologist into the operating room and into closer cooperation with the surgeon.

Biologic Basis

Radiosensitivity is the susceptibility of cells to injury by ionizing radiation. This injury may cause reproductive cell death by interrupting the cell's capacity to replicate indefinitely. Radiation can kill cells by interfering with critical cell functions unassociated with cell replication (“interphase death”); this is an important mechanism of radiation-induced cell death in lymphocytes and in normal salivary gland. The inherent radiosensitivity of most normal and transformed mammalian cells is remarkably similar, with doses of 110 to 240 cGy consistently reducing reproductive cell survival to 37 percent ($D_{0.37}$ dose). Therefore, differences in the rapidity and completeness of response of human tumors and normal tissues must be based on other factors, such as the capacity to repair sublethal damage, tissue oxygenation, cell cycle time and distribution, and repopulation.

Radiocurability is the ability of radiation to control a tumor permanently, allowing survival of the host. Tumor type, size, site, and extent have a greater influence on radiocurability than cellular radiosensitivity. *Radioresponsiveness*, or the rapidity of a tumor's response to radiation, may not correlate well with radiocurability. Epidermoid carcinomas of the oral cavity, larynx, skin, and cervix and adenocarcinomas of the breast, cervix, and prostate may be radiocurable despite relatively slow responses to radiation, and undifferentiated carcinomas often respond rapidly to radiation treatment but usually are not cured because of widespread tumor dissemination. However, differences in radioresponsiveness of tumors of a specific type and site may correlate with local tumor control.

Normal tissues have a greater capacity to repair injury than do tumor cells. Fractionation, or the division of a radiation dose into multiple smaller doses, allows recovery of this damage between radiation fractions. Laboratory and clinical studies indicate that an interfraction interval of at least 4 to 6 hours is necessary to permit maximal repair of sublethal injury. Because of their greater repair capacity, slowly dividing normal tissues usually are spared more than tumor cells by the use of relatively small fraction sizes, but rapidly dividing stem cell populations, such as bone marrow and mucosal surfaces, have less capacity for repair. The use of small doses per fraction improves the therapeutic ratio by reducing the late normal tissue effects of radiation more than the tumor effects. Although some protraction of treatment is necessary to permit repopulation of acutely responding normal stem cell populations and avoid intolerable acute normal tissue effects, treatment protraction also permits repopulation of tumor cells and should not be excessive. Empirical evidence suggests that daily fractions of approximately 2 Gy optimize the balance between these effects. Clinicians have been investigating the use of altered, hyperfractionated schedules that use two or three small fractions per day in an effort to further decrease late normal tissue complications without increasing the overall duration of treatment.

For similar reasons, cell killing can be modified by changes of the dose rate. To achieve tolerable individual treatment durations, external beam radiation therapy usually is delivered at a dose rate of 2 to 5 Gy/min. As the dose rate decreases, cell killing per unit of dose decreases. Low dose rates (i.e., less than 10 cGy/min) may favor repair in normal tissues. This effect is exploited clinically to improve normal tissue tolerance (particularly of the lung) in total-body irradiation and to allow delivery of very large total doses with low-dose-rate interstitial and intracavitary brachytherapy techniques.

Radiation-induced cell killing can be modified in other ways. Because molecular oxygen must be present for maximal cell killing by ionizing radiation, tumor cellular hypoxia can decrease the effectiveness of radiation therapy by as much as a factor of 3. This “oxygen effect” may explain the postirradiation persistence of tumor cells when there is necrosis or fibrosis.

The intrinsic radiosensitivity of cells can be increased by altering the target DNA, such as by replacing thymidine with halogenated pyrimidine analogues (BUdR, IUdR) during cell replication. Cell killing can be increased by inhibiting postirradiation repair processes. For example, the repair of DNA strand breaks can be inhibited by actinomycin D and doxorubicin and by heat (42 to 45°C). Unfortunately, current methods of altering the target DNA and inhibiting postirradiation repair are not selective for tumor cells and may not improve the therapeutic ratio.

Clinical Basis

Table 9-23 is a partial list of tumors that are treated with radiation. Irradiation may be the only anticipated treatment or may be combined with surgery and/or chemotherapy. The intent of treatment may be curative or palliative. In some clinical situations, surgery and radiation provide comparable rates of local tumor control with qualitatively different side effects. In these cases, the choice of treatment only can be made by a patient who has been carefully informed by a surgeon and radiation oncologist. Whenever curative treatment of malignancy is being considered, it is important that all the specialists (medical, surgical, and radiation oncologists) are involved before initiation of therapy. Previous surgical and chemotherapeutic management can significantly influence the therapeutic ratio of subsequent radiation therapy. Close cooperation from the beginning of therapy can often improve treatment outcome significantly. For example, careful marking of the margins of a tumor during surgery can help the radiation oncologist define a more accurate target volume and decrease the morbidity of therapy. Awkward placement of a surgical incision can dramatically increase the volume, complexity, and morbidity of subsequent irradiation. In some cases, the radiation oncologist can obtain valuable information by observing the operative field.

Decisions about the utility of radiation therapy are based on tumor-related factors (type, site, extent, typical natural history) and host factors (general condition, status of local and regional tissues). In almost all cases, definitive tissue diagnosis should be obtained before treatment to avoid inappropriately morbid treatment of benign conditions that mimic malignancy. On rare occasions when biopsy poses an unreasonable risk to the patient (e.g., tumors of the brainstem and optic tract), treatment may be initiated on the basis of strong radiologic diagnostic evidence. The introduction of CT and MRI should reduce diagnostic errors in these situations.

Histologic tumor type and grade may be useful pretherapeutic predictors of biologic behavior and radiocurability. The potential for local and regional tumor control is closely related to tumor size and the primary site. In most cases, radiation dose is limited by the tolerance of surrounding normal tissues. The probability of controlling a tumor with a tolerable dose of radiation is inversely proportional to its size and the number of proliferating clonogens that must be eliminated. Surgical tumor debulking procedures that leave gross residual disease are sometimes necessary to relieve tumor-related symptoms; they usually reduce the number of clonogens by less than one log, may increase tumor hypoxia, may decrease the tolerance of adjacent normal tissues, and rarely improve radiocurability. Large tumors are more likely to have spread regionally, sometimes requiring treatment of a larger volume with a consequent reduction of normal tissue tolerance.

The primary tumor site predicts biologic behavior and dictates which normal tissues will be affected by treatment. For example, small tumors of the glottic larynx rarely spread to regional nodes, and more than 90 percent of these tumors are cured with moderate doses of radiation to a small local field. By contrast, tumors of similar size originating a few millimeters away in the supraglottic larynx have a richer lymphatic supply, are associated with a greater likelihood of regional spread, and often require treatment with relatively large fields encompassing the regional lymph nodes. Even fairly large tumors of the cervix can be controlled locally with minimal risk of serious morbidity because of the high radiation tolerance of the uterus and vagina and the ability to deliver high doses with intracavitary therapy. Carcinomas of similar size in the upper abdomen are rarely controllable with radiation therapy alone because surrounding normal tissues such as liver, kidney, bowel, and spinal cord limit the deliverable doses of external beam radiotherapy. Intraoperative radiotherapy (the delivery of external beam radiotherapy directly to a tumor exposed during an operation) is currently being investigated as a possible means of increasing the radiation dose that can be delivered in such situations.

Goals

If the cancer is curable, a prolonged treatment course and a moderate risk of serious treatment-related morbidity are often accepted in an effort to overcome life-threatening disease. If the objective is palliation of cancer- related symptoms, treatment must be designed to minimize morbidity and inconvenience and maximize symptom relief. Even when seemingly indicated, radiation therapy could be inappropriate because of host factors. Debilitated or disoriented patients may not tolerate daily treatment. Local tissue changes induced by comorbid disease may cause the risk of curative treatment to outweigh the chance of benefit.

Dosage and Delivery

Treatment planning and delivery require close cooperation among health professionals, including medical physicists, dosimetrists, radiation therapy technologists, radiation therapy nurses, and radiation oncologists. The radiation oncologist begins by defining a target volume based on the patient's medical history, physical examination, radiologic studies, operative description, and pathology reports. This determination also considers the natural history of the disease, its anatomic routes of spread, and possible interactions among radiation, surgery, chemotherapy, and intercurrent disease. The therapeutic approach is designed to maximize the chance of tumor control and minimize the risk of treatment-related morbidity. In most cases, the primary tumor and adjacent area at risk for regional spread are graphically displayed and incorporated in a planned target volume.

Radiation can be delivered by multiple beams of photons or electrons, sometimes augmented by interstitial or intracavitary applications. Beam- shaping devices may be used to alter depth dose distributions and to shape the radiation field. The dosimetry data are incorporated in computer- assisted programs that allow rapid, accurate calculations of the desired options. The chosen treatment fields are simulated on the patient using a specialized machine built to the geometric specifications of the treatment machine but fitted with a diagnostic x-ray head and fluoroscope. The patient can be fitted with immobilization devices to improve the reproducibility of treatment. Marks on the patient's skin make it possible to reproduce the simulated fields during treatment. With careful immobilization and field localization, repetitive treatment delivery can be accurate to within a few millimeters.

Different sites in the same patient can be treated with different doses according to risk and the amount of tumor involvement. A relatively large volume including the primary tumor and surrounding areas at risk for harboring microscopic disease may be treated using a moderate dose. The primary tumor, grossly involved lymph nodes, or positive surgical margins may then be “boosted” to a higher dose with smaller external beam fields or with brachytherapy.

When the goal of treatment is palliation, the treatment dose and schedule are chosen to achieve symptom relief as quickly as possible, with little or no treatment-related morbidity. Because the dose is usually not pushed to normal tissue tolerance and the patient is not expected to survive to experience late effects of radiation, relatively large daily fractions may be used to shorten the overall treatment time. Treatment is given only to relieve symptoms or occasionally to prevent imminent problems (e.g., to prevent fracture in tumorous weight-bearing bones). Occasionally, aggressive radiotherapeutic treatment of hematogenous metastases may be indicated, particularly if the lesion is solitary and presents after a long disease-free interval.

Combination Modalities

Radiation therapy alone is curative in many clinical situations. Aggressive local or locoregional treatment yields high cure rates in many types of head and neck cancer, gynecologic malignancies, anal cancer, prostate cancer, Hodgkin's disease, and other neoplasms. In other cases, radiation is used in combination with surgery or chemotherapy. Radiation and surgery may be directed to the same site, e.g., when resection of a cancer of the hypopharynx is followed by irradiation, or when irradiation of a soft tissue sarcoma in an extremity is followed by surgery. Combined modalities may decrease the morbidity associated with either modality alone. Local tumor excision plus radiation therapy is an alternative to mastectomy for breast cancer. Treatment of soft tissue sarcomas with wide local excision and preoperative or postoperative irradiation achieves local control rates comparable to amputation but with preservation of the limb. In some cases, radiation and surgical treatment are directed to different sites, e.g., when orchiectomy is followed by irradiation of the retroperitoneal lymph nodes, or when a neck dissection follows interstitial irradiation of a cancer of the oral tongue.

Postoperative radiation improves local and regional control rates in many postsurgical situations. Because patients selected for postoperative therapy tend to be those with unfavorable clinical or histologic features, retrospective studies usually do not compare similar groups of patients and do not answer important questions about the influence of postoperative treatment on survival. However, even when the survival benefit of postoperative radiation is uncertain, treatment may be indicated to prevent local recurrence.

When surgery and radiation therapy are directed to the same site, the interval between them depends on a range of factors. Moderate-dose irradiation of a soft tissue sarcoma may be followed by resection in 10 to 14 days; rectosigmoid resection should be delayed 4 to 6 weeks after pelvic irradiation to allow for regression of edema and hyperemia. Most postoperative radiation therapy is delivered with relatively high doses directed to sites at high risk for persistent tumor. The optimal timing of postoperative radiotherapy depends on the type of surgical procedure, patient recovery, wound healing, and tumor characteristics and should be decided by the surgeon and radiotherapist. Unnecessary delays may allow time for tumor regrowth and decrease the efficacy of radiotherapy.

Planned combined treatment should not be confused with the use of one method after failure of another approach that often reduces the effectiveness of the second method. For example, irradiation of a tumor regrowing in tissues altered by surgery is likely to be ineffective because of decreased vascularity and increased tumor volume. It may have a higher morbidity because of abdominal adhesions that fix segments of bowel within the high-dose volume. Conversely, an aggressive course of radiation therapy that leads to fibrosis, loss of tissue planes, and decreased vascularity increases the incidence of complications from subsequent major surgery.

Side Effects

Any effective anticancer therapy can produce undesirable and occasionally dangerous side effects. Acute radiation-induced side effects can be distressing but can usually be managed conservatively and are almost always self-limited. The nature of these effects, summarized in Table 9-24, depends on the tissues included within the target volume.

The clinically important late sequelae of radiation therapy may not be apparent until months or even years after completion of treatment. The risk of late complications can usually be minimized (but not eliminated) by careful technique. In many situations, an effort to eliminate a small-to- moderate risk of major complications by reducing the radiotherapy dose will increase the risk of tumor recurrence.

The risk of second malignancies induced by ionizing radiation is small. In studies of more than 2,000 patients with head and neck cancer and 2,000 patients with cancer of the breast, no increase in the incidence of second cancers could be demonstrated in patients treated with radiotherapy. The increased incidence of leukemia in patients treated for Hodgkin's disease was strongly correlated with exposure to alkylating agents, although many of the patients also received radiation therapy. A very slight increase in the incidence of myelogenous leukemias (observed:expected = 1.4) was noted in 29,493 patients observed for 60,000 person-years after radiation for cervical carcinoma.

MANAGEMENT OF CANCER AT DISTANT SITES

Clinical Evaluation and Screening

If an initial screening appraisal indicates the need for a more extensive metastatic survey, then the presence and extent of metastases should be assessed by a comprehensive but cost-effective diagnostic evaluation. Components of this evaluation depend on the signs and symptoms of disease in a particular area (Table 9-25) and the goals of treatment (cure versus palliation). Laboratory tests, except for screening, should not be ordered unless the results would change the treatment plan. Other factors to be considered are the cost and availability of further tests, prognostic factors, and the natural history of the disease.

These principles are also important postoperatively in determining how long and how often to follow the patient and what types of screening tests should be used. In general, the patient should be evaluated every 3 to 4 months for the first 2 years, at 6-month intervals to the fifth year, and then at least once a year indefinitely. Chest x-rays and/or laboratory tests are obtained initially at 6-month intervals and then yearly or as indicated. The exact frequency of screening depends on the risk of recurrent metastatic disease; patients with very early cancer (e.g., melanomas less than 1.0 mm) might be evaluated at 6-month to 12-month intervals from the beginning, but those at high risk for metastases should be evaluated every 2 to 3 months during the first 2 years. Because a large proportion of recurrences are detected by the patients themselves, high-risk patients should be briefed on possible symptoms and urged to seek immediate medical attention should they appear.

History and Physical Examination

A hallmark of metastatic disease is a symptom complex that progresses in intensity or frequency. A careful history and physical examination are the most sensitive, specific, and cost-effective means of evaluating possible metastatic disease, short of a biopsy.

Laboratory and Radiological Tests

The chest x-ray should be used routinely for screening. Whole-lung tomograms or CT scans of the chest are useful for evaluating suspected pulmonary, pleural, or mediastinal metastases (Fig. 9-29). Serum liver function tests, including lactic dehydrogenase level, are important screening tools for metastatic disease. An isolated elevation of the serum alkaline phosphatase or lactic dehydrogenase level is presumptive evidence of metastatic disease. A CT or ultrasound scan of the abdomen should be obtained if physical examination or abnormal liver chemistry suggests intraabdominal metastases.

Some solid tumors produce circulating tumor secretory products that represent marker molecules. These tumor markers can be used to monitor for recurrence or assess the response to treatment (see Table 9-20). Marker molecules are not infallible, however, and a tumor cell population may emerge that fails to secrete the marker being monitored.

Bone and brain scans are not indicated for routine screening of occult metastatic disease because their diagnostic yield is low, except for those cancers that frequently relapse first in the bone, such as breast cancer and prostate cancer. A radionuclide bone scan is the most sensitive test for skeletal metastatic disease, but a careful history and directed radiographs are necessary to ensure that areas of uptake do not represent areas of old trauma or inflammation.

Positron emission tomography (PET) has become a widely used investigative and clinical tool for cancer detection and diagnosis, cancer staging, and cancer treatment monitoring. This use of PET technology is based on the observation that malignant transformation of cells is associated with an increased glycolytic rate (even in the presence of oxygen), which is correlated with uptake of fluorine-18 deoxyglucose (FDG). PET scanning can detect tumors as small as 10 mm. It is useful for detecting brain tumors, metastatic melanoma, and metastatic malignancies of the head and neck, breast, lung, and colon.

Pathologic Tests

The definitive diagnosis of metastatic disease is made by biopsy analysis. An excisional or needle biopsy procedure is relatively easy when the suspected metastasis is superficial. Deeper lesions may also be approached with a fine-needle biopsy. In many circumstances, a clinical diagnosis made by radiologic studies is sufficient, especially if metastases involve more than one site at the same time and the abnormality was absent on previous studies.

Cytologic examination of urine, sputum, cerebrospinal, peritoneal, or pleural fluid, or of bone marrow also may yield a diagnosis of metastatic disease, especially with specific symptoms referable to these areas. A common problem in the diagnosis of metastatic cancer is distinguishing a metastatic lesion from a primary anaplastic, undifferentiated carcinoma or lymphoma, e.g., when an anaplastic lung lesion appears in a patient who had a primary anaplastic or undifferentiated carcinoma or lymphoma. When an anaplastic lung lesion appears in a patient who had a primary melanoma removed 5 years previously and is a heavy cigarette smoker, should this be diagnosed as a metastatic melanoma or a lung carcinoma? Electron microscopy and immunostaining with antibodies of cytologic specimens and tissue specimens can be crucial in aiding the pathologist in the diagnosis.

General Principles of Treatment

The treatment of a patient with advanced cancer depends on the number and sites of metastases, their rate of growth, types of, and responses to, previous treatment, and the patient's age, overall condition, and desires. For example, vigorous treatment might be appropriate for a slowly growing solitary metastasis, but only symptomatic treatment, or none at all, might be used in a debilitated patient with multiple metastases for whom prior treatment has failed. The option of no treatment is particularly important in patients who are asymptomatic, terminally ill, or very old. For example, a physician may choose to observe an asymptomatic patient with slowly growing tumors in sites such as the lung. Quality of life is maintained in this instance, and treatment can be instituted when the size or number of metastases increases or when the patient develops symptoms. A patient should never be denied treatment when there is a reasonable expectation of success and an acceptable risk of toxic effects.

The number of organs or tissues containing metastases is the most significant factor predicting survival in patients with distant metastases. For example, the median survival is 7 months for melanoma patients with metastasis to one site, 4 months for those with metastases to two sites, and only 2 months for those with metastatic disease at three or more sites. The location of the metastases is also important. Sites associated with relatively favorable outcomes for melanoma and breast cancer include (in approximately descending order of frequency) the skin, subcutaneous tissue, distant lymph nodes, bone, and lung. Unfavorable outcomes are associated with metastases to the liver and brain. Metastases in favorable sites are associated with long-term survival (i.e., 2 to 5 years) in a small but measurable proportion of patients. If the patient has received previous therapy, such as chemotherapy or surgery for a systemic metastasis, the survival time is likely to be shorter. Patients in a debilitated state are less likely to withstand vigorous treatments than those who are not suffering symptoms of metastatic disease.

Defining the Goals, Benefits, and Risks of Treatment

The first goal of treatment is relief of symptoms. Treatment to relieve symptoms is worthwhile, especially when the benefit of symptom relief exceeds the risk of toxic effects and morbidity. Its efficacy can be monitored by subjective and objective assessment of the symptoms caused by the metastases.

The second goal of treatment is to prolong life. This has not been achieved in most patients with advanced cancer. One exception is the surgical removal of solitary metastases from skin and subcutaneous tissues or visceral sites such as the lung or brain (see Fig. 9-27). Curative surgical excision of solitary metastases of melanoma in visceral organs can lead to 5-year survival rates exceeding 40 percent.

Table 9-26 presents the treatment modalities for patients with metastatic disease. The choice of treatment should consider potential toxicity, requirements for hospitalization, frequency of treatments, and median time necessary to attain antitumor response. The physician must understand the potential risks of each treatment as they apply to a particular patient. Patients who have a good performance status (who are able to care for themselves and have no debilitating symptoms from their disease), good cardiac and pulmonary function, and adequate white blood cell and platelet counts are usually candidates for clinical trials. Available effective methods for relief of symptomatic disease should be considered first.

Patient Counseling

One of the greatest causes of anxiety for the cancer patient is uncertainty. Patients should be counseled so that they can participate in treatment decisions, cope with their disease

realistically, and arrange their personal lives. Although some patients do not want a full presentation of the facts, some discussion is warranted because it is difficult to initiate treatment without explanation. If the physician and the patient do not communicate honestly and openly, the patient could lose trust in the physician and might not accept subsequent treatment recommendations.

When counseling patients and their families, the physician must be sympathetic, realistic, and hopeful about the results of treatment. Patients need hope, and even in grim situations treatment can achieve certain goals, if only the relief of symptoms. In other words, the physician should continue treating the patient even when it is no longer possible to treat the cancer. The approach depends on the patient's prognosis, physical condition, and emotional stability. The patient's desires should be strongly considered when choosing among alternative forms of palliative treatment.

Surgery

Curative Procedures

Once a neoplasm has metastasized to a distant site, it should no longer be curable by surgical resection, but removal of metastatic lesions in the lung, liver, or brain has occasionally produced a clinical cure (Fig. 9-30). Therefore, in selected patients with slowly growing neoplasms, curative resection of the metastatic lesions may be indicated, especially if the metastasis is solitary. Observation for several weeks or months sometimes provides relevant information about the rate of tumor growth and the possibility of metastases emerging at other sites. All patients considered for curative resection must undergo an extensive workup to rule out metastatic spread. Among the components of this workup are a complete blood count and serum chemistry panel, applicable serum tumor markers, MRI of head, CT of chest, abdomen, and pelvis, and a bone scan, if applicable. Newer whole-body imaging studies, such as *PET* scanning, may eventually replace conventional radiologic techniques. Curative resection should be attempted only if the lesions are accessible and the procedure can be performed safely.

Palliative Procedures

Surgical procedures are sometimes indicated to relieve the symptoms or reduce the severity of disease, or to prolong a useful, comfortable life without attempting cure. A palliative operation that improves quality of life by relieving pain, hemorrhage, obstruction, or infection is justified when it can be done safely without great discomfort to the patient. Surgery that only prolongs a miserable existence does not benefit the patient. Some examples of palliative surgical procedures are colostomy, enteroenterostomy, or gastrojejunostomy to relieve obstruction; chordotomy to control pain; cystectomy for infected, bleeding tumors of the bladder; amputation for painful infected tumors in the extremities; simple mastectomy for carcinoma of the breast, even in the presence of distant metastases, when the primary tumor is infected, large, ulcerated, and locally resectable; colon resection in the presence of hepatic metastases; and biliary bypass procedures. Additionally, gastrostomy or jejunostomy may be undertaken to improve nutritional intake and administration of medications. Cerebral metastases that are not resectable by conventional surgical techniques can be ablated by stereotactic radiosurgery using a “gamma knife” that delivers focused radiation to specific brain lesions.

The decision to use palliative surgery depends on the site of the disease and the duration of anticipated survival. If the patient's life is measured in weeks, the surgical ablation of a metastasis is not justified, but longer anticipated survival may render excision of gross disease worthwhile.

Radiation Therapy

Irradiation has a role in the treatment of patients with advanced cancer, particularly those with symptomatic lesions. It is used as palliative treatment for patients with bone or brain metastases and for symptomatic lesions located in the skin, subcutaneous tissues, or lymph nodes. Radiation therapy using high-energy beams relieves the pain of bone metastases, often within 1 week. Cranial (whole-brain) and spinal cord irradiation is used for central nervous system metastases. Irradiation from high-energy proton beams using a linear accelerator also can effectively treat superficially located metastases in the skin or soft tissues.

Chemotherapy

Modern use of chemotherapy began in the 1940s with the administration of hormonal therapy using androgens and estrogens and the use of the alkylating agent nitrogen mustard. Throughout the 1950s, chemotherapy was given on empirical grounds, using doses and scheduling previously found to be successful in antimicrobial therapy. Daily oral administration of fixed drugs was generally used. Although responses were occasionally seen, clinical relapse was universal. In the early 1960s, Skipper founded the principles still used in designing chemotherapeutic trials:

- (1) A single cancer cell can grow into a lethal tumor mass.
- (2) The rate of tumor growth (tumor doubling time) slows with increasing tumor burden in the later stages of tumor growth.
- (3) Most chemotherapeutic agents exhibit log cell kill kinetics, and the same increment of log cell kill is seen with subsequent doses.
- (4) Tumor burden is inversely related to curability by chemotherapeutic agents.

The comparison of antimicrobial therapy for bacterial infections and chemotherapy for malignant disease has major limitations. An immunocompetent host recognizes bacterial antigens as foreign and mounts a defense. Antimicrobial therapy will kill bacteria or arrest bacterial proliferation until host antibodies and immune effector cells expand to eliminate remaining viable bacteria. By contrast, a cancer patient's immune system may not recognize tumor cells as foreign because tumors arise from normal cells through mutational events that uncommonly lead to significant cell-surface alterations. Even after significant cell kill by chemotherapy, the host's immune system may not recognize and attack the remaining viable tumor cells, allowing their regrowth (Fig. 9-31).

Systemic or regional delivery of diffusible pharmacologic agents can destroy or arrest tumor cells capable of proliferation. Currently available drugs are not selective for tumor cells; they affect all dividing and some quiescent cells. Since most agents act on one or more stages of cell cycle, cells and tissues with the highest growth fraction will be most affected. Chemotherapy attempts maximal tumor cell kill with minimal and acceptable toxicity to normal host tissues.

The primary target of chemotherapeutic agents is the tumor stem cell. Complete destruction of all tumor stem cells is essential for cure. Pioneering studies with L1210 murine leukemia, which has a 100 percent growth fraction, fostered the concept of log tumor killing. If a single dose of drug reduces the tumor cell number by 99 percent (a 2-log kill), then subsequent identical drug doses will decrease tumor cell number by the same log decrements. If a 1 cm tumor contains approximately 10^9 cells, then even a 99.999 percent cell kill (a 5-log kill) will leave 10^4 viable tumor cells. The so-called clinical complete response, defined as disappearance of all measurable disease, usually results in recurrent disease because several million viable, but undetectable, proliferating cells may still be present in one or more sites.

Chemotherapy was initially attempted to cure patients with macrometastatic disease. Although chemotherapy is still important for gross metastatic disease and may produce dramatic clinical responses, cures are rare and limited to some pediatric malignancies, hematologic malignancies, Hodgkin's and some non-Hodgkin's lymphomas, testicular neoplasms, and choriocarcinoma. Equal attention is being directed toward preoperative chemotherapy and postoperative treatment for presumed micrometastatic disease. Most human tumors do not undergo exponential growth in vivo but demonstrate a slowing of growth with progressive tumor size. The growth fraction peaks well before clinical detection. Smaller-volume disease is presumed to be more responsive to drugs active against proliferating cells and provides the rationale for treatment in a postoperative or adjuvant setting.

Resistance to Chemotherapeutic Agents

The concept of tumor cell log kill presumes that all cell subpopulations within a tumor are chemosensitive. Clinical drug resistance is responsible for the majority of chemotherapy failures. Resistant subpopulations of cells may exist de novo within tumors or may arise by spontaneous or induced mutations. Chemoresistant cells probably do not emerge until a tumor reaches a critical mass of approximately 10,000 to 10 million cells. This may take several years for many solid tumors. Above this mass, the likelihood of one or more chemoresistant clones increases exponentially with subsequent tumor growth. It is extremely important to begin chemotherapy as early as possible.

There appear to be several different general mechanisms of drug resistance for different classes of chemotherapeutic agents (Table 9-27). Tumors are often resistant to several agents that may be structurally unrelated. This *multidrug resistance (MDR)* can also occur by several different mechanisms. An MDR gene has been identified; its protein product, P-glycoprotein, causes the efflux of several different drugs from the intracellular space of target host cells in vitro. Drugs affected by overexpression of the MDR gene include several natural products, such as the anthracycline doxorubicin, the vinca alkaloids vinblastine and vincristine, actinomycin D, and the epidophyllotoxins, such as etoposide (VP-16). High levels of P-glycoprotein are found in normal cells lining the luminal spaces of the gastrointestinal tract, liver, and kidney, where this gene product is presumed to induce an active efflux of environmental toxins from the body. The efflux pump can be blocked in vitro by calcium channel blockers, such as verapamil, which is currently undergoing clinical study. Other compounds that are effective in reversing MDR include tamoxifen, cyclosporin A, a nonimmunosuppressive analog of cyclosporin (SDZ PSC 833), and an antifungal triazole (itraconazole).

The search for potent, selective modulators of *MDR* will promote a new generation of agents. Compounds that are active in vitro and in vivo include: RS-33295-198, SR33557, a new quinoline derivative (MS-209), a novel triazolinopiperidine (S9788), and an acridonecarboxamide derivative (GF 120918). Other mechanisms of MDR include: alterations in drug-conjugating enzymes, such as glutathione S-transferase; alterations in components of the cytochrome P-450 enzyme complex, such as aryl hydrocarbon hydroxylase; or alterations in drug transport enzymes, such as glucuronyl transferase. Additionally, MDR may involve allosteric alterations in DNA through interference with activity of topoisomerase II, an enzyme responsible for maintaining DNA's secondary and tertiary structure. Molecular probes are being developed to identify additional genes involved in MDR. Antibodies against P-glycoprotein are available and under investigation for diagnostic and therapeutic use. A new marker for MDR has been described. Specific glycosylated lipids appear to be expressed in high levels in drug-resistant cells and tumors. Detection of these lipids may signal the need for aggressive chemotherapy in combination with the new MDR modulators.

Classification of Chemotherapeutic Agents

Anticancer drugs may kill tumor cells, but the majority act by preventing cell division and cell proliferation. Most drugs affect one or more components of the cell cycle. The classification of anticancer drugs as non- cell cycle-specific, cell cycle-specific, or phase-specific is relative rather than absolute. DNA synthesis can be prevented by blocking the availability of purine and pyrimidine nucleotide precursors. DNA may be damaged by cross-linking with unstable alkyl groups. DNA transcription can be prevented by direct binding of drug to DNA. Mitosis can be arrested through binding of tubulin and prevention of mitotic spindle formation. Drug combinations are often based on the complementary effects of phase-specific agents on rapidly dividing cells, and non-cell cycle-specific agents on dividing and nondividing cells.

Alkylating Agents

Alkylating agents are non-cell cycle-specific agents that contribute an unstable alkyl group to cross-link nucleic acids (primarily DNA). The major effect is on cells in G1 or mitosis, but high doses may also have cytotoxic effects on cells in G0. Alkylating agents have antitumor activity against Hodgkin's and non-Hodgkin's lymphoma, leukemias, multiple myeloma, and carcinomas of the breast, endometrium, testis, and lung. Cyclophosphamide and cisplatin (CDDP) are used in a variety of solid tumors. Dacarbazine (DTIC) and ifosfamide have documented antitumor activity against soft tissue sarcomas and other neoplasms.

Nitrosoureas are a subgroup of alkylating agents with increased lipid solubility, and better central nervous system penetration. They also may act by carbamoylation, affecting DNA and RNA. They do not exhibit cross- resistance with other alkylating agents. There is a small but important incidence of secondary malignancies (especially leukemias) several years after therapy with nitrosoureas.

Antimetabolites

These agents interfere with DNA and RNA synthesis and are phase-specific for the synthesis phase of the cell cycle. An exception is *5-fluorouracil (5-FU)*, which is phase-specific and cell cycle-specific. Antimetabolites are most active in rapidly proliferating tumors such as the hematologic malignancies but also have wide applicability in many solid tumors, especially breast cancer and gastrointestinal malignancies. These drugs may include structural analogues of metabolites essential in normal growth and proliferation that become incorporated as a false message, such as the pyrimidine analogue cytarabine (Ara-C). Other actions include reversible or irreversible binding to key or rate-limiting enzymes in the synthesis pathways, such as the relatively irreversible binding of methotrexate to the enzyme dihydrofolate reductase (DHFR). Leucovorin (folinic acid or citrovorum factor) is the most stable form of folic acid and may be given to replenish the intracellular folate pool. It can be administered after a high- dosage methotrexate regimen to “rescue” normal tissues by repleting folate and bypassing the block at DHFR. If given in combination with 5-FU, leucovorin potentiates the antitumor effect of 5-FU by stabilizing the covalent bond of 5-FdUMP (the active metabolite of 5-FU) to the enzyme thymidylate synthase.

Plant Alkaloids

Derivatives of the periwinkle plant include the vinca alkaloids vinblastine, vincristine, and vindesine. All three inhibit mitosis by binding microtubules and causing arrest in metaphase. The vinca alkaloids have antitumor activity against Hodgkin's and non-Hodgkin's lymphomas, acute leukemias, breast, testicular, head and neck, renal, and cervical carcinomas, and other solid tumors. Although these three compounds share a similar chemical structure, they have a wide spectrum of clinical activity and toxicity.

Derivatives of the mandrake plant, the epidophyllotoxins etoposide (VP-16) and teniposide (VM-26), may act through inhibition of nucleoside transport and incorporation into DNA and RNA. These drugs show phase-specific activity for cells in G2 and late S phase. Clinical activity has been seen against lymphomas, leukemias, lung, bladder, prostate, and testicular carcinomas, hepatomas, and other solid tumors.

The plant alkaloids are prominent targets for the P-glycoprotein product of the *MDR* gene, resulting in clinical drug resistance.

Antibiotics

A wide spectrum of antitumor drugs have been isolated from microorganisms. These drugs are generally considered non-cell cycle- specific agents and appear to interfere with the synthesis and/or function of nucleic acids. Doxorubicin, bleomycin, mitomycin C, and dactinomycin are examples that have wide clinical application.

Doxorubicin was derived from *Streptomyces* species and has demonstrated activity against many solid tumors, such as soft tissue sarcomas, carcinomas of the breast, lung, esophagus, stomach, liver, bladder, prostate, head and neck, testis, and endometrium, among others. The mechanism of action appears to involve intercalation into DNA with triggering of topoisomerase II-mediated DNA cleavage to alter the tertiary structure of DNA. A close analogue, daunomycin, has shown antitumor activity against acute lymphoblastic and myeloblastic leukemia.

Bleomycin was derived from a *Streptomyces* species and causes single- and double-strand breaks in DNA but not RNA. It has antitumor activity against squamous cell carcinomas, lymphomas, testicular and lung carcinomas, malignant melanoma, soft tissue sarcomas, and mycosis fungoides.

Mitomycin C is a non-cell cycle-specific antitumor antibiotic that acts as an alkylating agent to cross-link DNA and inhibit DNA and RNA synthesis. Antitumor activity has been demonstrated against many gastrointestinal malignancies and breast, lung, cervix, and bladder carcinomas.

Dactinomycin (actinomycin D) is an antitumor antibiotic derived from a *Streptomyces* species. It prevents DNA and RNA synthesis by binding to the deoxyguanosine moieties of DNA. Single-stranded DNA breaks have also been demonstrated. It has antitumor activity against malignant melanoma, testicular tumors, choriocarcinoma, Wilms' tumor, neuroblastoma, retinoblastoma, and several sarcomas.

Miscellaneous

Tamoxifen citrate is a nonsteroidal antiestrogen with cytostatic effects mediated through competitive inhibition of the estrogen receptor (ER) and several other emerging non-ER pathways. Tamoxifen has a clear role in the therapy of hormone-dependent tumors of the breast and prostate. Recent large studies have demonstrated antitumor activity against hormone receptor-negative breast carcinoma and possible protection against contralateral and recurrent breast carcinoma. Additionally, some beneficial claims have been made regarding the possible prevention of osteoporosis and atherosclerotic heart disease through an ER effect. Although the ER- mediated mechanisms of tamoxifen have become more complex and confusing, the clinical applications have expanded as a result of the apparent benefits and the relatively low toxicity of outpatient oral therapy.

Levamisole is an antihelmintic drug recently used as an antitumor agent because of reported immunorestorative activity for functionally impaired macrophages and T lymphocytes. Two large, randomized, prospective trials have shown a survival benefit in levamisole combined with *5-FU* in the adjuvant treatment of modified Duke's C adenocarcinoma of the colon. Levamisole also is being used in combination with other agents for rectal carcinoma.

Drug Selection

The National Cancer Institute (NCI) has established three phases for testing new drugs prior to general availability. Phase I clinical trials are designed to determine the maximal tolerated dose (MTD) of drug, drug toxicity, and schedule for use in phase II trials. Phase I trials usually are limited to patients whose advanced disease has not responded to standard treatments. Schema frequently include three patients at each escalating dose until toxicity is observed. Patients are informed that the likelihood for a major clinical response is small, but responses are monitored and documented.

Phase II clinical trials attempt to estimate the response rate of specific tumor types to a particular drug. Fifteen to 25 patients with a specific tumor type are initially entered. If no responses are seen, the study is usually terminated. Phase II trials may or may not randomize patients. Patients with a good performance status and a minimum of previous treatment are preferred, to allow the highest probability of showing a favorable effect. Response to the drug must be measurable to allow quantitative assessment. Phase II trials assess the activity but not the efficacy of a particular drug. A minimal response rate of 20 percent is generally required to proceed to phase III trials.

Phase III trials determine whether a drug with known activity contributes significantly to the treatment of a disease. A comparative study with new treatment versus standard treatment(s) is used to determine the efficacy of the new treatment. The standard treatment may consist of no treatment or other chemotherapeutic agents. Large numbers of patients, long follow-up, and extensive clinical and biostatistical resources usually are required. Patients in different treatment groups must be comparable in every way, including stratification for known or suspected prognostic variables, such as sex, age, and tumor stage. A randomized, controlled trial does not use historical controls. It may have several treatment arms, but each patient is randomly assigned to only one arm. As the number of treatment arms increases, so does the number of patients required for statistically valid comparisons. Patient eligibility and entry, the timing and method of randomization, and the endpoints to evaluate response and toxicity are determined before the start of the study. The end point of many phase III trials is overall survival, but other common study end points included disease-free survival, tumor response rates, and palliation of symptoms. Phase III trials often require multi-institutional participation to accrue sufficient numbers of comparable patients over a reasonable period. Intra- and interinstitutional quality control is required to reach valid conclusions.

Dose and Timing

To achieve maximal tumor cell kill, the highest tolerated dose is given over the shortest possible time. The dosage is based on the MTD derived from phase I and II studies and must be tailored to a patient's performance status, medical illness, or organ dysfunction.

Drug dosing was traditionally described in milligrams per kilogram (mg/kg). The current, and more reliable, standard of drug dosing is in terms of body surface area, described as milligrams per square meter (mg/m²). There are many nomograms to convert mg/kg to mg/m². A simple and relatively accurate conversion multiplies the mg/kg dose by 40 to yield the mg/m² dose. Alterations are often made for patients whose actual weight is markedly different from their ideal weight, e.g., obese patients or patients with large volumes of “third-spaced” fluid collections, such as pleural effusions, ascites, or edema. These alterations must consider the drug's pharmacology and volume of distribution.

The interval between doses depends on a drug's toxicity. For most chemotherapeutic agents with bone marrow toxicity, leukopenia and thrombocytopenia become evident on a complete blood count by day 9 or 10 and are most pronounced between days 14 and 18. Recovery usually begins by day 21 and is approximately 90 percent by day 28. This provides the rationale for a 28-day course or cycle of marrow-suppressive agents. Bone marrow recovery time may be delayed in onset or prolonged because of previous chemotherapy, radiation therapy to marrow-producing bone, or inherent toxicities of a few specific drugs, but an interval that is long enough to assure the safe recovery of bone marrow production may also allow recovery and repopulation of remaining tumor cells. Administering combinations of drugs can overcome the problems of single-drug toxicity and drug-resistant cell populations.

Combination Therapy and Dose Intensification

Single-agent therapy has produced cures only in choriocarcinoma and Burkitt's lymphoma. The best strategy is to treat as small a tumor as possible, as early as possible, using a combination of drugs to minimize the emergence of resistant clones during therapy. Combination chemotherapy is an attempt to provide antitumor therapy to all resistant cell populations at the earliest possible time without increasing the toxicity to normal tissues. Combination therapy usually uses only those drugs that are also effective as single agents against a particular tumor. Drugs with cell cycle and non-cell cycle specificity can be combined to target both dividing and quiescent tumor subpopulations. Nonoverlapping toxicities are essential to avoid life- threatening injury to normal tissues.

Dose intensification has been promoted to allow comparisons of the relative effectiveness of each drug used in combination therapy and to optimize drug dose delivery during a course of combination chemotherapy. Relative dose intensity (RDI) is the proportion of drug delivered per unit of time (usually 1 week) relative to an arbitrary standard dose of that drug over the same interval. For example, during a 28-day cycle of combination therapy, a cyclophosphamide standard dose of 80 mg/m²/day by continuous infusion (a dose

intensity of 560 mg/m²/week) is compared with a cyclophosphamide test regimen of 100 mg/m²/day given only on days 1 to 14 (a dose intensity of 350 mg/m²/week). The RDI of the test regimen is 350/560 = 0.63. The test dose of cyclophosphamide is higher during the first half of the 28-day course, but the overall total dose per course is less than the standard dose.

The importance of optimizing dose intensification has been documented in laboratory animal models bearing curable tumors, in which as little as a 20 percent decrease in dose intensity of an effective drug can result in a 50 percent decrease in cure rate. Comparisons of dose intensity in combination drug trials for breast, ovarian, and colon carcinomas and for lymphomas have demonstrated a strong correlation between dose intensity and response rates.

Induction, Adjuvant, and Neoadjuvant Chemotherapy

Induction chemotherapy is the use of chemotherapy as the sole form of treatment for advanced disease. The patient usually has multiple sites of metastatic disease and is not a candidate for locoregional treatment by surgery or radiation. If induction chemotherapy is unsuccessful, a change to other chemotherapeutic agents or combinations, or salvage therapy, is considered.

Adjuvant chemotherapy is the use of regional or systemic chemotherapy after locoregional tumor elimination by surgery or radiation therapy. Adjuvant therapy attempts to eliminate residual micrometastatic disease and usually is limited to patients at moderate-to-high risk for local or distant recurrence. Since there is no tumor visible at the time of treatment, response can be evaluated only by monitoring rates of recurrence, disease-free survival, and overall survival. The choice of agents is based on response rates seen with advanced disease of the same histologic type. Therapy is intensive but limited in duration; any benefit against microscopic disease is most likely to occur during the first few courses, as long as adequate cytotoxic doses are given. Few studies show any added benefit to “maintenance” adjuvant chemotherapy.

Neoadjuvant or primary chemotherapy is the use of chemotherapy as the first treatment for localized solid tumors such as breast, gastrointestinal, and pediatric carcinomas, extremity sarcomas, and localized lymphomas, among others (Fig. 9-32). It has several advantages. First, it may reduce the size of large or locally advanced tumors, allowing a safer resection that spares surrounding normal tissues, as in breast conservation surgery, anal sphincter preservation with mid-to-low rectal tumors, and limb preservation with extremity sarcomas. Second, tumor responsiveness to chemotherapy can be determined while grossly or radiologically visible tumor is still present; agents that produce an initial complete or major partial response will be continued postoperatively. A third potential advantage is immediate treatment of possible micrometastatic disease. On the other hand, unsuccessful neoadjuvant chemotherapy can delay locoregional interventions, and tumor progression during this time may preclude safe resection or require sacrifice of additional normal surrounding structures to obtain adequate resection margins. Preoperative chemotherapy may confuse pathologic staging of resected tissues, complicating future treatment decisions and prognosis.

Drug Delivery

Intravenous administration, the most common route for chemotherapeutic agents, delivers uniform doses of one or more drugs to all capillary beds and end organs with known or presumed metastatic disease. Regional arterial chemotherapy and intraperitoneal chemotherapy are other common routes of delivery. Less common applications include intravesical bladder therapy, topical therapy of skin tumors, intrathecal therapy for meningeal disease, and intrapleural drug instillation for malignant pleural effusions.

Intraperitoneal Administration

Intraperitoneal chemotherapy delivers relatively high doses of a chemotherapeutic agent to the serosal surfaces of the peritoneal cavity. The use of soft, indwelling peritoneal catheters connected to a subcutaneous access port has greatly diminished the risk of catheter-related infections. Intraperitoneal chemotherapy might be considered when peritoneal metastasis or liver metastasis is the dominant pattern of failure. It has been used in abdominal neoplasms that demonstrate abdominal carcinomatosis or malignant ascites, such as ovarian, colorectal, pancreatic, and gastric carcinomas, retroperitoneal sarcomas, pseudomyxoma peritonei, and abdominal mesotheliomas.

The depth and rate of diffusion of chemotherapeutic agents through peritoneal serosa and into peritoneal tumor depends on their size, lipid solubility, and charge. A large molecule with a low lipid content and a positive or negative charge will have a lower rate of peritoneal surface absorption and a longer exposure to peritoneal surfaces. Drugs for intraperitoneal administration should have known tumoricidal activity, low peritoneal permeability to maximize the duration of peritoneal drug exposure, and rapid plasma clearance. Cisplatin is one of the most commonly used agents.

The drug must be evenly distributed throughout the peritoneal cavity to optimize exposure and prevent excessively high drug concentrations in focal fluid loculations. Equal drug distribution requires the absence of adhesions, necessitating the administration of therapy perioperatively and a duration of therapy limited to days. Additionally, studies of peritoneal cavity fluid distribution after transperitoneal catheter infusion using water-soluble contrast or radionuclides have demonstrated the need for instillation of 1 to 2 liters of crystalloid to assure equal distribution.

Intraperitoneal chemotherapy has been investigated for several years in the treatment of minimal residual disease in ovarian carcinoma, but an overall survival advantage has been difficult to demonstrate because of the lack of randomized, controlled clinical trials. The efficacy of this therapy in other intraabdominal, abdominal, and pelvic malignancies is under investigation. The ideal setting for intraperitoneal chemotherapy is at the time of laparotomy or laparoscopy in a patient with positive peritoneal cytology obtained by lavage. Intraperitoneal chemotherapy is a good option for patients with smaller tumor masses or patients at high risk for recurrence.

Regional Artery Administration

Regional arterial drug therapy to specific organ sites has been used for several primary and metastatic solid tumors, including head and neck carcinomas, primary and recurrent gastrointestinal and pelvic malignancies, extremity sarcomas, in-transit metastatic melanoma, and primary or metastatic hepatic tumors.

Limb perfusion in extremity melanoma employs vascular isolation of an affected extremity through external iliac or axillary arterial and venous cannulation and placement of a proximal limb tourniquet to minimize systemic drug leakage. A high concentration of drug with or without limb hyperthermia then can be perfused over approximately 1 h and extracted via an extracorporeal dialysate pump oxygenator before reestablishment of normal limb perfusion. The maximal drug dosage is limited by limb tissue toxicity and not by systemic drug levels, and the procedure is accompanied by the considerable additional morbidity associated with operative limb vascular access and compromised wound healing.

Chemoembolization, used in the treatment of primary hepatomas or neuroendocrine tumors that have metastasized to the liver, delivers drug-impregnated pellets or sponge spheroids that temporarily or permanently occlude end-arteries selectively supplying tumor-bearing liver, slowing tumor blood flow, and increasing the time of drug exposure to the tumor. The same decrease in flow has been accomplished with internal or external, surgically placed, temporary hepatic artery occlusion catheters, which can be manipulated to cause tumor ischemia after hepatic arterial infusion of drug.

The high tissue extraction of some drugs by hepatic parenchyma has also been exploited. The fluoropyrimidines demonstrate high hepatic extraction, with removal of more than 90 percent of floxuridine (FUDR) and 75 to 85 percent of 5-FU on a single pass after hepatic artery bolus infection. Clinical trials using hepatic arterial FUDR have demonstrated that biliary sclerosis (apparently irreversible) and elevation of hepatic enzymes represent the dose-limiting toxicities seen with hepatic intraarterial FUDR administration.

Drugs such as doxorubicin and mitomycin C that have some demonstrated tumoricidal activity against primary hepatomas and metastatic colon carcinoma may also be given by hepatic arterial infusion. Systemic toxicity, however, becomes dose-limiting because the hepatic extraction of these drugs is approximately 25 percent and 12 percent, respectively. Attempts to increase the extracorporeal extraction of these and other drugs from the venous effluent of the liver and other regional infusion sites is under investigation using regional venous filtration techniques.

Measuring Response

The clinical response to adjuvant chemotherapy for micrometastatic disease is determined by the rates of recurrence, disease-free survival, and overall survival. The clinical response to induction or neoadjuvant chemotherapy for visible, palpable, or radiologically measurable tumors is determined by the change in tumor mass. A *partial response* is generally a 50 percent or greater reduction in summed measurable tumor mass. Each tumor mass is measured as the product of the two greatest perpendicular diameters. A partial response is occasionally subdivided into minor responses (less than 50 percent size reduction) and major responses (more than 50 percent size reduction, but less than a complete response). This subdivision is clinically insignificant because only a complete response has the potential for cure. A *complete response* requires total disappearance of tumor on physical examination and radiologic studies for at least 4 weeks. A complete clinical response is likely to be followed by early relapse if chemotherapy is not continued long enough to eliminate any micrometastatic disease. *Tumor progression* is defined as a greater than 50 percent increase in summed measurable tumor mass. *Stable disease* indicates no change in tumor mass, size reduction less than a partial response, or any increase in size less than progression.

Lack of change in tumor size does not always indicate a lack of tumor response. Many large tumors may undergo necrosis, fibrosis, or granuloma formation with marked destruction of viable tumor cells, but minimal or no change in size. The increasing use of neoadjuvant chemotherapy has enabled histologic evaluation of pretreated tumor and normal tissue. The biologic tumor response to chemotherapy is estimated by the percentage of visibly unaffected tumor cells identified through histologic examination of sectioned tumor.

Side Effects and Toxicity

Some degree of drug toxicity during the administration of chemotherapy is not only expected but often is desirable because it indicates a cellular damage response. With rare exceptions, the maximal tolerated dose of most chemotherapeutic agents is sought to achieve the highest tumor cell kill. Close attention to normal tissue injury is essential to optimize the risk/benefit ratio and assure that the treatment is not worse than the disease.

Patterns of organ toxicity have been well described for the different classes of drugs (Table 9-28). The degree of toxicity depends on drug concentration, duration of exposure, and host response. Anticipated drug toxicity is based on the nonspecific damage caused by most chemotherapeutic agents to rapidly proliferating normal tissues, such as bone marrow stem cells, gastrointestinal crypt lining cells, and hair follicles. Within the same class of drugs, differential organ injury may be seen. Rare idiosyncratic systemic drug reactions within any drug category may cause unexpected and occasionally life-threatening toxicity.

Objective and subjective measurements of toxicity are serially recorded to allow the delivery of maximal tumoricidal drug doses with minimal and reversible injury to normal tissues. The Eastern Cooperative Oncology Group (ECOG) and other groups have developed scales that attempt to standardize toxicity criteria for measurement and reporting in clinical trials. The severity of signs, symptoms, and alterations in laboratory tests are graded for all major organ systems. Grade 0 denotes no toxicity, and grade 4 usually denotes life-threatening toxicity. The duration, chronicity, and reversibility of toxicity are also important. Clinical trials often have provisions for dose reduction or cessation of drug therapy for toxicity higher than grade 2 or 3, depending on the organ system affected.

Biologic Therapy

Biologic therapy is the administration of any biologic molecule or multimolecular complex and includes immunotherapy and gene therapy. The most common type of biologic therapy for cancer is administration of recombinant cytokines, such as *interleukin-2 (IL-2)* and *interferon-alpha (IFN- α)* , alone or in conjunction with chemotherapy (biochemotherapy). Of increasing interest are cancer vaccines, a form of active specific immunotherapy, and gene-based interventions. Biologic therapy of cancer is generally used in a multimodal regimen to increase the efficacy of surgery, radiation, or chemotherapy. For example, biologic agents may be administered preoperatively to patients with measurable and surgically resectable metastatic disease; this allows the clinician to determine clinical response in a measurable tumor, compare pathologic response to clinical response, and examine the in vivo effects of biologics on host-tumor relationships. Alternatively, biologic agents can be administered after curative resection to destroy residual micrometastatic disease and prevent recurrence.

Considerable effort has been devoted to immunotherapy of cancer. Immunotherapy assumes that cancer progression results from failure of the host immune defenses to recognize and reject the tumor. Biologic agents augment the immune response with the goal of blunting tumor progression. Theoretically, the immune system may be activated or reactivated to attack and destroy tumor specifically, leaving normal tissue largely unaffected.

Implicit in the ability of the immune system to recognize and attack neoplastic cells is the existence of immunogenic tumor-associated antigens. Evidence that human tumors are immunogenic comes primarily from investigations using melanoma, which is one of the most immunogenic solid tumors. Blood from melanoma patients contains antibodies against tumor antigens as well as cytotoxic T cells (CTL) that can destroy tumor cells in vitro. Clinical studies indicate that approximately 3 to 15 percent of all cutaneous melanomas are first diagnosed as lymphatic or visceral metastases without evidence of a primary tumor, which suggests that the immune system has caused complete regression of the primary melanoma. Histopathologic evidence of regressive changes has been reported in up to 58 percent of primary melanoma specimens. On rare occasions, there has been spontaneous, complete regression of metastatic disease.

In most experiences, induction of antitumor antibody affords little protection. In rodent models, immunity to solid tumors cannot be transferred to normal syngeneic hosts using serum or purified antibody preparations. By contrast, T lymphocytes play a critical role in the rejection of solid tumors in these models. In addition to the adoptive specific immunity afforded by T lymphocytes, natural killer (NK) cells can lyse a wide variety of tumor and virus-infected cells without the antigen-specific receptors used by T or B cells.

Rapidly emerging advances in the basic mechanisms of cell-mediated immunity provide new strategies for biologic therapy based on the prospect that the host immune system may be manipulated either in vivo or ex vivo to reject neoplastic outgrowth. Molecular biology and cell cloning enable investigation of a new level of the biology of host-tumor relationships and development of biologic agents to administer to cancer patients. No single possibility has universal application, but as our understanding of the host immune response to autologous tumor antigens in various neoplastic diseases improves, the strategies for biologic approaches to treatment will improve vastly.

Recombinant Cytokines

Table 9-29 lists the cytokines currently in use for biotherapy of cancer. Early trials with nonrecombinant interferon in patients with advanced melanoma demonstrated a major response rate of 14.4 to 22.6 percent. Subsequent trials using recombinant *IFN-a* for metastatic melanoma showed a major response rate of about 23 percent (range 14 to 28 percent). *IL-2*, the T cell growth factor, induces a major response, particularly in patients with metastatic melanoma and metastatic renal cell carcinoma, but not in patients with breast cancer, colon cancer, or lymphoma.

The treatment of melanoma was significantly changed when Kirkwood and associates reported that high-dose *IFN-a-2b* significantly prolonged both relapse-free and overall survival rates after surgical resection of high-risk primary melanoma (American Joint Committee on Cancer (AJCC) stage II-B) or regional lymph node metastases (AJCC stage III). IFN-a-2b administered intravenously (20 million U/m²/day) for 4 weeks and then subcutaneously (10 million U/m² three times a week) for 48 weeks increased median disease-free survival by 1 to 1.7 years and median overall survival by 2.8 to 3.8 years, compared with observation. Interferon treatment also increased the rate of 5-year survival by 24 percent.

This important study was the first randomized controlled trial to show a significant benefit of adjuvant therapy in prolonging relapse-free and overall survival of high-risk melanoma patients. On the basis of the results of the study, the Food and Drug Administration approved *IFN-a-2b* for postoperative adjuvant therapy in melanoma patients at high risk of systemic recurrence.

The intravenous administration of the cytokines is not without significant toxicity, however, and it is clear that they trigger a cascade of effects that results in other lymphocyte activities as well as the direct effects of *IL-2* on other tissues. Some of the side effects are similar to those seen with septic shock. Concern over the toxicity of intravenously administered cytokines has promoted the development of methods to target cytokines and reduce systemic effects. Cytokines are administered aggressively and frequently because of their short half-life in the blood. In addition, cytokines are not always distributed to all areas of the body and often do not traverse the blood-brain barrier. An innovative approach is the incorporation of interferon and adjuvants into liposomes that direct the cytokine to a specific host cell population. In animal models, liposomes are ingested by macrophages. Degradation of the “packaged” therapy results in intracellular release of adjuvants, such as *muramyl tripeptides (MTP-PE)* , and the interferon, which activates macrophages to become cytotoxic against cancers in vitro and in vivo. A randomized study in dog osteogenic sarcoma showed significantly greater response and survival rates in animals that received the MTP-PE liposomes as compared with those that received empty liposomes. This novel approach has been extended to human studies, beginning with a pilot study of patients with osteogenic sarcoma and melanoma.

Combination biologic therapy is now in clinical trials for treatment of most major human cancers. The multiagent concept is plausible because: (1) multiple immune abnormalities are most likely to occur in cancer patients; (2) there is heterogeneity in immune response (nature of lymphocytes, role of antibody, presence of macrophages) relative to the site of the metastases; and (3) combinations of agents with different mechanisms of action are more likely to augment individual aspects of immune response additively or synergistically in a diverse population of cancer patients. For example, combinations of *IL-2* and *IFN-a* elicit a higher rate and more durable response time for metastatic melanoma than either cytokine alone. Combinations of tumor antigen, lymphokines, and cyclophosphamides are intended to activate tumor-specific immunity, promote effector T cell proliferation, and down-regulate suppressor T cells.

Biochemotherapy uses cytokines such as *IL-2* and *IFN-a* in combination with chemotherapeutic agents such as *5-FU* with the goal of enhancing antitumor activity. Various biochemotherapeutic regimens are being examined in patients with metastatic colorectal cancer, lung cancer, renal cell carcinoma, and melanoma. Cytokines also have been used as supportive therapy to allow higher doses of chemotherapy. For example, *granulocyte-macrophage colony-stimulating factor (GM-CSF)* has been administered with erythropoietin to allow acceleration and dose escalation of chemotherapy with cyclophosphamide, epidoxorubicin, and 5-FU in patients with advanced breast cancer.

Immunotherapy

Immunotherapy is a logical adjunct for the treatment of subclinical microscopic disease after definitive cancer surgery, radiation therapy, or chemotherapy for the following reasons: (1) patients who have only small foci of cancer cells remaining after destruction of the major tumor bulk are the most likely to benefit from immunotherapy, because the tumor mass that must be destroyed is smallest at that time; (2) the specificity of the immune response provides a possible therapeutic tool that has selectivity for small numbers of cancer cells not possible with any other therapeutic modality; (3) patients with disease in earlier stages are more likely to respond to immunotherapeutic maneuvers, since the cancer patient's general immune competence is greatest when the disease is localized and is often impaired after metastasis; and (4) immunotherapy should complement rather than interfere with currently available methods of cancer therapy. Because both irradiation and chemotherapy are immunosuppressive, the use of immunotherapy in combination with these therapeutic modalities must be carefully controlled.

Table 9-30 describes the types of immunotherapy. The rational application of these immunotherapies to human cancer will depend on a better knowledge of tumor-associated antigens in human neoplasms and methods for increasing the immune response against these agents. Their use should be limited to cancer facilities, where the effects of this form of treatment can be scientifically evaluated.

Active Specific Immunotherapy (Cancer Vaccines)

The clinical use of cancer vaccines was initiated at the turn of the century, prompted by the success of vaccines against infectious disease. Unlike vaccines against infectious disease, which are administered prophylactically, cancer vaccines are generally administered after the advent of disease. Both types of vaccine utilize attenuated whole cells, cell walls, specific antigens, or nonpathogenic strains of living organisms to stimulate the patient's immune system to fight the disease. The specific goals of active immunotherapy with cancer vaccines are to overcome the immunosuppression produced by tumor-derived factors, to stimulate specific immunity that will destroy tumor cells, and to enhance the immunogenicity of tumor-associated antigens (TAA).

Several observations support the potential value of active specific immunotherapy for the treatment of cancer. These include (1) vaccine- induced immunity against cancer in animal models, (2) the regression and eradication of tumors injected directly with immunostimulants, (3) occasional regression of noninjected tumors after the intralesional injection of *bacille Calmette-Gu érin (BCG)*, and (4) the development of antitumor antibodies.

The goals of active immunotherapy are to combat the decreasing immunocompetence that results from interaction of tumor-related products with components of the immune system and to eliminate the diverse clones within a tumor cell population by augmenting cellular and/or humoral immunity. Unlike chemotherapy, vaccine immunotherapy has no direct cytotoxicity on tumor cells; its clinical effect depends on the immune system and is slower in onset but tends to be more durable (Table 9-31).

Adoptive Immunotherapy

In adoptive immunotherapy, immune lymphoid cells are transferred to a recipient to mediate tumor destruction. Rosenberg and colleagues pioneered the study of adoptive immunotherapy using lymphokine- activated killer (LAK) cells, which are cytolytic lymphocytes generated in the presence of *IL-2*. These cytolytic cells can kill a wide range of fresh and cultured human cancer cells, but not normal cells. Human LAK cells do not have T cell markers and are not *MHC-* or antigen-specific in their killing.

Clinical trials using autologous LAK cells (obtained by repeated leukopheresis and in vitro IL-2 expansion) and systemically administered IL-2 produced clear, objective responses in some patients with bulky metastatic cancer. Some evidence suggests that LAK cells may be more important in renal cell carcinoma than in melanoma. Subsequently, the method was developed for isolating *tumor-infiltrating lymphocytes (TIL)* from human melanoma and renal cell carcinoma; after proliferation ex vivo in the presence of IL-2, the TIL were returned to the patients and IL-2 therapy administered concurrently. In preliminary studies, response rates of up to 40 percent were obtained. Unlike LAK cells, the TIL may have an important role in the therapy of local and regional disease, while LAK cells may remain useful in the treatment of hematogenous dissemination and some renal cell carcinomas. TIL-based immunotherapy is an active area of research, and TILs also are being investigated in conjunction with gene therapy.

Nonspecific Immunotherapy

Certain substances, such as mixed bacterial toxins and fractions of the tubercle bacillus, nonspecifically enhance host resistance to most viral, fungal, and bacterial agents. Although the exact mechanism is unknown, these agents appear to stimulate immune response to a wide variety of antigens, including tumor antigens.

Interest in a nonspecific immunotherapy was revived more than 20 years ago using attenuated bovine tuberculosis bacillus (bacille Calmette-Guérin, BCG). Some tumor regressions were observed, but consistent responses in any one treatment group were difficult to achieve. There are several possible mechanisms to explain tumor regression following BCG injection; specific and nonspecific immune reactions were probably involved. The tumor cells may be killed as “innocent bystanders” during the delayed cutaneous hypersensitivity reaction that occurs when lymphocytes and macrophages attack BCG dispersed throughout the tumor nodule. This is supported by the observation that the intratumor injection of BCG works only in patients who can be sensitized to BCG, as shown by their delayed cutaneous hypersensitivity reaction to purified protein derivative.

In addition to the nonspecific effect, a specific immune response to melanoma-associated tumor antigens occurs in some patients because an associated rising titer of antimelanoma antibody is observed after BCG immunotherapy. Sequential biopsies of tumor nodules taken after BCG inoculation reveal that the regression of these nodules is associated with granulomatous infiltration of lymphocytes, monocytes, and fibroblasts surrounding and infiltrating the melanoma cells. The regression of melanoma nodules not injected with BCG is accompanied by the appearance of lymphocyte infiltrates within the regressing melanoma tumor nodules. The specific antitumor effect may result from more lymphocytes and macrophages coming into contact with the tumor cells and promoting antigen presentation. Conversely, it may work via the effector limb of the immune response by bringing greater numbers of stimulated and unstimulated lymphocytes to the tumor.

Other nonspecific agents include *Corynebacterium parvum*, *Bordetella pertussis*, *MTP-PE*, methanol-extractable residue of BCG, bacterial endotoxins, and polynucleotides. Another form of nonspecific immunotherapy involves the use of agents capable of restoring depressed immune responses. Several agents have been proposed, including thymic hormones, such as thymosin, and the antihelminthic drug levamisole.

Passive Immunotherapy

The systemic use of tumor-specific antiserum is laden with theoretical and practical problems. Passive immunotherapy is effective only in suppressing small numbers of tumor cells and must work in concert with host effectors (e.g., complement, macrophages, antigen-dependent cellular cytotoxicity) to effect a cytotoxic action on target cells. In addition, only antibodies of certain classes and subclasses can interact effectively with certain cellular effectors. Most of the better-characterized human tumor-specific antisera are murine monoclonal antibodies that, because of their antigenicity, have limited applications in human beings.

Immunotoxins are tumor-specific antibodies that are attached to toxic molecules. This concept, first proposed by Ehrlich a century ago, uses the antibody molecule to preferentially localize anticancer agents in the vicinity of tumors. It obviates the need for the host to supply effector cells or complement to mediate tumor destruction. Monoclonal antibodies are preferred to heterologous antiserum because they permit the use of homogeneous, purified antibodies of defined specificity. A wide range of toxic molecules has been tested in vitro and includes radioactive isotopes, traditional cancer drugs, and plant and bacterial toxins. Recombinant DNA technology now permits the creation of hybrid or chimeric immunotoxin molecules in which the Fc fragment of the immunoglobulin molecule has been replaced by a polypeptide toxin sequence. Immunotoxins are currently undergoing clinical trials, although their overall therapeutic efficiency in clinical oncology is unproved.

Gene Therapy

Gene cloning has introduced a new era of biologic therapy that will have an impact on human clinical trials in the coming years. An example is the studies by Hellstrom and colleagues, in which a cloned gene for one of the major human melanoma tumor-associated antigens (p97) was introduced into the genome of a vaccinia virus. The virus then expressed both immunogenic viral antigens, plus the weaker immunogenic human melanoma antigens on the surface. This reagent will be tested in human studies in the near future. A novel approach is transfection of human *TIL* with genes for producing cytokines, such as tumor-necrosis factor (TNF). The ability to transfect cytokine genes into human *TIL* suggests adaptive cellular therapy with genetically transfected cells capable of producing high concentrations of tumor-necrosing factor or other lymphokines at the tumor site. This would deliver high concentrations of cytokine to the tumor site while sparing the vascular compartment of the otherwise deleterious effects of high-dose, systemic cytokine. Gene transfection also may be used to augment the immunogenicity of tumor vaccines. Preclinical studies demonstrate enhanced cellular immunity when whole-cell vaccines are transfected with IL-2, IL-4, IL-7, and/or *GM-CSF*. Several phase I and phase II clinical trials of cytokine-transduced cellular vaccines for patients with advanced melanoma are under way.

Other approaches to gene therapy of cancer are antisense oncogene and tumor suppressor gene therapy, which attempt to correct genetic disorders of cancer by suppressing the abnormal expression of proliferative genes. Potent regulatory antisense DNA sequences can prevent the translation of molecules that down-regulate host immunity to tumor.

Management of Distant Metastases at Specific Sites

Lung, Pleura, and Mediastinum

Two of the most common initial sites of metastasis are the lungs and pleura. Although pulmonary metastases are generally asymptomatic, they may cause persistent cough, shortness of breath, and/or chest pain. An irritating, dry, and unproductive cough may progress to hemoptysis.

A standard chest x-ray is sufficiently sensitive and cost-effective for screening all cancer patients and frequently will reveal hilar and mediastinal adenopathy in those with pulmonary metastases. Although pulmonary tomograms or CT scans have too low a yield and too high a cost to be justified when the chest x-ray is normal, they are of value in evaluating suspicious chest lesions or in determining whether the metastatic disease seen on the chest x-ray is present elsewhere in the chest. Tomograms can detect lesions as small as 6 mm in diameter, and CT scans can identify lesions as small as 3 mm. Their increased sensitivity is offset by decreased specificity; the false-positive rate can be as high as 15 percent. The choice between the two tests often depends on available facilities, cost, and skills of the radiologist. Neither test is indicated unless the presence of pulmonary metastases would alter the treatment plan or unless a better definition of lesions is required for entry into a research protocol.

Bronchoscopy with biopsy may be considered when the etiology of a pulmonary lesion is in doubt (e.g., metastatic disease, fungal disease, benign tumor, or bronchogenic carcinoma), especially when symptoms suggest bronchial involvement (e.g., a productive cough or a centrally placed or cavitory lesion). A scalene lymph node biopsy is indicated for palpable nodes. Mediastinoscopy is indicated if the chest x-ray, tomogram, or CT scan reveals abnormal mediastinal nodes that are accessible through the instrument. Thoracentesis or pleural biopsies may be helpful when evaluating effusions. Fine-needle biopsy of a pulmonary lesion under CT scan guidance may be useful in selected instances to establish the histologic diagnosis. Video-assisted thoracoscopy is increasingly used both diagnostically and therapeutically in the staging and treatment of lung cancer. This technique permits visualization of the entire visceral, parietal, and mediastinal pleural surfaces and excisional or incisional biopsy for establishing diagnosis. If the diagnosis remains in doubt, an exploratory thoracotomy may be necessary, especially for a solitary lesion, because some patients will have potentially curable primary lung cancer.

The treatment approach is determined by the location and number of thoracic metastases and by the patient's overall status. Nearly all patients with disseminated cancer develop metastases in the chest prior to death. Some neoplasms, such as melanoma, metastasize preferentially to the lungs. Surgical resection may be indicated even when there is more than one pulmonary metastasis. In most surgical series, the median survival is 17 months, with a 5-year survival rate of 20 to 25 percent. Some patients live more than 10 years. Resection of a solitary pulmonary metastasis provides a higher rate of 5-year survival than resection of primary bronchogenic carcinoma of the lung. Criteria for resection include absence of metastases at other sites, control of the primary tumor, potential for complete resection, and a long tumor doubling time. CT scans should be obtained preoperatively because the number of lesions demonstrated by CT scanning is often greater than that shown by chest x-ray. Lung parenchyma should be conserved during resection. Most metastases occur just below the pleura, and a wedge of tissue removed by segmental resection suffices. Stapling, electrocautery, and laser surgery can be useful. Lobectomy and pneumonectomy usually are not indicated.

Patients who are ineligible for surgery, such as those with multiple slow-growing tumors, might be monitored but receive no treatment while they are asymptomatic. If the pulmonary metastases progress rapidly, chemotherapy can be considered, especially if multiple visceral metastases exist at other sites or if the patient has disease symptoms. A pleural effusion associated with lung cancer does not rule out resection; malignant pleural disease must be documented. If patients are symptomatic, they can be treated by thoracentesis, tube thoracostomy, or thoracoscopy with chemical pleurodesis. Closed chest tube drainage and sclerosis often are necessary for effective control of recurrent symptomatic effusions. Talc diluted in normal saline solution is the agent of choice for chemical pleurodesis. Talc can also be delivered via video-assisted thoracoscopy. Fever is a common side effect, and chest pain, dyspnea, and pulmonary infiltrates sometimes occur. Pleural effusions are generally associated with a short survival time.

Tumor Doubling Time

The growth rate of a tumor can be expressed by the time the tumor doubles in volume. The tumor doubling time (TDT) is an accurate and reproducible measure of biologic aggressiveness that can be used to determine the indications for surgical resection. TDT represents the balance between the intrinsic proliferative rate of the tumor cell and the patient's immune defense mechanisms.

TDT measurement is especially useful in treating patients with pulmonary metastases because neoplasms tend to be peripherally located and discretely identified on chest radiographs. It is quite easy to obtain accurate serial chest x-rays that can be used to measure the changing diameters of the lesion. The greater and lesser diameters are averaged and then plotted against time on semilogarithmic paper. The slope of the line drawn between any two points represents the rate of tumor growth. The horizontal distance between any two doubling points represents the TDT in days.

TDT may vary from 8 to 600 days, but most tumors double in 20 to 100 days. Patients with a short TDT have aggressive, fast-growing metastatic lesions; patients with a long TDT might have nonaggressive lesions that would be responsive to surgery. TDT is an important prognostic tool for selecting surgical candidates. TDT also can be used to monitor the effects of chemotherapeutic agents and compare different therapeutic regimens. Patients with pulmonary metastases can be divided into three survival groups according to TDT. Those patients with TDT less than 20 days are not recommended for surgery; it is likely to be ineffective and will not result in long-term survival. Patients with a TDT of 20 to 40 days are not ineligible for surgery, particularly if a slowing of TDT is observed after preoperative chemotherapy; their long-term survival rates are not much improved by surgery alone. Patients with a TDT of 40 days or more can have long-term survival after resection of the pulmonary lesion. Sarcoma patients with a TDT of more than 40 days were found to have significant palliation from pulmonary resection and remained free of disease for as long as 5 years; patients with a TDT of less than 20 days did not significantly benefit from resection of metastatic lesions.

Liver, Biliary Tract, and Spleen

Hepatic metastases can occur in many patients with metastatic disease, especially those with gastrointestinal malignancies and breast cancer. There are no reliable and accurate tests for early detection of liver metastasis, and no common symptoms and physical signs. The patient might experience decreased appetite with loss of weight followed within weeks by general lassitude and debility. The loss of appetite can precede a clinically palpable liver by a month or more. Conversely, the patient may have an easily palpable liver and feel perfectly well. As the liver disease progresses, nausea, vomiting, jaundice, night sweats, and fever can develop.

A history and physical examination and serum liver chemistries with appropriate tumor markers are the most cost-effective screening tests. Elevated levels of lactic dehydrogenase or alkaline phosphatase in the presence of normal or only slightly elevated levels of serum glutamic- oxaloacetic transaminase or bilirubin suggests liver metastasis. An elevated lactic dehydrogenase level is a clinically useful and relatively specific indicator for metastatic melanoma.

Suspected liver metastasis should be confirmed by ultrasonography, dynamic CT scan, or CT portography. Radionuclide liver scanning and hepatic arteriography are less frequently used. The choice depends on the availability and cost, the interpretive skills of the radiologist, and the generation of equipment used, which is especially important in CT scanners and ultrasound units. Abdominal CT scans are more accurate and reliable than ultrasonography and radionuclide liver scans for evaluation of liver masses. *PET* scans also are increasingly useful in detecting metastatic disease. Hepatic metastases are not detected by radiologic tests until they are more than 1 cm in diameter.

Angiography is used only when the differential diagnosis cannot be established by noninvasive techniques, when the information gained would affect the treatment decision, or when hepatic resection is contemplated. Biopsy usually is not necessary to confirm the diagnosis of liver metastasis. In the few instances in which biopsy confirmation is essential to treatment decisions, a needle biopsy can be performed percutaneously with CT or ultrasound guidance, or by laparoscopy, or during laparotomy.

Some patients with isolated liver metastases from colorectal cancers can benefit from surgical resection. Those patients with a solitary metastasis or metastases located in one lobe are often successfully treated with resection, and approximately 25 percent will survive for 5 years. Most liver metastases are not amenable to surgical excision. Systemic chemotherapy or hepatic arterial chemotherapy is the most common intervention for patients with nonresectable hepatic metastasis, and response rates vary. Cryosurgery might offer effective palliative treatment for patients with nonresectable primary or metastatic hepatic malignancies; in certain cases, extended survival has been reported with the potential for cure. Other treatments include hepatic artery embolization, chemoembolization, radiation therapy, and alcohol injection.

Brain and Spinal Cord

Many cancers, particularly breast cancer, lung cancer, and melanoma, metastasize to the brain, a common cause of death. Headache and mental deficits are the most common symptoms of brain metastasis. Headache resulting from brain metastasis characteristically begins as a mild morning headache. As the condition progresses and the intracranial pressure increases, the headache will persist longer into the day and become more severe. It is usually generalized, but it may be slightly worse in the frontal or occipital region and is sometimes associated with visual changes. The most common physical sign of brain metastasis is a focal neurologic deficit; seizures are common. The presence of papilledema is a helpful sign, but its absence is not useful diagnostically.

The best tests for diagnosing intracerebral metastasis are MRI and CT with contrast enhancement. MRI, a technique that depends on the intrinsic paramagnetic properties of biologic tissue, is generally the preferred test to detect and stage brain and spinal metastases. MRI can detect tumors that cannot be detected by CT scans. MRI has an advantage particularly for lesions at the base of the skull and in the posterior fossa, because only a weak signal is generated from the adjacent bone. MRI scans appear to have a higher tissue sensitivity, including a better ability to distinguish hemorrhage from tumor, than CT scans.

The accuracy and sensitivity of these scans make it unnecessary, in most cases, to perform a radionuclide brain scan or electroencephalogram unless there are some equivocal findings. A carotid arteriogram may be indicated to rule out vascular abnormalities. A lumbar puncture and cerebrospinal fluid analysis to diagnose meningeal involvement occasionally is necessary (after a CT scan) for a patient who has a cranial nerve palsy, bladder dysfunction, or nonlocalizing or bilateral neurologic signs and symptoms.

The mainstay of initial treatment is corticosteroids, the most effective of which is dexamethasone (up to 100 mg/day). Dexamethasone reduces edema around the tumor and temporarily helps to relieve symptoms in the majority of patients. The steroid dose should be tapered over 2 to 4 weeks as tolerated and the therapy stopped after definitive treatment unless the patient's symptoms intensify during steroid withdrawal. Steroids often do not help patients with rapid neurologic deterioration because this condition usually reflects intracerebral hemorrhage around the metastasis. Chemotherapy is not usually effective for brain metastasis.

Surgical excision followed by cranial irradiation is the treatment of choice for a solitary, surgically accessible metastasis. Tumor excision by means of a craniotomy is relatively safe; it alleviates symptoms in most patients and prevents further neurologic damage in patients with demonstrable metastases. The treatment may be considered in some patients who have disease at other sites plus symptomatic brain metastases because their estimated life span can exceed 3 months, and their neurologic status usually improves.

Patients treated with open brain surgery and fractionated radiotherapy have a better outcome than those treated with radiation alone, but many patients do not have surgically accessible cerebral metastases. In these cases, stereotactic radiosurgery using the "gamma knife" may offer the best chance of prolonged survival.

Bone

Bone metastases are common in patients with advanced breast or prostate cancer but infrequent in patients with gastrointestinal cancers. They are medullary in location and destructive in nature. The pain from bony metastases is typically nocturnal at first, becoming persistent, progressive, and localized, and it can become quite severe.

Bone metastases are frequently diagnosed in symptomatic patients, but occasionally they are seen incidentally on radiographs (e.g., rib metastases on routine chest x-ray) or a bone scan prompted by an elevated serum alkaline phosphatase level in the absence of liver metastasis. They are generally osteolytic in appearance on radiography and provoke little if any bone formation, but some patients with prostate cancer have osteoblastic bone metastases. Axial metastases account for up to 80 percent of bony lesions and are most common in the spine and rib. When bony lesions involve the vertebral body, there are often compression fractures that may lead to neurologic symptoms, such as radicular back pain, paresthesia or paresis of the legs, and urinary retention. Only about 10 percent of lytic lesions occur in weight-bearing bones, but these could result in pathologic fractures.

The radionuclide bone scan is the initial test for evaluating suspected bone metastases. Its sensitivity is reportedly 50 to 80 percent greater than radiographs alone, but bone scan abnormalities are nonspecific and must be correlated with radiographic study (x-ray or CT scan) and patient history (fractures, trauma, arthritis, etc.) to distinguish between benign and malignant causes. A bone biopsy might be necessary to establish the diagnosis before instituting treatment.

The treatment of bone metastases depends on the degree of symptoms, the location and magnitude of the lesions, and the patient's life expectancy. The goals of therapy are to relieve pain and maximize ambulation. Patients without symptoms should be monitored to assess the progression of their lesions but should receive no major treatment unless they become symptomatic.

Symptomatic metastases frequently involve non-weight-bearing bones, particularly the spine and ribs. In these cases, irradiation of the lesions usually provides relief. The radiation fields should be restricted to those lesions responsible for the symptoms. Symptomatic bone lesions only occasionally respond to systemic chemotherapy, but bone metastases from breast cancer sometimes respond well to hormonal therapy.

Symptomatic metastases in weight-bearing bones (e.g., the femur) require special consideration. If the lesion is large, and if there is evidence of cortical destruction, prophylactic stabilization and irradiation are sometimes used when the patient's life expectancy is at least 2 months. Stabilization includes operative metallic bone fixation (e.g., with intramedullary rods), joint replacement, repair with methyl methacrylate, or external braces or a cast. Radiotherapy is generally given postoperatively. Alternatively, the lesion might be treated with radiation alone, but the patient must be closely monitored for evidence of pathologic fracture. Unless the surgical risk is high or the patient's expected life span is short, pathologic fracture of a weight-bearing bone should be stabilized, maximizing the patient's quality of life and decreasing hospital or nursing home costs. Patients with fractures of the vertebrae that have compressed the spinal cord require prompt treatment to avert paralysis. The treatment may require decompressive laminectomy and postoperative irradiation, or irradiation alone, depending on the extent of the disease and the patient's overall medical condition.

Skin, Subcutaneous Tissues, and Distant Lymph Nodes

These are the most common sites of distant metastases and are often the first sign of hematogenous spread. Skin and subcutaneous metastases generally are 0.5 to 2.0 cm in diameter and are readily detectable by physical examination. Distant lymph node metastases can occur in any area. The more superficial nodal metastases are easily diagnosed by physical examination. Metastases within the thorax usually can be detected on chest x-ray, with CT scans or tomograms used as confirmatory tests, but abdominal nodal metastases are generally detected by CT or ultrasound scans. These lesions usually are treated, especially when symptomatic and isolated, by surgical excision (if superficially located) or radiation therapy. Other treatments include intralesional immunotherapy with *BCG*, regional chemotherapy (isolated perfusion of the extremities), or systemic chemotherapy or hormone therapy, especially when there are multiple lesions or simultaneous visceral metastases.

Gastrointestinal Tract

Some cancers, particularly melanoma, can metastasize to the gastrointestinal tract, often to multiple sites. Early gastrointestinal involvement usually causes persistent but nonspecific complaints, such as epigastric distress, nausea, anorexia, or weight loss. The most common clinical manifestations are the result of chronic bleeding (anemia, anorexia, and weight loss), obstruction of the small bowel (abdominal pain, nausea, and vomiting), and acute bleeding (hematemesis or melena). For example, a patient with melanoma metastatic to the gastrointestinal tract typically presents with a symptom complex of gastrointestinal complaints, guaiac- positive stool, and anemia.

Gastrointestinal metastases are difficult to detect by radiologic studies, the routine use of which is not indicated for screening. Follow-up clinical examination should include a guaiac test of a stool specimen. Although subject to limited specificity, a positive result is an indication for further investigational studies using barium enema, colonoscopy, esophagogastroduodenoscopy (EGD), or CT scans of the abdomen and pelvis. Newer studies, such as *PET* scans and radiolabeled monoclonal antibody scans, might be helpful in identifying the general location of a metastasis. Persistent anemia without an identified lesion may indicate further studies, such as small bowel enteroclysis, radionuclide scans, or angiography, with possible embolization to control persistent bleeding from nonoperable lesions.

Repeated blood transfusions might be indicated for persistent anemia in patients who have nonresectable visceral metastases or for whom a thorough workup failed to reveal a source of bleeding. Chemotherapy can be considered for patients with multiple gastrointestinal lesions, but surgical excision should be considered for *isolated* gastrointestinal metastases if the patient's condition is good. Surgery is recommended for most patients who have acute complications of obstruction, massive bleeding, or perforation because the alternative is to allow the patient to die. The final decision depends on the patient's overall clinical condition, but symptoms can be successfully alleviated in most cases, and survival after surgical excision averages 4 to 8 months.

PSYCHOLOGIC MANAGEMENT AND REHABILITATION

The physician can ease the cancer patient's fear of the disease by free and open communication. Psychologic support and education are necessary for the patient to deal with any disability that can result from therapy. Examples include training in the care of a stoma following curative surgery for colonic and rectal cancer and referral to lay groups associated with the American Cancer Society for counseling the anxious patient with an altered body image resulting from mastectomy.

It is impossible to predict the exact course of any malignant tumor. Patients with a poor prognosis are occasionally cured by aggressive therapy, and spontaneous regressions are sometimes observed in patients with metastases. In contrast, some patients with apparently localized disease can die of disseminated cancer in a few months. Uncertainty about the future is one of the most difficult adjustments that cancer patients and their families face. It is reassuring to emphasize that the chances for cure improve each month after successful treatment of the primary neoplasm, particularly for tumors such as squamous cell carcinoma of the lung or oropharynx. Other, more slowly growing neoplasms, such as carcinoma of the breast and malignant melanoma, can recur after disease-free intervals of 10 to 20 years, though the chances of recurrence also decrease with time. Recognition that cancer is a chronic disease is an important aspect of management. Long-term, consistent follow-up provides opportunities for reassurance and usually can ensure detection of recurrence at an early stage.

Some patients suspect the worst but do not want to hear the truth from their physician. However, a lie is never appropriate, even if requested by the family. Untruths often create barriers between patients and their families that can lead to psychologic isolation of patients, who are unable to discuss their fears and anxieties with those they need most. Gentle and optimistic truth is generally the best approach, even when primary cancer therapy has failed and the patient is judged incurable. Realistic and consistent support is actually more important to the patient and family at this stage of the disease than earlier. There is increasing evidence that patients tolerate the process of dying much better when sustained by the physician's continuing concern and active support.

Some incurable patients are unable to accept the realities of the situation. In this case, it is essential that a responsible family member be informed. The duration of the incurable patient's life is so uncertain that predictions should be avoided. If, as frequently happens, the relatives insist on some estimate, a combined minimum-maximum prognosis, such as from 6 months to 2 years, will help the family accept this uncertainty.

The basic aim in caring for the patient with advanced cancer is to prolong useful life, but not useless suffering. The patient should be permitted to die with dignity when active therapy can no longer be of benefit.

SURGERY IN THE FUTURE

Increasing evidence suggests that surgery should be the last, rather than the first, therapeutic intervention. Surgery undertaken after chemotherapy and radiation therapy can remove cancer cells that are resistant to these modalities. Chemotherapy and radiation therapy before surgery can shrink the tumor mass to be excised and improve the possibility of organ preservation. There have been promising results from preliminary trials using these concepts in bone and soft tissue sarcomas, locally advanced breast cancer, and other neoplasms.

Surgery may be considered a form of immunotherapy. Any therapeutic maneuver that lowers tumor burden can reverse tumor-associated immunosuppression, altering the immune balance in favor of the patient. Tumor-associated immunosuppression is caused by antigens shed from a growing neoplasm into the blood and by humoral factors produced by or in response to the neoplasm. Cancer surgery effectively removes the cancer cell mass that produces the immunodepression and allows the patient's immune responses to recover. Once the tumor mass has been removed, the patient's immune system can deal with any clinically silent micrometastases. Surgical resection of apparently localized tumors can favorably affect the host-tumor relationship and may even cure the patient with subclinical distant metastases. By reducing the number of tumor cells, it also increases the curative potential of systemic therapies.

Early Detection

In many neoplasms, prognosis depends on the status of the lymph node basin draining the primary tumor. The extent and timing of lymph node dissection is controversial. Sentinel lymphadenectomy, a promising technique for early detection of nodal disease, is under investigation in multicenter trials. Detection of the sentinel node (i.e., the first lymph node draining a primary tumor) was introduced for melanoma and is being applied to breast carcinoma and other neoplasms. Initially the technique relied on injection of a vital blue dye at the tumor site and visual tracking of this dye along the lymphatics to the nodal basin. Sentinel node mapping is now facilitated by adding a radiolabeled isotope to the dye and monitoring its path by a handheld gamma probe.

Radioimmunoguided surgery (RIGS) using murine monoclonal antibody (MAb) enables surgeons to localize previously undetected malignant lesions and disseminated disease intraoperatively. It is especially effective in localizing primary and secondary colorectal, gastric, and ovarian carcinomas. For RIGS, MAb is conjugated to ¹²⁵I, which has a half-life of about 60 days. Radiolabeled MAb is administered intravenously approximately 2 weeks before surgery. To avoid radioactive uptake by the thyroid gland, a supersaturated solution of potassium iodide is administered 2 days before the MAb and continued for 3 weeks. During surgery, the surgeon manipulates the gamma-detecting device, consisting of a detection crystal, a preamplifier, and a signal processor with a digital readout. This device produces audible and numerical displays as it encounters radiolabeled tumor cells. It allows the surgeon to define tumor margins, to seek out malignant lesions that might have escaped previous detection by CT scan or plain chest x-ray, and to examine thoroughly those sites that might contain tumor cells.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 10 - Transplantation

Hans W. Sollinger
Anthony M. D'Alessandro
Mark H. Deierhoi
Munci Kalayoglu
Allan D. Kirk
Stuart J. Knechtle
Jon S. Odorico
Bruce A. Reitz
David D. Yuh

INTRODUCTION

Transplantation is the process of taking a graft—cells, tissues, or organs—from one individual—the donor—and placing it into another individual—the recipient, or host. If the graft is placed into its normal anatomic location, the procedure is called *orthotopic* transplantation (e.g., heart transplants, liver transplants). If the graft is placed in a different site, the procedure is called *heterotopic* transplantation (e.g., kidney transplants, pancreas transplants). The genetic relationship between donor and recipient is described by the following terms: *syngeneic* refers to transplantation between individuals of a genetically identical strain, or between identical twins. In this situation, the genetic makeup of donor and recipient are identical, and rejection does not occur. Transplantation between genetically different members of the same species is referred to as *allogeneic* transplantation, such as transplantation of tissues or organs between two different strains of experimental animals (e.g., rat strain ACI to Lewis). In human beings, all transplants except those between identical twins are allotransplants. Transplants between members of different species (e.g., pig to man) are referred to as *xenogeneic* transplants. If a graft (e.g., skin) is transplanted from one site to a different site within the same individual, it is referred to as an *autotransplant*.

Although attempts at transplantation date back to ancient times, the impetus for modern transplantation was World War II and the Battle of Britain. Royal Air Force pilots often were severely burned when their planes crashed. The mortality rate associated with burns corresponds to the size of the area of skin that has been injured, and the survival rate can be improved if burned skin is replaced. For this reason, British doctors turned to skin transplantation from other human donors as a mode of therapy. However, attempts to replace damaged skin with skin from unrelated donors were uniformly unsuccessful. Over a matter of several days, the transplanted skin would undergo necrosis and fall off. This problem led many investigators, including Nobel Prize winner Sir Peter Medawar, to study skin transplantation in animal models. His work, performed in collaboration with Gibson, formed the basis of modern transplantation biology. Medawar could establish that the failure of a skin graft to “take” was the result of a process later termed *immunological rejection*, which was mediated by the recipient's white blood cells. Later studies by Gowens revealed that lymphocytes play a major role in transplant rejection.

In the late 1940s Medawar became aware of an observation made by Dr. Ray Owen from the University of Wisconsin. Owen had observed that nonidentical cattle twins who shared a placenta in utero were chimeric for red blood cells carrying the genetic makeup of the nonidentical twin for the rest of their lives. When skin grafts were exchanged between these two nonidentical twins, they did not undergo rejection. Medawar concluded from these experiments that antigen delivered to the immature individual during intrauterine development or shortly after birth would result in long- term immunological tolerance. Billingham, Brent, and Medawar tested this hypothesis by injecting blood cells from mouse strain A into neonatal animals of mouse strain B. Later, when mouse strain B animals received skin grafts from mouse strain A, the grafter were accepted indefinitely, without the need for immunosuppressive therapy. The concept of inducing stable tolerance to prospective recipients of organ transplants is still the main focus of transplantation research today, almost 50 years after Medawar's initial discovery.

It became obvious during the early 1950s that kidney transplants in human beings undergo rapid and violent rejection as early as a few days after transplantation, and it is not surprising that the first successful kidney transplants were performed between identical twins. This early series of live-donor kidney transplants was performed in Boston by Murray, Hume, and Merrill. Primitive attempts to prolong kidney transplant survival between nonidentical individuals included radiation therapy and the use of 6-mercaptopurine. Later, with the addition of corticosteroids, improved survival was achieved. In the early 1960s Woodruff described the immunosuppressive effect of antilymphocyte serum. This antiserum was induced by the injection of donor strain lymphoid tissue into other species (rabbit, goat, horse). Thus an antiserum was generated that destroyed or functionally incapacitated the recipient's lymphocytes.

Once clinicians were confident that adequate immunosuppression was available, organ replacement for end-stage organ failure entered its early investigational phase. In 1963 the first liver transplant was performed by Starzl in Denver, and although this series remained largely unsuccessful, many of the technical principles that still guide liver transplantation were established. A few years later, Sir Roy Calne in Cambridge initiated the second-largest liver transplantation program in the world. Today more than 200 programs worldwide perform liver transplantation.

The first pancreas transplant was performed in 1966 in Minneapolis by Kelly, Lillehei, and Merkel. For many years, pancreas transplantation remained an unsuccessful mode of therapy for the treatment of end-stage diabetes, but in the early 1980s, with the refinement of surgical techniques and the introduction of new immunosuppressants such as cyclosporin A, the results in pancreas transplantation improved. The 1-year survival rate for simultaneous pancreas-kidney transplants in the United States is approaching 80 percent. One year after the first pancreas transplant, the first heart transplant was performed by Christiaan Barnard in Cape Town, South Africa. The first recipient survived only for a short time, but the second recipient survived for almost 2 years, and this success stimulated a worldwide surge in heart transplants. Many of these transplants, however, were performed in centers inadequately prepared, and results remained poor for several years. It is to the credit of the systematic scientific work by the Stanford group and its leader, Dr. Norman Shumway, that the foundation was laid for the current success achieved in heart transplantation.

Rejection of the transplant remained a major obstacle until the late 1970s, when a new drug, cyclosporin A, discovered by Jean Borel, revolutionized transplantation. After the introduction of cyclosporine, graft survival rose 15 to 30 percent. Shortly thereafter, the first monoclonal antibody to become commercially available for clinical use, OKT3, was tested for its effect in treating acute transplant rejection. Clinical trials led by Cosimi, from Massachusetts General Hospital, showed that this antibody was highly effective in reversing acute rejection in renal transplants, a finding that was later confirmed for heart, pancreas, and liver transplants.

The decade from 1985 to 1995 was characterized by the introduction of many new immunosuppressive agents. The original description of tacrolimus (FK506, Prograf) by Ochiai in Japan was followed by extensive clinical trials demonstrating that this agent was as potent as cyclosporine and in some instances demonstrated superiority in reducing the incidence of rejection. A new antimetabolite, mycophenolate mofetil (CellCept) was first tested at the University of Wisconsin, and other large-scale clinical trials demonstrated that this drug, in combination with cyclosporine, further reduces the incidence of rejection episodes.

Other new agents in clinical trials include rapamycin (sirolimus), mizoribine, and leflunomide. It is unclear whether these drugs will exhibit a benefit beyond the drugs that are used in current clinical practice. Monoclonal antibodies for the treatment of acute rejection or for use in induction therapy are being tested.

More than 40 years of clinical experience in organ transplantation has proved that the side effects of prolonged immunosuppressive therapy greatly contribute to morbidity and mortality after transplantation. The ultimate goal, as set out by Medawar, Brent, and Billingham, remains the induction of long-term stable immunological tolerance without prolonged immunosuppressive therapy. In recent years animal models have been established that allow scientists to test new strategies to accomplish this goal. These strategies will soon be tested in human subjects.

TRANSPLANT IMMUNOLOGY AND IMMUNOSUPPRESSION

Allan D. Kirk and Hans W. Sollinger

No field of medicine has been so intimately related to scientific advances as transplantation. Since the landmark technical developments of Alexis Carrel in 1902 that allowed vascular anastomoses to be routinely performed, it has been clear that transplantation of tissues between genetically nonidentical individuals fails. This failure has been broadly termed *rejection*, and its study has fueled most of the major discoveries in immunology. Components of the immune system are widely recognized not only for their importance in graft rejection but also for their roles in shock, tumor growth, autoimmune disease, and the systemic response to trauma. An understanding of immunology is critical to a thorough understanding of the biology of modern surgery.

Terminology

An *epitope* is the molecular unit of specific immune recognition. It is generally a carbohydrate or peptide moiety with a defined stereochemical configuration. The word *antigen* is used to describe an epitope containing molecules that can be bound by one of two types of lymphocyte receptors: the *T-cell receptor* (*TCR*) of T cells, or the *antibody* (or *immunoglobulin*) of B cells. Antigens may contain several epitopes.

Organs transplanted between genetically nonidentical individuals of the same species are termed *allografts* (the word *homograft* was used in earlier literature). The degree to which an allograft shares regulatory molecules of the immune system with the recipient is referred to as the *histocompatibility* of the graft. This is a description of the similarity of a cluster of genes on chromosome 6 known as the *major histocompatibility complex* (*MHC*, known as *human leukocyte antigen* [*HLA*] in human beings). The structure of the MHC has been described in detail. Two different classes of MHC gene products are produced, *class I* and *class II*. The importance of MHC gene products stems from their *polymorphism*. Unlike most genes, which are identical within a given species, polymorphic gene products differ in detail while conforming to the same basic structure. Since MHC is polymorphic, it can serve as an antigen to another individual. Antigens derived from different MHC molecules within the same species are called *alloantigens*. Allografts that are matched to their recipient at HLA are referred to as HLA-identical allografts, and those matched at half of the HLA loci are termed *haploidentical*. HLA-identical allografts differ genetically and are distinguishable from *isografts*. Isografts are organs transplanted between identical twins and are immunologically inconsequential. *Xenografts* are organs transplanted from one species to another (they were formerly described as heterografts).

Historical Background

The phenomenon of nontechnical loss of transplanted tissues was recognized by many investigators in the early 1900s, beginning with Carrel and Guthrie. Carrel was awarded the Nobel Prize for his initial contributions. The routine technical success of vascularized allografts was closely followed by Murphy's observation that lymphocytes were critical to the process. In the 1940s Peter Medawar's classic studies of skin-grafted rabbits demonstrated that rejection was a genetically controlled, donor-specific event mediated by lymphocytes and monocytes and governed by multiple transplant-related antigens. He showed that immunity could be invoked by prior exposure to donor tissues other than the transplanted tissue, that the immunity was remembered by the host immune system, and that active cell division was required for host response amplification. Building on the observations of Owen, Medawar and his colleagues also demonstrated that transplant immunity was acquired during ontogeny rather than being an innate property of the host, and

that the barrier to transplantation was not absolute. Like Carrel's, Medawar's advance was recognized by the Nobel Committee. Mitchison further demonstrated the importance of lymphocytes in the rejection process by transferring transplant immunity with leukocyte transfusion. In a period of 10 years, the problem was defined, the effector identified, and the solution envisioned. These investigations heralded the beginning of modern transplant immunology and have served as the basis for all subsequent work.

The next two decades were marked by extraordinary progress in the understanding of the genetic basis of immune recognition. Landmark investigations on the immunogenetics of the MHC began with Gorer and Snell's description of the H-2 system in mice, a genetic locus that segregated with transplant tumor survival. This erythrocyte-based system was soon shown by Amos to exist on leukocytes and to elicit an antibody response toward the antigens encoded by H-2. This antibody response became critical to establishing the correlate of H-2 in human beings. Dausset used this finding, together with his observation that antibodies could also be generated by antigen exposure via transfusion or pregnancy, to establish the first serologically based typing system for human transplant antigens, the Mac antigens. This system used serum from multiparous females to lyse cells based on differences in the cell surface antigens of lymphocytes. These differences were subsequently shown to be based on polymorphisms of the class I (HLA-B) molecules. This system was followed quickly by the 4a4b system of van Rood and the LA system of Payne. Snell and Dausset shared a Nobel Prize for their observations, perpetuating the pattern of fundamental discoveries in transplant immunobiology.

Class II MHC polymorphism also was defined in this period. Bach demonstrated that lymphocytes would proliferate in response to contact with lymphocytes of a different genetic background. This phenomenon, known as the *mixed lymphocyte culture (MLC)*, was found to be based on differences in class II MHC molecules and was defined as HLA-D.

These findings set the stage for a series of workshops from 1964 to 1970 organized to integrate the growing number of systems into a single, unified system. Stimulated by the increasingly rapid pace in defining HLA polymorphism, methods for rapid serologic definition of HLA were developed and broadly applied, primarily Terasaki's lymphocytotoxicity assay and refinements of Bach's MLC. This resulted in the current definition of HLA as six loci on chromosome 6, encoding two related cell surface molecules: class I (HLA-A, B, and C) and class II (HLA-DR, DP, and DQ).

In the 1970s advances in molecular biologic techniques allowed the genetic basis of HLA polymorphism to be established, leading to Bjorkman and Brown's discovery of the three-dimensional structure of HLA antigens. These methods allowed the fundamental aspects of antigen recognition to be defined. Particularly important were the Nobel Prize-winning contributions of Tonegawa that showed that antigen receptors were generated by complex rearrangements of the somatic genes in T and B cells. These studies have allowed further examination of the fundamental nature of neonatal tolerance first described by Medawar.

The characterization of HLA catalyzed numerous fundamental investigations in immunology in the 1980s. The function and structure of the T-cell receptor, the importance of adhesion molecules and costimulatory signaling in T-cell activation, and the elucidation of the soluble mediators of cell activation, cytokines, were products of this period. They have led to a reasonably comprehensive picture of the mechanism of allograft rejection and the physiologic functions of lymphocytes.

The close link between the basic scientific community and pioneer transplant clinicians has been underscored by the rapid application of new discoveries to the problem of rejection. Clinical observations have aided greatly in understanding the importance of basic observations. After the genetic basis for transplant rejection was established, Murray applied the concept by successfully transplanting a kidney between identical twins. The importance of lymphocyte division also was quickly recognized and exploited. Elion and Hitchings developed a 6-mercaptopurine analogue, azathioprine, which inhibited lymphocyte nucleic acid metabolism. It was applied first, experimentally, by Calne and Murray, and then, clinically, by Murray to allow the first successful allografts in human subjects. Elion and Hitchings were honored by the Nobel Committee in 1988 and Murray in 1990.

The most significant advance in immunosuppression was the discovery of cyclosporine by Borel in 1976. This selective inhibitor of T-cell activation was quickly recognized by White and Calne and applied to renal transplantation. The release of this drug in 1983 made clinical transplantation routinely successful not only for kidneys but for other organs as well. The success of cyclosporine has benefited T-cell biology too, because it stimulated much of the investigation necessary for understanding cytokine function and overall transmembrane receptor signaling.

Today's armamentarium to combat rejection is large and varied. It is now rare for an allograft to be lost to acute rejection. Future advances must focus on the chronic diseases of a growing transplant population. Ideally, the concept of specific immunological tolerance first described 50 years ago will soon be understood and will allow for transplantation without immunosuppression.

Overview of Immunity

Immunity in vertebrates has evolved into two distinct, but complementary, branches to combat disease: innate and acquired immunity. These two types of immunity differ in the degree of specificity with which their receptors or proteins bind to pathologic agents. Innate systems use germ line-encoded proteins that are limited in specificity but are broadly reactive against common components of pathogenic organisms, e.g., lipopolysaccharides on gram-negative organisms or other glycoconjugates. Acquired immunity is based on antigen receptors formed by germline gene rearrangement, which leads to highly specific binding interactions, primarily the activity of B and T cells. These rearrangements occur during ontogeny to ensure that the receptors released into the periphery are adapted to the MHC molecules of each individual. MHC polymorphism leads to a receptor repertoire that is unique to each individual. Its evolution has ensured that the population is adapted to deal with an extraordinary array of pathogens and has minimized the possibility of an all-encompassing plague.

The hallmark of acquired immunity is *specific* recognition and elimination of nonself. Highly specialized mechanisms for distinguishing pathogenic organisms and transformed cells from normal tissues have evolved to facilitate this goal. The pathogen is recognized as a specific entity, not just as nonself, and a record of that encounter is retained for more rapid response to future encounters, a phenomenon known as immunological memory. Antigens are peptides, or carbohydrates, containing epitopes, and they are the portion of a pathogenic cell or foreign body that is recognized by the acquired immune system. Antigen recognition is mediated by lymphocytes, T cells, and B cells, each of which is endowed with receptors for detection of specific antigens. T cells protect the cells of the body against alterations by mutation or viral infection (cellular immunity) and bind peptide antigens that have been processed by the body's cells. B cells provide protection against extracellular infectious organisms and foreign material (humoral immunity) and recognize antigens in their native unprocessed state. Both B and T cells rearrange the germ line configuration of their genome during ontogeny to form as many as 10^{11} clones of cells, each dedicated to the recognition of one specific antigen. An individual's ability to recognize a given antigen is determined before birth.

T cells use a specific receptor, the T-cell receptor (TCR), to recognize processed peptide antigens bound to MHC molecules. Self cells "contaminated" with peptide antigens are identified for elimination. This binding event is strengthened by *adhesion molecules* that improve the binding affinity and communicate information about the nature of the binding to the T cell. Parenchymal cells express class I MHC molecules. These class I molecules display peptides from within, e.g., peptides from normal cellular processes or from viral replication, which are bound by T cells expressing an adhesion molecule with special affinity to class I, the CD8 molecule. Hematopoietic cells also express class II MHC molecules. These molecules display peptides that have been phagocytized from surrounding extracellular spaces and bind to T cells complemented by an adhesion molecule with affinity for class II-CD4. Under physiologic conditions, CD4+ T cells are first alerted of an invasion of the body by *antigen-presenting cells (APCs)* and activate CD8+ T cells to search the body for cells that have been infected by this invader. Many other adhesion molecules control the movement of immune cells through the body, monitor their trafficking to specific areas of inflammation, and nonspecifically strengthen the TCR:MHC binding interaction.

B cells bind soluble antigens and secrete soluble forms of their receptor, known as *antibodies*, to bind these foreign molecules. Material that is bound by an antibody is opsonized (flagged) for destruction by cells of the innate arm of immunity—phagocytic cells lacking the ability to distinguish self from nonself—primarily macrophages, monocytes, and *polymorphonuclear leukocytes (PMNs)*. This also improves the efficiency by which APCs engulf and present pathogens to T cells. Antibody-bound surfaces activate a destructive enzymatic cascade known as the *complement* system. This leads to destruction of the membrane to which the complement is bound and further opsonization for APC uptake. Specific immunity is complemented by innate defense mechanisms that, once activated, destroy in a more indiscriminate manner.

The entire immune process is facilitated by a means of amplifying the response of one cell to one antigen. *Cytokines* (known as interleukins [IL]) (see Table [10-1](#)) are polypeptides that are released by many cell types and activate or suppress adjacent immune cells. Other soluble chemicals released during an immune response increase blood flow to the area and improve the exposure of the area to lymphocytes and the innate immune system. The prototypical cytokine of T-cell activation is IL-2.

While each component of the immune system has a defined role in the defense against disease, the immune response to an allograft is multifaceted and complex. This is reflective of its serendipitous nature. There has been no selective pressure to evolve a means of counteracting engrafted tissue. The immune response to an allograft is the result of incompatibility between the recipient's receptor repertoire and the donor's MHC polymorphisms. Effector mechanisms that have evolved to counteract viral, fungal, and bacterial infection, as well as those in place to prevent malignancy and autoimmunity, all come into play after transplantation. The innate and acquired arms of the immune system are involved in rejection, and the process of counteracting rejection limits the defense against other pathogenic entities.

Rejection, like physiologic immunity, can be divided into humoral and cellular mechanisms. Humoral rejection of a graft can be the result of antibodies existing in circulation prior to exposure, or antibodies acquired following exposure. This type of rejection involves the binding of antibody specific to the donor tissue MHC discrepancies, and subsequent activation of the complement cascade and cell lysis. Cellular rejection is the result of T-cell incompatibility between the donor and recipient. T cells recognize the MHC molecules of the donor themselves (either intact or processed and presented by APCs) as antigens, and identify the cells as contaminated and worthy of elimination. These two categories of rejection are not mutually exclusive.

Genetic and Structural Characteristics of Transplant Antigens

The antigens primarily responsible for human allograft rejection are those encoded by the HLA region of chromosome 6 (Fig. [10-1](#)). The polymorphic proteins encoded by this locus that directly affect transplant rejection are class I molecules (HLA-A, B, and C) and class II molecules (HLA-DR, DP, and DQ). Class I genes with limited polymorphism (E, F, G, H, and J) are not currently typed. Other genes encoded by HLA are the tumor necrosis factors a and b, components of the complement cascade (class III molecules), the heat-shock protein HSP 70, and genes necessary for class I and class II presentation of peptides to the body's T cells (peptide transporter proteins TAP 1 and TAP 2, and proteasome proteases LMP 2 and LMP 7). While other polymorphic genes, referred to as *minor histocompatibility antigens*, exist in the genome, these antigens are more important in bone marrow transplantation than in solid organ transplantation. Even HLA-identical individuals are subject to rejection on the basis of these minor differences. The blood group antigens of the ABO system also must be considered polymorphic transplant antigens, and their biology is critical to humoral rejection.

Each class I molecule is encoded by a single polymorphic gene that is combined with the nonpolymorphic protein b₂-microglobulin (b₂M, from chromosome 15) for expression. The polymorphism of each class I molecule is extreme, with 30 to 50 alleles per locus. Class II molecules are made up of two chains, a and b, and individuals differ not only in the alleles represented at each locus but also in the number of loci present in the HLA class II region. The polymorphism of class II is increased by combinations of a and b chains and a hybrid assembly of chains from one class II locus to another. As the HLA sequence varies, the ability of various peptides to bind to the molecule and be presented for T-cell recognition changes. This extreme diversity is thought to improve the likelihood that a given pathogenic peptide will fit into the binding site of these antigen-presenting molecules, preventing a

single viral agent from evading detection by T cells of an entire population. The importance of class I and class II structure is underscored by the relationship of HLA allotypes to many viral and autoimmune diseases.

While the structure of HLA is becoming increasingly complex, the clinical importance of the region is easily understood by simple Mendelian genetics. Recombination within the locus is uncommon, occurring in approximately 1 percent of molecules, and therefore the HLA type of the offspring is predictable. The unit of inheritance is the haplotype, which consists of one chromosome 6, and one copy of each class I and class II locus (HLA-A, B, C, DR, DP, and DQ). The genetics of HLA are particularly important in understanding clinical living-related donor (LRD) transplantation. Each child inherits one haplotype from each parent; the probability of a sibling's being HLA-identical is 25 percent. Haploidentical siblings occur 50 percent of the time and completely nonidentical siblings 25 percent of the time. Parents are haploidentical with their children. The degree of HLA match also can improve if the parents are homozygous for a given allele, giving the same allele to all children. If the parents share the same allele, the likelihood of that allele's being inherited improves to 50 percent. The inheritance of HLA also is affected by linkage disequilibrium, or a propensity of certain class I and class II genes to be located on the same chromosome.

The three-dimensional structure of class I molecules (HLA-A, B, and C) was first elucidated in 1987 and is shown in Fig. 10-2. Class I is expressed as a single MHC-encoded, transmembrane alpha chain, in combination with β_2 M. The alpha chain has 3 domains, alpha 1, 2, and 3. The critical structural feature of class I molecules is the presence of a groove formed by two α helices mounted on a β -pleated sheet in the alpha 1 and alpha 2 domains. Within this groove, a nine-amino acid peptide, formed from fragments of proteins being synthesized in the cell's endoplasmic reticulum, is mounted for presentation to the body's T cells. Almost all of the significant sequence polymorphism of class I is located in the region of the peptide binding groove and in areas that directly contact T cells. It is this variation in sequence at the HLA:TCR interface that is the essence of alloreactivity. Class I molecules are found on all nucleated cells except neurons.

The three-dimensional structure of class II molecules (HLA-DR, DP, and DQ) was inferred by sequence homology to class I and was eventually proved by x-ray crystallography. The structural features of class II molecules are strikingly similar to those of class I molecules (see Fig. 10-2). In particular, the groove and platform for peptide binding are present and are held away from the cell surface by two subunits. Class I is a single chain of three domains that relies on a non-MHC-encoded subunit for expression (β_2 M); class II is formed by two MHC-encoded chains, alpha and beta, each with two domains. The peptide binding site for class II is formed by the alpha 1 and beta 1 domains. This site is filled with a peptide derived from endocytosed proteins (as opposed to proteins formed by the cell, as is the case for the groove of class I). These peptide fragments are the result of proteolytic degradation of captured extracellular material and are substantially larger (up to 23 amino acids long) than the peptide found in class I. Like class I, the sequence polymorphism of class II is located at this TCR interface region. Class II molecules are found primarily on cells of the innate immune system, particularly phagocytes, such as dendritic cells, macrophages, and monocytes, but can be up-regulated to appear on other parenchymal cells by cytokines released during an immune response or injury.

Another important sequence-related difference between class I and class II molecules is located on the supporting subunits (alpha 3 domain for class I and beta 2 domain for class II). A loop extends outward and serves as a binding site for accessory molecules associated with the T-cell receptor. The TCR accessory molecule CD8 selectively binds the loop of class I while the accessory molecule CD4 binds the loop of class II. In this way T cells geared toward the initial recognition of intruders and subsequent amplification of the immune response (CD4+ helper T cells) are targeted to bind the cells with the ability to capture and present these antigens. Similarly, T cells that survey the body's parenchyma for signs of entrenched intracellular pathogens and destroy infected cells (CD8+ cytotoxic T cells) are outfitted to perform this duty.

Biology of Transplant Antigens

The physiologic role of MHC molecules is twofold: to provide a mechanism for T-cell inspection of parenchymal cells, and to provide an interface between innate immune cells (APCs) and T cells. For the structural reasons detailed above, class I molecules serve the first role and class II molecules serve the second. Organ transplantation is not a physiologic process but rather an artificial situation. T-cell responses to either class of MHC molecule can generate a rejection episode.

The assembly of class I is dependent on association of the alpha chain with β_2 M and occupation of the peptide groove with a native peptide. Incomplete molecules are not expressed. All peptides made by a cell are candidates for presentation, although sequence alterations in this region favor certain sequences over others. Human class I presentation occurs on all nucleated cells in contact with blood, allowing the acquired immune system to inspect and approve of ongoing protein synthesis. T cells develop on the basis of their ability to bind without activation to self MHC molecules presenting self peptides. Alterations in the peptide content, for example, by introduction of viral peptide synthesis, cause activation. In the case of transplantation, this alteration is possible not only with the peptide but also with the presenting molecule itself. T cells interpret any binding event not established as appropriate during their development as worthy of presenting cell destruction. Class I molecules are bound only by T cells expressing CD8, and CD8+ T cells have enhanced cytotoxic capabilities for infected cell destruction.

Class II molecule assembly requires association of an alpha chain and a beta chain in combination with a temporary protein called the *invariant chain*. This third protein covers the peptide binding groove until the class II molecule is out of the endoplasmic reticulum and is sequestered in a forming endosome. Proteins that are engulfed by a phagocytic cell are degraded at the same time as the invariant chain is removed, allowing peptides of external sources to be associated with, and presented by, class II. In this way the acquired immune system can inspect and approve of proteins that are in free circulation. The T cells that bind class II molecules are CD4+ and have enhanced abilities for activation of CD8+ T cells and antibody-producing B cells. When an inappropriate peptide is detected, CD4+ T cells release cytokines to recruit CD8+ cells into the area to inspect nearby cells for intracytoplasmic presence of the offending peptide. B cells also are stimulated to release antibody to bind the offending peptide and aid in its clearance by the innate immune system. The cytokines released by CD4+ cells, particularly interferon- γ , also induce expression of class II molecules on local cells and increase expression of class I molecules locally. This increases the chances that infected cells will be detected. In the case of transplanted organs, as with class I, an abnormal peptide or the foreign class II molecule itself can lead to T-cell activation. Ischemic injury at the time of transplantation accentuates the potential for T-cell activation by foreign MHC by causing up-regulation of both class I and class II molecules.

While matching donors and recipients with regard to HLA type has been shown to improve outcome after kidney, heart, and pancreas transplantation, no such correlation exists in liver transplantation. Matching may reduce overall survival. The reasons for this lie in the dualistic nature of HLA in the pathophysiology of liver disease. T-cell-mediated rejection of the liver is mechanistically the same as with other organs; therefore, rejection is reduced with improved HLA compatibility. The physiologic role of HLA is the presentation of viral peptides to T cells to initiate destruction of virally infected cells. HLA compatibility potentiates the inflammation during viral reinfection after transplantation for viral hepatitis and increases the chances for clinical recurrence of the original disease. Similarly, T-cell-mediated autoimmune diseases, e.g., primary biliary cirrhosis, are etiologically based on T-cell recognition of HLA-presented peptides. Recurrence of autoimmune diseases may be potentiated as well. Further knowledge regarding specific disease states worsened by certain HLA matches may be useful for selective typing in the future.

Clinical Definition of Transplant Antigens

There is no doubt that matching the recipient and the donor as closely as possible with regard to their MHC genes reduces the risk of acute rejection. However, given the time constraints, pretransplant matching is feasible only in cadaveric renal transplantation and living-related allografts. Other organs are MHC-typed retrospectively. As immunosuppression has improved, the relative importance of MHC matching, even for renal allotransplantation, has decreased. When determining the destination of an organ, significant emphasis is placed on the recipient's physical condition and time on the waiting list. The degree of match, however, should be considered when allocating renal allografts and when evaluating organ dysfunction after allotransplantation.

Historically, MHC polymorphism has been defined with the use of two biologic assays: the lymphocytotoxicity assay and the MLC. Both assays define MHC epitopes but do not comprehensively define the entire antigen or the exact genetic disparity involved. Techniques now exist for precise genotyping that distinguishes the nucleotide sequence of an individual's MHC.

The lymphocytotoxicity assay is performed by taking serum from donors with anti-MHC antibodies of known specificity, and mixing it with lymphocytes from the individual in question. Rabbit complement is added with a vital dye, such as trypan blue, which is not taken up by cells with intact cell membranes. If the antibody binds to MHC, it activates the complement, leads to cell membrane disruption, and stains the cell with the dye. Microscopic examination of the cells can determine if the MHC antigen exists on the cells. This assay has been extremely valuable, but it is limited by the nature of the reagents. Antibodies bind to epitopes. The presence of one epitope does not preclude the presence of other epitopes that may differ from the desired MHC type. Many antibodies *cross-react* with MHC antigens other than the one with which they were raised. The pattern of these cross-reactivities is reasonably well established, and the use of a large panel of antibodies allows reasonable definition of the genetic locus in question.

The MLC is performed by incubating recipient T cells with irradiated donor T cells in the presence of ^{3} H-thymidine. If the cells differ at the class II MHC locus, the recipient cells proliferate and incorporate the radionuclide into the new cells. This incorporation can be detected and quantified. While class II polymorphism is detected by this assay, it takes several days to complete one assay (unlike the lymphocytotoxicity assay, which takes 4 to 6 h). The use of MLC as a prospective typing assay is limited to living-related donors. The antigen receptor is again the biologic reagent of this assay, and the genetic basis for the reaction can only be inferred from a series of reactions.

The sequencing of the class I and class II HLA loci has allowed several genetics-based techniques to be used for histocompatibility testing. These include restriction fragment length polymorphism (RFLP), oligonucleotide hybridization, and polymorphism-specific amplification using the polymerase chain reaction and sequence specific primers (PCR-SSP). Of these methods, the PCR-SSP technique is most commonly employed for class II typing. Serologic techniques are the predominant method for class I typing because of the complexity of class I sequence polymorphism. Sequence polymorphisms that do not alter the TCR:MHC interface are unlikely to affect allograft survival; the enhanced precision of molecular typing may provide more information than is clinically relevant.

Genetic and Structural Characteristics of Transplant Antigen Receptors

Receptors

Two cell types have evolved with the ability to specifically bind to antigen: T cells and B cells. Their receptors are similar in genetic development but differ in the types of antigens bound. T-cell antigen receptors bind peptide antigens that have been processed by cells and combined with MHC molecules, and B-cell antigen receptors, also known as antibodies, bind antigens in their native conformation free in the extracellular fluid. T-cell receptors are fixed; B-cell receptors can be secreted and act at locations remote from the cell.

T-Cell Receptor

The formation of the TCR is fundamental to the understanding of alloreactivity and self nonreactivity. T cells are formed in the fetal liver and bone marrow and migrate to the thymus during the first trimester of fetal development. At this stage, they have no TCR or accessory molecules. Upon entering the thymus, T cells undergo a remarkable rearrangement of the DNA that encodes the TCR (Fig. 10-3). This series of genetic deletions and splicing events is driven by two recombination activating genes, RAG-1 and RAG-2. Four distinct loci (α and δ on chromosome 14, and β and γ on chromosome 7), each made up of a highly polymorphic variable (V), junctional (J), and/or diversity (D) region and a well-conserved constant (C) region, are involved. The recombination event randomly joins C, V, D, and J regions together to form a functional α , β , γ , or δ chain. The γ and δ loci recombine first,

and if recombination is successful, a gd TCR is formed. If this event is not successful, then the α and β regions recombine to form an ab TCR. Approximately 95 percent of cells progress to express an ab TCR. The order of genetic rearrangement recapitulates the evolution of the TCR, as T cells expressing the gd TCR have more primitive functions, including recognition of heat-shock proteins, activity similar to natural killer (NK) cells, and MHC recognition.

Regardless of the genes used, individual cells recombine to express a single TCR with a single specificity. The rearrangements occur randomly in all T cells, resulting in a population of T cells capable of binding 10⁶ different specificities—essentially all combinations of MHC and peptide. These developing T cells also express CD4 and CD8, increasing the binding repertoire of the population to include either class I or class II MHC molecules. If this population were released unmodified, they would quickly bind to all self cells, activate, and destroy the individual.

To avoid the release of self-reactive T cells, developing cells undergo a process after recombination known as *thymic selection*. Cells initially interact with the MHC-expressing cortical thymic epithelium. If binding does not occur to self MHC, the cells are useless to the individual because they would be unable to bind and function in the periphery. All nonbinding cells undergo *apoptosis*, or programmed self-destruction, a process called *positive selection*. Cells surviving positive selection then move to the thymic medulla and lose either CD4 or CD8. If binding to self MHC in the medulla occurs with an unacceptably high affinity and apoptosis results, this is called *negative selection*. The precise nature of this affinity threshold most likely involves interaction with hematopoietic cells that reside in the thymus and simulate APCs. The only cells released into the periphery are those that can bind self MHC without activation. Any foreign peptide encountered alters the affinity that has been preordained in the thymus, resulting in T-cell activation. MHC molecules that were not part of the T cell's thymic education will bind the TCR with unacceptable affinity and lead to activation. This phenomenon defines alloreactivity.

In addition to thymic selection, mechanisms exist for peripheral modification of the T-cell repertoire. Much of this is in place for removal of T cells after an immune response and down-regulation of activated clones. A molecule known as Fas (CD98) is expressed on activated T cells. Under appropriate conditions, binding of this antigen to its ligand leads to apoptosis. This method is dependent on TCR binding and the activation state of the T cell. The Fas ligand, in addition to its role in down-regulation, can serve as a molecular barrier to T-cell invasion of certain immunologically privileged sites, e.g., testes. Complementing this deletional method to TCR repertoire control are nondeletional mechanisms that selectively anergize (make unreactive) specific T-cell clones. One prominent receptor group mediating this function is the CD28:B7 pair. TCR binding leads to activation only if the costimulatory molecule B7 is bound to its ligand CD28, generally found on APCs. In the absence of binding, the cell is turned off until exogenous IL-2 is added. TCR binding that occurs to self in the absence of appropriate antigen presentation or active inflammation fails to lead to self reactivity.

The manipulation of the TCR repertoire by central (thymic) and peripheral mechanisms is responsible for the phenomenon described by Medawar as *tolerance*. Its understanding will yield the methods required for specific allograft acceptance.

The end result of TCR formation is a heterodimeric transmembrane receptor with a site for binding to peptide: MHC combination. The receptor is combined with an accessory-binding molecule (CD4 or CD8) and a transmembrane-signaling complex known as CD3.

Antibody

Antibody, also called immunoglobulin (Ig), is formed in B cells much the way the TCR is in T cells, although maturation occurs in the bone marrow, not in the thymus, and continues in the periphery. Five different heavy-chain loci (μ , γ , α , ϵ , δ) on chromosome 14 and two light-chain loci (κ on chromosome 2 and λ on chromosome 2), each with V, D, J, and C regions, are brought together randomly by the RAG-1 and RAG-2 apparatus to form a functional antigen receptor. Antibodies have a basic structure of four chains, two of which are identical heavy chains, and two of which are identical light chains. The heavy-chain usage defines the Ig type as being either IgM, IgG, IgA, IgE, or IgD. This structure forms two identical antigen binding sites brought together on a common region known as the Fc portion of the antibody (Fig. 10-4). The Fc portion is bound by Fc receptors on phagocytic cells of the innate immune system, facilitating phagocytosis, followed by destruction of the antigen and processing of antigenic peptides. The Fc portion of IgM, and some classes of IgG, also serve to activate complement. A mechanism for regulating B-cell tolerance exists, although our understanding of it is less complete than that for T-cell tolerance.

Unlike the TCR, the Ig loci undergo continued alteration after B-cell stimulation to improve the affinity and functionality of the secreted antibody. In an alteration known as isotype switching, Ig genes change their initial heavy-chain gene usage from the IgM type (used for initial and baseline responses against common carbohydrate antigens), to one of four types, each providing heightened specialization for a given purpose. IgG becomes the most significant soluble mediator of opsonization and is the dominant antibody resulting from allostimulation. IgA is formed for mucosal immune responses, IgE for mast-cell mediated immunity, and IgD as a primary cell-bound antibody form. Once a clone of cells is activated, the D and J regions of the utilized gene undergo random additions through the action of terminal deoxynucleotide transferase (TdT), to slightly altering the gene coding for the antigen binding site. This process, called affinity maturation, results in clones that have altered antigen affinity. Those with increased antigen affinity are retained for a more vigorous response in the event that this antigen is reencountered. Individuals with prior exposure to an alloantigen are more likely to have clones of B cells that have mutated to form a gene complex expressing an IgG with extremely high affinity. To avoid a vigorous humoral rejection of the graft, screening for these antibodies must be done before transplantation.

Biology of Transplant Antigen Recognition and Destruction

T-Cell Activation

T cells can respond to transplant antigens directly, through TCR binding to foreign MHC molecules expressed on transplanted tissues, or indirectly, by encountering APCs that have phagocytosed fragmented allograft tissues and processed the antigens for expression on self MHC. Regardless of the source of the activating MHC, the ensuing activation of the cell proceeds in the same way as it does for physiologic T-cell function. This process has been exploited at almost every critical step to prevent acute rejection, and its understanding is key to the rational use of immunopharmacologic agents.

Initial T-cell binding to an APC or endothelial cell is nonspecific and mediated by adhesion molecules. These molecules, including ICAM-1, VCAM-1, LFA-1, and other integrin family molecules are all up-regulated upon APC activation. This increases the likelihood that a T cell with pertinent specificity will bind and become activated. Once nonspecific adhesion has formed, MHC recognition can occur.

Because the number of potential antigens is high and self antigens are likely to vary minimally from foreign antigens, the nature of the TCR binding event has evolved such that a single interaction with an MHC molecule is not sufficient to cause activation. A T cell must register a signal from approximately 8,000 TCR ligand interactions with the same antigen before a threshold of activation is reached. Resting T cells have low TCR density, which usually is accomplished by sequential binding and internalization over several hours. Transient encounters are not sufficient.

The TCR transmits its signal to the cell by initiating the activity of intracytoplasmic protein tyrosine kinases (PTKs). These PTKs include p56^{lck} (on CD4 or CD8), p59^{Fyn}, and ZAP70, the latter two of which are associated with the TCR-associated transmembrane protein complex called CD3 (Fig. 10-5). Repetitive binding signals eventually activate phosphokinase- γ (PLC- γ 1), which in turn hydrolyses the membrane lipid phosphatidyl inositol biphosphate (PIP₂), releasing inositol triphosphate (IP₃) and diacylglycerol (DAG). IP₃ binds to the endoplasmic reticulum, causing a release of calcium that induces calmodulin to bind to and activate calcineurin. Calcineurin dephosphorylates the critical cytokine transcription factor NF- κ B, prompting it to initiate transcription of IL-2. IL-2 is then released and binds to the T cell in an autocrine loop, potentiating DAG activation of protein kinase C (PKC). PKC is important in activating many gene regulatory steps critical for cell division.

In addition to TCR engagement, a second confirmatory signal is required for T-cell activation. This is most likely the role of the T-cell surface molecule CD28. It binds to molecules on APCs known as B7 molecules (B7-1 or CD80 and B7-2 or CD86). Binding of CD28 directly potentiates the TCR-initiated tyrosine phosphorylation, allowing more efficient signal transduction and lowering the number of binding events required for T-cell activation from 8,000 to about 1,500. B7 expression appears to distinguish MHC-expressing cells with potent APC function (so-called “professional” APCs, such as dendritic cells and macrophages) from cells with poor ability to present antigen. A second receptor, CTLA-4, acts to negatively regulate CD28, while the surface molecule gp39 binds to CD40 on B cells and macrophages to up-regulate B7 and potentiate costimulation. Costimulation can control whether a TCR signal results in activation or quiescence. This control plays a role in self-tolerance.

Activation also is dependent on the presence of a “danger” signal. By requiring an indication that the binding event is occurring in a location undergoing injury, spontaneous T-cell activation is prevented. This signal may be given by dendritic cells activated by phagocytosis of injured or opsonized material and likely is closely tied to the CD28/CTLA-4 costimulation pathway.

T-Cell Amplification

Once activation occurs, cytokines, particularly IL-2 and interferon- γ (IFN- γ), create a potent milieu, recruiting other T cells into the response and potentiating clonal expansion. Cytokines, particularly IFN- γ , also induce the nonlymphoid tissues to express class II molecules and to up-regulate expression of class I molecules. B-cell activation also is mediated through cytokine secretion. Cytokines are responsible for many of the systemic symptoms of fever and malaise associated with severe graft rejection.

T cells, once activated, develop into one of two phenotypes based on cytokine expression. T cells mediating cytotoxic responses, such as delayed-type hypersensitivity (DTH), express IL-2, IL-12, IL-15, and IFN- γ and have been called Th1 cells. T cells supporting the development of humoral or eosinophilic responses express IL-4, IL-5, IL-10, and IL-13 and have been called Th2 cells. This dichotomy is supported by the fact that many Th1 cytokines have a common receptor chain called γ that is not used by Th2-type cytokines. While extensive study has shown that differential cytokine expression alters the character of the effector response on a population scale, the predominance of a given cytokine is based on skewed expression along a continuum rather than absolute expansion of a single cell type. This is particularly true in the presence of immunosuppressive agents. Although evolved physiologic responses to pathogens have been shown to align themselves into discrete cytokine profiles, this division is less valid during transplant rejection. Immunosuppression produces artificial patterns of gene expression, resulting in nonphysiologic patterns of cytokine secretion. Responses also vary on the basis of the organ targeted for rejection. Kidneys are infiltrated by T cells in the context of a Th1-type milieu plus IL-10, but livers have a significant eosinophil infiltration and a large amount of IL-5. The patterns of inflammation seen during allograft rejection result from cytokine-mediated amplification but vary considerably according to other factors.

In the late phases of rejection, the inflammatory response recruits cells with nonspecific cytotoxic activity to the organ. T cells expressing the gd TCR and other T lineage cells, such as lymphokine-activated killer cells, are activated to destroy surrounding tissue in an MHC-unrestricted fashion. IL-8, released by activated macrophages and T cells, also recruits PMNs to the scene to remove necrotic tissue.

T-Cell-Mediated Cytotoxicity

T cells assume one of two roles: that of an amplifier or that of a cytotoxic effector. The amplification role generally is performed by CD4⁺ cells, because these cells are most suited for communication with class II-expressing APCs. Cytotoxicity is best mediated by CD8⁺ cells, since they bind to the MHC of all nucleated cells. In the artificial situation presented by organ transplantation, both CD4⁺ and CD8⁺ cells can mediate cytotoxicity. This occurs by a Ca²⁺-dependent secretory mechanism or a Ca²⁺-independent binding mechanism. The Ca²⁺ flux that occurs with activation also causes the exocytosis of cytolytic granules. These granules contain a lytic protein called *perforin* and serine proteases called

granzymes. Perforin polymerization in the presence of extracellular Ca^{2+} forms defects in the target cell's membrane, allowing granzyme activity to lyse the cell. In the absence of Ca^{2+} , T cells can induce apoptosis of a target cell. Apoptosis is a programmed death that involves fragmentation of the nuclear contents. It occurs when a surface Fas is cross-linked by its ligand. Cytotoxic T cells up-regulate Fas ligand upon activation.

After activation, non-MHC-restricted, T-cell-mediated cytotoxicity occurs. This allows the utilization of T cells activated in the area of inflammation without a TCR specific to the antigen to take part in protective, or in the case of transplant rejection, detrimental immunity. The mechanisms of direct T-cell-mediated destruction of microorganisms or non-MHC-expressing tissue remain poorly defined. T cells expressing the gd TCR are prominent effectors of non-MHC-restricted cytotoxicity. This type of activity is induced by high concentrations of IL-2, which may be relevant in the local milieu of a rejecting allograft. This type of killing may be involved in rejections precipitated by infections or trauma to the transplanted organ and is certainly elicited during vigorous MHC-directed rejection.

B-Cell Activation and Clonal Expansion

B cells recognize antigen in its native form without the requirement for processing and presentation on MHC molecules. Surface antibody cross-linking by antigen leads to B-cell proliferation and differentiation into a plasma cell. Like the T cell, the threshold for B-cell activation is high. This can be lowered by a factor of 100 by costimulation signals received by the transmembrane complex CD19/CD21. B cells also can internalize antigens bound to surface antibody and process them for presentation to T cells. They can receive signals from T cells via CD40 mediated binding to the T cell's gp39. This signal up-regulates expression of B7 molecules on B cells and facilitates antigen presentation and T-cell costimulation. As such, B cells can bind antigen in circulation and initiate a T-cell response to deal with antigen incorporated into tissues of the body. Plasma cells (activated B cells) are distinguished histologically by their hypertrophied Golgi apparatus. They secrete large amounts of monoclonal antibody (antibody with a single specificity). During the activation process, the specificity of the antibody is altered by affinity maturation, as detailed previously.

In addition to being secreted after exposure to an antigen, antibody can be present as part of a natural repertoire in circulation for initial response to common pathogens. Antigen exposure generally leads to B-cell affinity maturation and isotype switching, and produces high-affinity IgG antibodies. Naturally occurring antibodies are IgM antibodies with low affinity and are generally thought to respond to a broad array of carbohydrate epitopes found on many common bacterial pathogens. Natural antibody is responsible for ABO antigen responses and discordant xenograft rejection.

Antibody-Mediated Cytotoxicity

Antibody facilitates the destruction and removal of antigenic cells. Once bound to an antigen, antibody serves as an anchoring site for the complement component C1q. In a pathway known as the *classical complement activation cascade*, two antibody Fc portions bind and initiate C1q-mediated activation of C3, the central activating enzyme of the complement cascade. This leads to formation of the membrane attack complex (MAC) of polymerized C5, C6, C7, C8, and C9, resulting in disruption of the target's cell membrane and lysis. In addition, several by-products of C3 cleavage serve as chemoattractant agents for phagocytic cells and as opsonins potentiating antigen phagocytosis. Antibody also can serve as an opsonin directly. Most phagocytic cells have receptors for the Fc portion of IgG and actively engulf antibody-coated targets in a process known as *antibody-dependent cellular cytotoxicity* (ADCC).

Antibody binding to the endothelium, and the subsequent activation of complement, also alters the activation status of the endothelial cell. This leads to cellular retraction and exposure of the underlying matrix, which in turn potentiates platelet activation and aggregation. Endothelial activation also alters its usually anticoagulant environment in favor of a procoagulant one. Heparin sulfate is shed, as is thrombomodulin. This prevents thrombomodulin-mediated activation of protein C and the interaction of activated protein C with protein S. The result is microvascular thrombosis, a hallmark of the two antibody-mediated graft rejections: hyperacute rejection and acute vascular rejection.

Clinical Rejection Syndromes

Rejection has been classified as hyperacute, acute, or chronic on the basis of the general pathophysiology and effector arm of the immune system involved. Of these, only acute rejection can be successfully reversed. Though hyperacute rejection is untreatable, it is mostly a preventable phenomenon. Chronic rejection remains a difficult problem.

Hyperacute Rejection

Hyperacute rejection (HAR) is caused by presensitization of the recipient to an antigen expressed by the donor. It develops in the first minutes to hours following graft reperfusion. Antibodies in circulation prior to transplantation, the result of prior exposure to donor-type alloantigens or ABO blood group antigens, bind to the donor tissue. This initiates complement-mediated lysis and induces a procoagulant state, resulting in immediate graft thrombosis. Exposure usually is in the form of prior transplant, transfusion, or pregnancy. While there is no treatment for hyperacute rejection, a thorough understanding of its cause has resulted in preoperative screening tests, namely the lymphocytotoxic crossmatch and ABO typing. These two tests identify donor-to-recipient combinations in which HAR could be avoided. The crossmatch is performed by mixing cells (nonactivated T cells that express class I, but not class II, MHC antigens on their surface) from the donor with serum from the recipient in the presence of complement. Lysis indicates the presence of antibodies directed against the donor. Many variations of this test exist, each detecting many antibody types against a variety of donor antigens. However, only detection of IgG antibodies directed against class I MHC molecules represents a positive test and a contraindication to transplantation. Preoperative verification of proper ABO matching and a negative crossmatch effectively prevents hyperacute rejection in 99.5 percent of transplants.

A delayed variant of HAR known as *vascular rejection* also is mediated by humoral factors. Vascular rejection occurs when offending alloantibodies exist in circulation at levels undetectable by the crossmatch assay, even though presensitization has taken place. Reexposure leads to restimulation of the memory B cells responsible for the donor-specific antibodies. The result is initial graft function, followed by deterioration on or about postoperative day 3. Enhanced immunosuppression with steroids, combined with nonspecific antibody depletion with plasmapheresis, occasionally is successful in attenuating vascular rejection. Vascular rejection also can occur through de novo antibody synthesis spurred on by T-cell-dependent B-cell activation.

Acute Rejection

Acute rejection is caused primarily by T cells and evolves over a period of days to weeks. It can occur at any time after the first 5 postoperative days, but is most common in the first 6 months, and is the inevitable result of an allotransplant unless immunosuppression directed against the T cell is used. To initiate acute rejection, T cells bind antigen via their TCR either directly or after phagocytosis of donor tissue and re-presentation of MHC peptides by self APC. This leads to cell activation as described previously. The result is a massive infiltration of the graft of T cells, with destruction of the organ (Fig. [10-6](#)).

Kidneys can be preserved long enough to allow organ allocation to the recipient to be the most closely matched to the MHC of the donor. Other organs cannot be matched in this way. The incidence of acute rejection declines with decreasing MHC disparity, but any mismatch puts the patient at risk for T-cell-mediated graft destruction and mandates T-cell-specific immunosuppression. Like HAR resulting from humoral presensitization, presensitization at the T-cell level results in an accelerated form of cellular rejection mediated by memory T cells.

Treatment leads to successful restoration of graft function in 90 to 95 percent of cases, and failure to treat results almost uniformly in graft loss. Prompt recognition of acute rejection is imperative. Most acute rejection episodes for patients on modern immunosuppression therapy are asymptomatic until the secondary effects of organ dysfunction occur. By this time the rejection is well-entrenched and difficult to reverse. For this reason, monitoring for acute rejection must be intense, particularly during the first year after transplantation. Unexplained graft dysfunction should prompt biopsy and evaluation for the lymphocytic infiltration and graft parenchymal necrosis characteristic of acute rejection. Liver acute rejection is characterized by eosinophil infiltration in addition to lymphocytic infiltration.

Chronic Rejection

Unlike acute and hyperacute rejection, chronic rejection (CR) is poorly understood. Its onset is insidious, occurring over a period of months to years, and because the pathophysiology is not well defined, it is untreatable. Heightened immunosuppression is not effective in reversing or retarding the progression of chronic rejection. Its distinction from acute rejection by biopsy is important. Acute rejection can be superimposed on CR, with treatment for the acute rejection leading to partial return of graft function. Histologically, CR tends to exhibit anomalies; regardless of the organ involved, it is characterized by parenchymal replacement by fibrous tissue with a relatively sparse lymphocytic infiltrate. Those organs with epithelium show a dropout of the epithelial cells and endothelial destruction. Chronic rejection is related to the recognition of MHC by the recipient's immune system, but nonimmunologic factors probably have a role in its progression. The time course, histology, and refractory nature of CR suggest that direct cell-mediated tissue destruction is not a primary mechanism. Cumulative effects of mild subclinical immune recognition by several limbs of the immune system, and the resulting exposure to soluble factors, including fibrogenic cytokines, eventually takes its toll on the fragile epithelium and endothelium. Chronic rejection requires retransplantation.

Immunosuppression

Without some attenuation of the immune system, all allografts eventually would be destroyed. The redundancy and plasticity of the innate and acquired immune systems has prevented any single manipulation from specifically preventing graft destruction. As allograft recognition is mediated by an immune system formed for the detection and elimination of pathogenic microbes or, to some extent, malignantly transformed cells, manipulations altering this system do so at the expense of a vital defense network. No immunosuppressive intervention is allograft-specific, nor do all maneuvers preventing allograft loss put the recipient at increased risk for infection or malignancy. Rational, selective use of several immunosuppressive agents acting through different synergistic mechanisms is required to successfully prevent rejection without completely removing the body's defenses.

For all organs, the events occurring at the time of transplantation are the most critical in establishing the state of immune unresponsiveness necessary for long-term graft survival. For this reason, immunosuppression is extremely intense in the early postoperative period and subsequently tapers. This initial conditioning of the recipient's immune system is known as *induction* immunosuppression. It usually involves deletion of the T-cell response completely and cannot be maintained indefinitely without lethal consequences. Medications used to prevent acute rejection for the life of the patient are called *maintenance* immunosuppressants. These agents are well tolerated if dosed appropriately. All have side effects that increase the risk of infection and malignancy. Immunosuppressants used to reverse an acute rejection episode are called *rescue* agents. They are the same as the agents used for induction therapy.

Corticosteroids

Corticosteroids, in particular the glucocorticoid effect of steroid preparations, remain a central tool in the prevention and treatment of allograft rejection. Used alone, they are ineffective in preventing allograft rejection, but in combination with the other agents described below they significantly improve graft survival. Higher doses of steroids also are used as a rescue agent to treat acute cellular rejection. Although steroids have a desirable immunosuppressive effect, they can contribute significantly to the morbidity of transplantation.

Despite the use for decades of steroids as an antiinflammatory agent, the mechanism of the immunosuppressive effect of glucocorticosteroids has not been completely elucidated. Glucocorticosteroids bind to an intracellular receptor after nonspecific uptake into the cytoplasm. The receptor-ligand complex then enters the nucleus, where it acts as a

DNA-binding protein and increases the transcription of several genes, the most important of which is probably the gene for IkBa. This protein binds to and prevents the function of NF-κB, a key activator of proinflammatory cytokines. In doing so, steroids prevent the primary mechanism by which lymphocytes amplify their responsiveness. The resulting effects are predictably diverse. Steroids block transcription of IL-1 and TNF-α. They also block IFN-γ production, PMN migration, and lysosomal enzyme release by PMNs. Phospholipase A₂, and consequently the entire arachidonic acid cascade, is inhibited. The up-regulation of MHC also is muted by steroids. By blocking the response of leukocytes to chemotactins, and by inhibiting vasodilators, such as histamine and prostacyclin, steroids exert a profound effect to dampen the inflammatory response and ultimately to blunt T-cell proliferation. They do not have a significant influence on antibody production.

Hydrocortisone (cortisol) is the major endogenous glucocorticoid. Numerous synthetic steroids have greater antiinflammatory potency and less mineralocorticoid effect than cortisol. In addition, the plasma half-life of each preparation varies. The most commonly used form of steroid is prednisone or its intravenous substitute, methylprednisolone.

The adverse effects of steroid therapy are numerous. Patients receiving maintenance steroids have a suppressed hypothalamic-pituitary-adrenal axis. When undergoing the stress of severe illness or injury, these patients need coverage with a steroid preparation in the amount equal to their endogenous adrenal capacity. This should not exceed 50 mg hydrocortisone every 8 h. The acute complications include impairment of glucose tolerance, delayed wound healing, salt and fluid retention that may exacerbate hypertension, and CNS effects, such as insomnia, depression, nervousness, and euphoria. These events usually can be ameliorated by tapering the steroid dose. Chronic side effects of corticosteroids include Cushing's syndrome (central obesity, acne, striae, hirsutism, altered facies), cataracts, muscle wasting, and growth retardation in prepubertal children. Patients show increased propensity toward peptic ulceration. Osteoporosis results from the combined effects of the inhibition of bone matrix formation and of intestinal absorption of calcium. The current trend in transplantation is to minimize the dose of steroids used (limiting the side effects) by adding two to four other immunosuppressive agents. There also are subsets of patients who may not require maintenance steroids, such as those with HLA-identical donors. In addition, a patient's rejection history should be considered; patients who have survived a year without a rejection episode can be considered for withdrawal from steroids.

Antiproliferative Agents

Azathioprine

The antimetabolite azathioprine was the first immunosuppressive pharmaceutical agent used in organ transplantation, and it remains a part of many maintenance immunosuppressive protocols. Azathioprine undergoes hepatic conversion first to 6-mercaptopurine (6-MP), and then to 6-thio-inosine monophosphate (6-tIMP). These derivatives inhibit DNA synthesis by alkylating DNA precursors and inducing chromosomal breaks through interference with DNA repair mechanisms. They also inhibit the enzymatic conversion of *inosine monophosphate (IMP)* to adenosine monophosphate (AMP) and *guanosine monophosphate (GMP)*. The primary effect is to deplete the cell of adenosine. The effects of azathioprine are relatively nonspecific; it acts not only on proliferating lymphocytes and PMNs but on all rapidly dividing cells.

Azathioprine effectively inhibits rejection when given as a maintenance agent but, unlike steroids, has no value as a rescue or induction agent. It is generally given in a dose of 1 to 3 mg/kg/day. Its primary toxicity is directed at the bone marrow, gut mucosa, and liver. The dose is decreased as the total white blood cell count falls, or it is stopped when leukopenia develops (usually less than 3,000/mL). In the presence of infection, azathioprine generally is withheld so as not to impair the physiologic response of the immune system. Because azathioprine is a hepatotoxin, its use is limited during liver dysfunction, and it is rarely used in liver transplantation. Mycophenolate mofetil is gaining popularity as a more specific purine antimetabolite.

Mycophenolate Mofetil

Mycophenolate mofetil (MMF) (RS-61443) is a potent immunosuppressive agent approved for use in adults. It is a morpholinoethyl ester of mycophenolic acid (MPA), an established noncompetitive, reversible inhibitor of IMP dehydrogenase. This modification improves the bioavailability of MPA.

Physiologic purine metabolism requires that GMP be synthesized for subsequent synthesis of guanosine triphosphate (GTP) and deoxyguanosine monophosphate (dGTP). GTP is required for RNA synthesis and dGTP for DNA synthesis. GMP is formed from IMP by IMP dehydrogenase, and therefore MMF prevents a critical step in RNA and DNA synthesis. Of major importance is the presence of a "salvage pathway" for GMP production in most cells *except* lymphocytes (hypoxanthine-guanine phosphoribosyl transferase–catalyzed GMP production directly from guanosine). MMF exploits a critical difference between lymphocytes and other body tissues, including PMNs, to produce relatively selective immunosuppressive effects. MMF blocks the proliferative response of both T and B lymphocytes, inhibits antibody formation, and prevents the generation of cytotoxic T cells. MMF also strongly suppresses in vitro B-cell memory responses.

MMF does not cause nephrotoxicity or hepatotoxicity at immunosuppressive doses, and its marrow-suppressive potential is less than that of azathioprine. Mild gastrointestinal irritation, in the form of diarrhea and nausea, has been reported. Because of its low toxicity in human beings and its antiproliferative effect on T and B cells, MMF was evaluated in a multicenter clinical trial of renal allotransplantation, comparing it in combination with prednisone and cyclosporine to azathioprine, prednisone, and cyclosporine. The results showed that MMF decreased the rate of biopsy-proved rejection or treatment failure from 48 to 31 percent and decreased the need for antilymphocyte agents in rescue therapy from 20 percent to a range of 5 to 10 percent, depending on the dose of MMF used. MMF has replaced azathioprine in patients with high risk of rejection. MMF also is effective as a rescue agent. In a comparison of high-dose methylprednisolone to MMF for the treatment of biopsy-proved rejection, MMF was shown to reduce the incidence of graft loss by 45 percent and the risk of recurrent rejection by 50 percent. MMF may interact with tacrolimus, potentiating the effect and possibly the side effects of this drug.

Calcineurin Inhibitors

Cyclosporine

In 1976 Borel demonstrated the T-cell–specific immunosuppressive properties of cyclosporin A, a cyclic endecapeptide isolated from the fungus *Tolypocladium inflatum* Gams. In 1983 cyclosporine was approved for use in the United States. Its use resulted in dramatic improvements in the results of all organ transplants, but particularly in hepatic and cardiac transplantation. It has remained a mainstay immunosuppressant in most maintenance regimens.

Cyclosporine's mechanism of action is mediated primarily through its ability to bind to cytoplasmic protein cyclophilin. Cyclophilin is a *cis-trans* peptidyl- prolyl isomerase critical for the proper folding of proteins (also known as a rotamase or an immunophilin). This function of cyclophilin is not critical to the immunosuppressive effects of cyclosporine but may be related to the toxic side effects of the drug. The cyclosporine-cyclophilin complex binds with high affinity to the calcineurin-calmodulin complex and blocks its role in the calcium-dependent phosphorylation and activation of the transcription- regulating factor NF-AT. This prevents the transcription of the IL-2 gene and other genes critical for T-cell activation. The role of cyclophilin in cellular cytotoxicity is probably interrupted. Cyclosporine reversibly inhibits T- lymphocyte–mediated immune responses, but it does not prevent antigen recognition by T cells, and its effects can be overcome with exogenous (or in the case of an ongoing rejection episode, ambient) IL-2. For this reason, once IL-2 is present in the graft cytokine milieu, cyclosporine is ineffective. Cyclosporine therefore works solely as a maintenance agent and is ineffective as a rescue agent.

Cyclosporine is a cyclic polypeptide, insoluble in aqueous solutions but soluble in lipids and organic solvents. A microemulsion formulation is now available that has improved cyclosporine's lipophilic nature. It usually is given orally but can be administered intravenously by slow infusion. While the gastrointestinal absorption of cyclosporine depends on bile flow (a significant concern in liver transplantation), the absorption of its microemulsion form is not. Cyclosporine is metabolized by the hepatic cytochrome P-450 enzymes, and blood levels are therefore increased by inhibitors of cytochrome P-450 (ketoconazole, erythromycin, calcium channel blockers) and decreased by cytochrome P-450 inducers (rifampin, phenobarbital, phenytoin). Liver failure slows the clearance of cyclosporine, because 90 percent of its metabolites are cleared in the bile. Renal failure does not alter cyclosporine clearance significantly, because only 6 percent of the dose given appears in urine.

Cyclosporine causes dose-related nephrotoxicity, because it has a vasoconstrictor effect on proximal renal arterioles. Cyclosporine tends to increase vascular resistance in the kidney and may delay the resolution of acute tubular necrosis or hepatorenal insufficiency. An additional renal effect is an idiosyncratic reaction producing hemolytic uremic syndrome. Hyperkalemia also may result from its effects on the proximal and distal renal tubules. Long-term use causes a 30 percent reduction in renal function, which usually can be reversed by discontinuing the drug. Hypertension also is a common adverse effect but usually can be effectively treated. Cyclosporine frequently causes neurologic side effects consisting of tremors, paresthesias, headache, depression, confusion, somnolence, and, rarely, seizures. Hypertrichosis of the face, arms, and back is seen in about 50 percent of patients. Gingival hyperplasia also may occur. Hepatotoxicity with elevation of bilirubin and serum aminotransferases may occur, as well as elevation of skeletal alkaline phosphatase. Renal toxicity usually is noticeable before hepatotoxicity occurs.

Tacrolimus

Tacrolimus (FK506) is a macrolide produced by *Streptomyces tsukubaensis* that was discovered in 1984 in Japan during a search for new immunosuppressive agents. Kino and coworkers first demonstrated its in vitro immunosuppressive properties in 1987. Tacrolimus, like cyclosporine, blocks the effects of NF-AT, prevents cytokine transcription, and arrests T- cell activation. The intracellular target is an immunophilin protein distinct from cyclophilin, known as FK-binding protein (FK-BP), the effect of tacrolimus is thus additive to that of cyclosporine, and the use of tacrolimus with cyclosporine produces prohibitive toxicity. Tacrolimus is 100 times more potent in blocking IL-2 and IFN-γ production than cyclosporine. Like cyclosporine, the effects of tacrolimus are relatively T-cell–specific, but in addition to its role as a maintenance agent, tacrolimus has shown promise as a rescue agent. The side effect profile of tacrolimus is similar to that of cyclosporine with regard to renal and hepatic toxicity. Neurotoxicity, in the form of tremors and mental status changes, is somewhat more pronounced, as is its diabetogenic effect. Cosmetic side effects are reduced substantially. Tacrolimus is extremely effective for liver transplantation and is the drug of choice. Enthusiasm for its use in children, particularly infants, has been tempered by a high rate of posttransplant lymphoproliferative disorder (PTLD), approaching 40 percent in some series. This may reflect early dosing inexperience.

Antilymphocyte Preparations

Antilymphocyte Globulin

Antilymphocyte globulin (ALG) is a polyclonal serum produced by inoculating heterologous species with human lymphocytes, collecting the plasma, and purifying the IgG fraction. Thymocytes rather than lymphocytes are sometimes used, and this is designated as *anti-thymocyte globulin (ATG)*. The most commonly inoculated species is the horse, and ATGAM is the most widely used preparation. Sera from rabbits and goats also are available. ATGAM targets the central mediator of acute rejection, the T cell, by coating multiple epitopes on this cell type and promoting their clearance through complement-mediated lysis, opsonin-induced phagocytosis, and internalization of key surface receptors. Because ATGAM is polyclonal (composed of antibodies with many specificities), antibodies to both T and B cells, and probably to other cells, are present. Only a small fraction of the serum is estimated to be biologically active against T lymphocytes. One prominent side effect is severe thrombocytopenia, which may result from cross-reactivity with platelets and therefore may limit the use of the drug.

The usual dose of ATGAM is 10 to 20 mg/kg/day, given intravenously, and the dose may be adjusted according to toxicity. The usual duration of therapy is 10 to 14 days, but this may vary. ATGAM should be given through a central line to avoid phlebosclerosis and delivered through a filter to prevent administration of insoluble aggregates that may develop during storage. Infusion should be relatively slow (over 4 to 6 h) to limit the intensity of any adverse reactions. Any history of allergy to horse or horse products should be elicited from the patient before ATGAM is used. Intradermal skin testing is available, but anaphylaxis is rare.

ATGAM is most commonly used as part of a multidrug induction immunosuppression protocol in renal transplantation with cyclosporine, azathioprine or mycophenolate mofetil, and prednisone. Limiting the dosage of each drug, and thus limiting the side effects, is the rationale for multidrug therapy. ATGAM often is used sequentially with cyclosporine or tacrolimus in the early postoperative period after renal transplantation to avoid the nephrotoxicity of these drugs early in the transplant course.

Most of the side effects of ATGAM are from its heterologous origin and heterogenous composition, but major side effects are rare, and the drug is well tolerated by most transplant recipients. The most common symptoms are the result of transient cytokine release after antibody binding. Chills and fevers occur in up to 20 percent of patients. These symptoms usually are transient and can be ameliorated by pretreatment with steroids, antihistamines, and antipyretics. The symptoms generally diminish in intensity as the treatment progresses. A skin rash characterized by large, raised, erythematous wheals on the trunk, neck, and proximal extremities may develop in up to 15 percent of patients. Treatment does not need to be discontinued because of these symptoms, which usually can be treated effectively with an antihistamine. Thrombocytopenia and leukopenia do require an alteration in treatment. Thrombocytopenia occurs to some degree in most patients, and leukopenia occurs less frequently. Hemolysis also may occur. When using ATGAM it is imperative that daily white blood cell and platelet counts be performed. Decreasing the dose or temporarily discontinuing therapy usually is accompanied by rapid recovery. If thrombocytopenia is the problem and therapy must be continued, pooled platelet transfusions prior to administration may be used. Because antilymphocyte preparations profoundly inhibit T lymphocytes, they also suppress cell-mediated immunity. The use of ALG has been associated with an increase in the reactivation and development of primary *cytomegalovirus (CMV)* infections. In addition to CMV infections, herpes simplex virus (HSV), Epstein-Barr virus (EBV), and varicella infections may occur more frequently after therapy with antilymphocyte preparations.

OKT3

Monoclonal antibodies have very specific targets. The technology for the creation of effective amounts of monoclonal antibodies stems from the hybridoma work in the late 1970s by Kohler and Milstein. The only commercially available preparation of monoclonal antibodies for use in organ transplantation is the murine monoclonal antibody to the signal transduction subunit on human T cells (CD3), known as Orthoclone OKT3. OKT3 is the product of a hybridoma formed from a myeloma cell and a primed murine lymphocyte capable of producing this antibody; the cell lines are cultured in mouse ascites and the final product is purified.

There are several ways in which OKT3 is thought to have its effect. The CD3 determinant is a cluster of transmembrane proteins found on the surface of all mature T lymphocytes. Because OKT3 binds to the CD3 determinant, it prevents signal transduction of the T-cell receptor antigen binding event and arrests amplification of a rejection episode. After the administration of OKT3, there is a rapid decrease in the number of circulating T lymphocytes. This is partially a result of opsonization and clearance by the reticuloendothelial system of the OKT3-lymphocyte complex. Several days after administration, there is a return of T cells expressing the accessory binding molecules CD4 and CD8, but lacking the T-cell receptor. Another way in which OKT3 exerts its effect is by down- regulation of the T-cell receptor complex, producing a “blind” T cell incapable of binding to antigen. In addition to interfering with the generation of cytotoxic T cells and the modulation of cell surface proteins, OKT3 blocks the cytotoxic activity of already activated T cells through inappropriate activation and degranulation. This is perhaps its most important function, but it leads to substantial side effects.

T-cell–derived cytokines have evolved as activators of adjacent cells. Their release is strongly polarized to the side of the cell actively engaged in cell- to-cell contact. The most potent T-cell activator, IL-2, exerts most of its effect in an autocrine loop. Pan-activation of the body’s T cells leads to transient activation and systemic cytokine release. This is similar to, but greater in magnitude than, the effect of the superantigen staphylococcal exotoxin, the etiologic agent of toxic shock syndrome. Administration of OKT3 leads to a profound, systemic cytokine release syndrome that can result in hypotension, pulmonary edema, and, rarely, fatal cardiac myodepression. In approximately 2 percent of patients, the inflammatory response manifests itself as aseptic meningeal inflammation. Administration of high-dose methylprednisolone prior to OKT3 administration is required to blunt this adverse response, but rarely is the response averted altogether. The syndrome abates with subsequent dosage as the target cells available for degranulation are consumed or exhausted.

The usual adult dose of OKT3 is 5 mg/day for 10 to 14 days. Unlike ATGAM, which must be given slowly and into a central line, OKT3 is given as a bolus injection over 1 min and may be given through a peripheral line. Pretreatment with methylprednisolone before administration of OKT3 as well as the use of antihistamines and antipyretics will help to limit the adverse reactions. Measurement of serum levels of OKT3 is possible but not done routinely. More commonly, the percentage of CD3-positive cells is determined by flow cytometry. The presence of less than 10 percent CD3- positive cells usually is associated with therapeutic efficacy; more than 10 percent CD3-positive cells usually indicates the presence of anti-OKT3 antibodies, which may limit the therapeutic effect. These antibodies may be directed against structural regions of the antibody (constant or variable regions) or the actual binding site (idiotypic) of the OKT3 antibody; they usually arise after a prolonged course of OKT3 or after the cessation of therapy. Levels of these antibodies can be measured. There is evidence that the use of azathioprine and other immunosuppressive agents during treatment with OKT3 may reduce the formation of anti-OKT3 antibodies. Occasionally, increasing the dose of OKT3 can overcome the effect of these antibodies.

OKT3 was first used as a rescue agent to treat acute renal allograft rejection. It is vastly superior to conventional steroid therapy in reversing rejection and improving allograft survival, but its side effects and the limiting nature of the antimurine antibody response have served to limit its use to the treatment of steroid-resistant rejection. The use of OKT3 in induction immunosuppressive protocols for kidney transplants is common. There are compelling data supporting the use of OKT3 in immunological high-risk patients, such as those requiring renal retransplants, to improve graft survival. OKT3 is used extensively in heart transplantation and may be used to reverse steroid-resistant rejection in liver transplant recipients.

OKT3, like other antilymphocyte preparations, causes a high reactivation rate of cytomegalovirus. Epstein-Barr virus infection leading to lymphoproliferative disorders also has been reported with the use of OKT3. Pediatric patients treated with OKT3 are at a higher risk of becoming infected with the varicella virus. It is not uncommon to see a rise in serum creatinine after the first several doses of OKT3 when treating acute renal allograft rejection. This may be a result of lysis of T lymphocytes within the renal tubules. The creatinine usually rises for 2 to 3 days and falls as treatment progresses.

New Immunosuppressive Agents

Rapamycin

Rapamycin is a macrolide antibiotic derived from *Streptomyces hygroscopicus*. This agent was described initially as an antifungal agent and was subsequently found to have immunosuppressive properties. Structurally, rapamycin is similar to tacrolimus, and the two have been shown to antagonize each other's biologic activity. Both interact with the cytoplasmic protein FK-BP, but rapamycin does not affect calcineurin activity, because its binding sterically hinders engagement with this molecule. As a result, rapamycin does not inhibit the expression of NF-AT or of IL-2, but it impairs signal transduction by the IL-2 receptor (IL-2R) through interaction of the rapamycin/FK-BP complex with the cytoplasmic protein RAFT-1, a critical kinase in IL-2R–associated activation. In doing so, the p70 S6 kinase cascade is arrested and T cells are prevented from entering the S phase of cell replication. This has the advantage of interrupting the T-cell activation pathway even if IL-2 is present from an exogenous source or ongoing rejection. Other receptors also are affected, including those for IL-4, IL-6, and platelet-derived growth factor (PDGF). Another benefit may lie in rapamycin's ability to antagonize B-cell lymphomas, and potentially it may be used to prevent the development of posttransplant lymphoproliferative disease.

At very low doses rapamycin dramatically prolongs mouse heart and skin allograft survival, as well as rat cardiac allograft survival. It is significantly more potent than cyclosporine. In vitro studies also have demonstrated marked synergy among rapamycin, cyclosporine, and steroids, allowing a dramatic reduction in dosages while inhibiting lymphocyte proliferation. These findings are important, because the toxicity of these agents should be reduced with combined therapy. Nephrotoxicity has not been observed, but hypertriglyceridemia may be a significant side effect.

Deoxyspergualin

The antitumor antibiotic spergualin was isolated from *Bacillus laterosporus* in 1981. Deoxyspergualin (DSG) has the strongest antiproliferative properties, and it is an immunosuppressant. The mechanism of DSG is not completely understood, but there is evidence that it is immunosuppressive via a predominantly antimonocyte-antimacrophage effect. It does prevent the nuclear translocation of NF-kB, perhaps through its association with the heat-shock proteins HSP 70 and HSP 90. This interaction inhibits the release of IκB, possibly mimicking a critical part of the steroid mechanism of action. The close relationship with HSPs, combined with evidence that the inhibitory effect of DSG on cytotoxic T cells can be abolished by the MHC–up-regulating cytokine IFN-γ, suggests that the primary effect of DSG is to inhibit antigen presentation or the costimulatory function of APCs.

DSG is effective in prolonging allograft survival in animal models of renal, pancreas, liver, and heart transplantation. Studies in renal transplantation suggest that it will be effective in this setting. Although severe gastrointestinal disturbances in canines have been noted, these may be species-specific. A trial of DSG in renal transplant recipients with rejection reported only mild gastrointestinal disturbances. Headache, fatigue, perioral numbness, and decreased white blood cell and platelet counts and hematocrit level also were noted. None of these was severe enough to justify discontinuing therapy. The overall response rate was 79 percent in 34 cases of rejection. Additional study is required to elucidate the mechanism of action of DSG and its place in the immunosuppressive armamentarium.

Brequinar

Brequinar sodium (BQR) is an immunosuppressive agent that selectively inhibits T- and B-cell proliferation. It functions by inhibiting the enzyme dihydroorotate dehydrogenase, thus interfering with pyrimidine synthesis. BQR, like mycophenolate mofetil, inhibits lymphocyte DNA synthesis and interrupts the rejection process at the level of clonal expansion. Prolongation of heart, kidney, and liver allograft survival was shown in several experimental models, and the effects of BQR are synergistic with cyclosporine. Primary side effects are related to leukopenia.

Anti-IL-2 Strategies

One approach to achieving more specific immunosuppression is to target antigens expressed only on activated lymphocytes. The interleukin-2 receptor (IL-2R) is one such antigen. While a low-affinity version of the IL- 2R is constitutively expressed on T cells, the high-affinity form required for clonal expansion is dependent on the up-regulation of a 55-kD subunit known as Tac. Experience with anti-Tac monoclonal antibodies has been encouraging. Induction with anti-Tac has prevented and reduced the frequency of early rejection episodes when used in combination with cyclosporine. Because anti-Tac and cyclosporine are synergistic, lower doses of cyclosporine could be used, thereby decreasing nephrotoxicity. A rat IgG2a monoclonal antibody, 33B3.1, also demonstrated excellent results in preventing posttransplant rejection, even without cyclosporine. Once rejection occurred, however, 33B3.1 was effective in reversing only 60 percent of rejection episodes. Experience has been reported with BT563, also of murine origin. In a double-blind placebo-controlled study using BT563 as an induction agent with cyclosporine and prednisone, freedom from rejection at 3 months was observed in 89 percent of the treated group, compared to 72 percent of the control group. While these results are comparable to other induction strategies with ATGAM or OKT3, the side effect profile was more favorable. This approach combines the specificity of monoclonal therapy with the targeting of a nonactivating ligand, avoiding the profound cytokine release syndrome seen with OKT3, and the promiscuous cell lysis seen with ATGAM. Longer follow-up is required to fully evaluate the efficacy of these agents.

Costimulation Blockade

The prevention of rejection has been based on a strategy of pharmacologic suppression of T-cell activation or T-cell ablation. Although this approach is effective, it is not specific for

the allogeneic response; rather it pan- suppresses all T-cell immunity. A more comprehensive understanding of T- cell biology may allow specific control of immunity by exploitation of the requirements for dual T-cell stimulation. Once the TCR is engaged, the T cell can follow one of two paths: activation, through CD28 binding to B7, or anergy, in the absence of CD28 binding. Interruption of the CD28 pathway of costimulation may selectively anergize only those cells undergoing binding to the allograft, leaving nonreactive cells unaffected. Increasing amounts of literature document a tolerogenic effect of specific biologic blockade of the CD28 or CD40 pathway. In rodents, simultaneous blockade of CD28 and CD40 at the time of transplantation can allow for indefinite survival of cardiac and skin allografts without the need for any subsequent immunosuppression. Agents under evaluation include antibodies directed against B7-1 and B7-2, CD28, CD40, gp39, and a fusion protein, CTLA4-Ig, which links the receptor CTLA4 to the Fc portion of IgG.

Xenotransplantation

The success of transplantation has led to its widespread application, but the indications for organ replacement have expanded far beyond the available donor supply. Most people who could benefit from a transplanted organ are unable to receive one, and those who do receive one usually have a significant waiting time and undergo a deterioration before operation. Many investigators are examining the feasibility of xenotransplantation to improve the availability of organs. There are two types of xenografts, discordant and concordant. The immunology of these two types of xenografts differs markedly.

Concordant Xenografts

For human beings, concordant grafts are derived from closely related species, such as Old World monkeys and apes. The critical element defining an animal as concordant is the assembly of carbohydrate antigens on the cell surface. In particular, the enzyme galactosyl transferase is absent in these animals, preventing the synthesis of the N-linked disaccharide galactose α(1-3)-galactose [GAL α(1-3)GAL]. Rather, they express typical blood group antigens of the ABH system. The antibodies present in circulation of potential human recipients can be predicted by ABH typing, avoiding the problem of hyperacute rejection (HAR) that occurs for allografts. With HAR removed as a threat, the typical mechanisms of graft rejection are left, including acute cellular rejection, acute vascular rejection, and, presumably, chronic rejection.

Most of the critical molecular elements responsible for antigen presentation and T-cell–mediated rejection are evolutionarily conserved in mammals. Xenogeneic MHC polymorphism is restricted to the antigen-binding region, with the areas responsible for accessory molecule binding intact. Costimulatory and adhesion molecules required for adequate T-cell interaction with a target cell function well across the species barrier. For these reasons, cellular rejection and T-cell–dependent humoral rejection proceed similarly to an allograft completely mismatched at the HLA locus.

Experimental models of concordant xenograft rejection, and occasional ventures into the clinical arena, have demonstrated that concordant xenotransplantation is feasible with available immunosuppressive pharmaceutical agents. Prevention of concordant xenograft rejection would require induction with a T-cell–specific biologic agent and chronic cellular suppression, including calcineurin inhibition, an antiproliferative agent such as mycophenolate mofetil, and steroids. Results matching those seen for poorly matched allografts could be expected, including the complications from the immunosuppression required for engraftment. With the possible exception of neonatal cardiac transplantation, the use of endangered, intelligent species with little ability to be bred or genetically altered, could not meet the need for readily available, size-matched organs. Widespread application of concordant xenografts would quickly deplete the supply of nonhuman primates, particularly when a loss rate extrapolated from poorly matched allografts is taken into consideration. There also is significant concern that zoonotic transfer of disease, in particular, retroviral illness, would put the patient and the public at undue risk. It is anticipated that xenotransplantation will primarily involve discordant species.

Discordant Xenografts

Discordant grafts are derived from distantly related species, such as New World monkeys and nonprimates. When transplanted into human beings, these organs rapidly undergo hyperacute rejection. The primary cause of the rejection varies from species to species but generally involves preformed IgM antibodies directed against atypical carbohydrate residues, such as GAL α(1-3)GAL, acting in concert with complement. These antibodies might serve a protective role against cross-reactive bacteria. Vigorous T-cell–mediated rejection also can occur because human T cells can be bound and activated by discordant MHC. Induced antibody responses similar to a vigorous vascular rejection are seen. In addition to the typical patterns of cellular rejection, there also is a substantial detrimental effect of natural killer cells and other innate cellular elements.

Because of the spectrum of effectors involved in discordant rejection, intervention at all levels of innate and acquired immunity will be required for successful engraftment. A multimodal approach will be required. Significant advances have been made, particularly in the pig, toward establishing a xenogeneic source of donor organs. Several groups have transgenically altered pigs to express human complement regulator proteins, such as CD59, decay-accelerating factor, and membrane cofactor protein. Other investigators have altered the balance of carbohydrate processing to favor production of typical H-antigen synthesis over GAL α(1-3)GAL. Added to conventional immunosuppression, these interventions as well as several inhibitors of complement, including anti-C5 monoclonal antibodies, cobra venom factor, and soluble complement receptor type 1, have been successful in abrogating HAR and extending survival in primates from minutes to weeks. While clinical application remains several years away, the flexibility afforded by genetic engineering will allow for custom organs to be available as necessary.

VASCULAR AND NERVE GRAFTS

Hans W. Sollinger

Vascular grafts usually are used as autografts. Their predominant use is for replacement of damaged or atherosclerotic arteries. Vein grafts are the optimal conduit for coronary artery bypass grafting. The saphenous vein is harvested from one or both legs of the individual undergoing coronary revascularization. Careful handling of the vein grafts, avoiding mechanical trauma and overdistention during flushing, is mandatory. Vein grafts may undergo similar changes to the patient's own coronary arteries as time passes. Vein grafts also are used for reconstruction of the arterial supply of the lower extremities. Femoral popliteal, or even distal bypass grafting, can be achieved with the use of appropriate vein grafts. In patients with injuries to the arterial systems, vein grafts can be used as patch grafts to cover a defect of limited size. In selected cases, vein grafts may be used for the creation of long-term dialysis access, but because of repeated sticks by large bore dialysis needles, aneurysms develop frequently, limiting the life span of these conduits.

Autografting of arteries is used occasionally for arterial replacement of a diseased renal artery. The hypogastric artery is the best size-matched artery. Allogeneic grafting of arteries is rare in normal individuals because it requires systemic immunosuppressive therapy. Sometimes allogeneic arterial substitutes can be used for the repair of renal transplant stenosis. The authors have replaced the diseased or stenotic renal artery on several occasions with segments of common or external iliac artery from a cadaveric donor source. There has been no incidence of rejection, aneurysm, or restenosis. Allogeneic arterial conduits are the ideal replacement material for transplant renal artery stenosis.

Nerve autotransplants are used to bridge defects after traumatic injury to important motor nerves. Allografts in the form of liophilized nerve grafts have been attempted, but are less successful. The transplanted nerve serves as a route along which the patient's own nerve can regenerate. Attempts have been made to transplant vascularized nerve grafts utilizing microsurgical techniques.

PANCREAS

Jon S. Odorico and Hans W. Sollinger

The discovery of insulin in 1921 was hailed as the cure of diabetes. By preventing death from diabetic coma and controlling overt symptoms, insulin rapidly led to increased life expectancy among diabetics and altered the natural history of diabetes. But as diabetic patients lived longer, complications developed in other vital organs. Diabetes is now the leading cause of renal failure in the United States and commonly contributes to blindness, debilitating neuropathies, and accelerated atherosclerosis. Numerous studies, particularly the Diabetes Control and Complications Trial (DCCT), have demonstrated the beneficial effect of intensive glucose control on arresting the development and progression of end organ complications. Experimental animal studies have shown that kidney transplants from normal to diabetic animals develop histologic findings characteristic of diabetes, but kidney transplants from diabetic to normal animals showed disappearance or stabilization of these lesions. These observations suggest that it might be possible to prevent or ameliorate systemic complications of diabetes by achieving more precise glucose control.

Alternative methods of insulin replacement therapy are under investigation. These include insulin pumps, bioartificial pancreases (vascularized, implantable devices seeded with pancreatic islets and protected from the patients' immune system by a semipermeable membrane), and pancreatic islet allotransplantation. None of these modalities has been successful enough to warrant more widespread application. It is understood that Type I diabetes is an autoimmune disease, and attempts have been made to *prevent* the development of diabetes by modulating the host's immunity with cyclosporine. In a Canadian trial, cyclosporine given to newly diagnosed diabetics with residual insulin secretion was moderately successful in preventing diabetes. This approach is limited by the inability to reliably identify patients in the prediabetic state. The most successful approach to restoring normal long-term glucose homeostasis in Type I diabetes is whole-organ pancreas transplantation.

Historical Background

The first successful clinical pancreas transplants were performed by Kelly and Lillehei at the University of Minnesota in 1966. Their procedure employed transplantation of a segmental graft to the iliac fossa with ligation of the pancreatic duct. This technique resulted in a significant incidence of graft pancreatitis, and it became apparent that a method for handling the exocrine secretions of the graft was necessary. In 1973 Gliedman and associates suggested using the urinary tract for exocrine pancreatic drainage. They reported anastomosing the pancreatic duct to the ureter. Groth in Stockholm pursued the development of enteric drainage techniques, employing a Roux-en-Y anastomosis with temporary exteriorization of pancreatic juice via a small catheter.

In 1982 investigators at the University of Wisconsin developed the concept of direct drainage of exocrine secretions into the urinary bladder. This technique evolved from a direct duct-to-mucosa anastomosis to a duodenal button and eventually to a duodenal segment. Using a whole pancreaticoduodenal allograft with anastomosis of the duodenal segment to the bladder is the safest and most popular method of handling exocrine secretions, with over 90 percent of all pancreas transplants worldwide performed in this fashion. However, there has been a resurgence in the United States in using enteric drainage techniques.

Indications

Proper patient selection is crucial to the success of pancreas transplantation. Type I insulin-dependent diabetics younger than 45 years of age are potential candidates (Table [10-2](#)). Death in diabetics is often secondary to myocardial infarction and stroke; therefore, patients with significant coronary artery disease are excluded. Patients older than 25 years should undergo thallium stress testing, and if reversible ischemia is identified, coronary angiography is performed. Diabetics with major amputations secondary to severe peripheral vascular disease or severe visual impairment are not appropriate candidates, because they are unlikely to derive improvement in function at the end organ level. For major organ transplant necessitating long-term immunosuppression, contraindications include untreated malignancy, active infection, and HIV seropositivity.

Pancreas transplantation is performed in three different sets of circumstances: *pancreas transplantation alone (PTA)* in the nonuremic diabetic with minimal or no evidence of diabetic nephropathy; *pancreas transplantation after successful kidney allografting (PAK)* , and *pancreas transplantation performed simultaneously with a kidney transplant (SPK)* in

the uremic patient.

Approximately 90 percent of pancreas transplants performed in the United States are SPK transplants. SPK transplantation is the procedure of choice in the diabetic uremic patient with only mild and potentially reversible secondary complications of diabetes. One exception to this is the young diabetic patient for whom a suitable living-related renal transplant is available. A living-related renal transplant offers excellent long-term results with less immunosuppression than is required for SPK transplantation. After a successful renal allograft, the patient may be offered a PAK graft. PAK transplants should be offered to young diabetics with two or more end-organ complications of diabetes, who have undergone prior living-related or cadaveric renal transplantation, and who are currently stable, with excellent renal function. Two or more diabetic complications in diabetics with normal native renal function would be an indication for a PTA transplant. Other indications for PAK or PTA grafting in nonuremic diabetics include evidence of early histologic changes of diabetic nephropathy in their native or transplanted kidneys, and evidence of glucose hyperlability with frequent episodes of hypoglycemia without overt symptoms. Patients with marginal renal transplant or native kidney function should be strongly considered for SPK transplants because the high doses of cyclosporine or tacrolimus (FK506) used postoperatively after PAK and PTA transplants can precipitate renal failure.

Operative Procedure

The most widely used duct management technique for pancreas transplantation is bladder drainage, in which a duodenal segment is anastomosed side-to-side to the dome of the bladder (Fig. 10-7). Bladder-drained pancreas grafts reported to the *International Pancreas Transplant Registry (IPTR)* had significantly better long-term graft survival rates than grafts whose ducts were managed by enteric drainage, but there has been a renewed interest in draining the exocrine secretions enterically. Advantages of bladder drainage include the ability to use urinary amylase determinations as a screening test for rejection, avoidance of an enteric anastomosis and spillage of bowel contents, and reduced potential for peripancreatic infections. Advantages of the enteric drainage technique include avoidance of the postoperative urologic complications that occur in up to 30 percent of patients, avoidance of chronic dehydration and the need for bicarbonate replacement, and early removal of the Foley catheter. A retrospective comparison of pancreas transplants with bladder versus enteric drainage noted equal efficacy, graft survival rates, and morbidity rates.

The pancreatic graft harvested en bloc with the liver is first separated and prepared in cold saline solution at 4°C. Vascular reconstruction of the pancreatic blood supply is performed at this time with a donor iliac artery “Y” graft (Fig. 10-8). The duodenum segment is shortened to approximately 10 to 12 cm, taking care to avoid injury to the head of the pancreas and the ampulla of Vater. Peripancreatic fat is trimmed and neurovascular tissue in the region of the celiac axis is ligated carefully. The spleen is then removed by ligating the splenic artery and vein in the splenic hilum. If necessary, the common bile duct can be opened and cannulated to confirm patency of the ampulla of Vater. Arterial reconstruction is carried out by end-to-end anastomosis of the donor iliac artery “Y” graft to the splenic artery and the superior mesenteric artery. Careful handling of the pancreas is crucial, because rough handling leads to posttransplant pancreatitis.

The recipient operation begins with a midline transabdominal incision. Intraabdominal placement of the graft is associated with a lower incidence of lymphoceles and peripancreatic fluid collection than retroperitoneal placement. The right colon, small bowel, and sigmoid colon are reflected medially and superiorly, allowing access to the iliac vessels. The ureters are mobilized bilaterally, and the common, external, and internal iliac arteries and veins are extensively mobilized. The pancreas usually is placed into the right iliac fossa, and the kidney, if transplanted simultaneously, is implanted on the left side. The venous anastomosis is performed first. A tension-free venous anastomosis is essential to minimize the risk of vascular thrombosis. This can be achieved by anastomosing the portal vein of the pancreas graft to the distal inferior vena cava or to a completely mobilized iliac vein, obviating the need for an interposition graft.

The arterial anastomosis is then performed between the reconstructed donor iliac artery “Y” graft and the common iliac artery. There is no need for administration of systemic heparin or low-molecular-weight dextran during the anastomoses, which may be associated with a higher risk of postoperative bleeding complications.

After completion of the vascular anastomoses, the clamps are released slowly to minimize a precipitous rise in pressure within the graft. Delayed graft reperfusion minimizes edema and blood loss and allows an opportunity to ligate bleeding pancreatic vessels. Shortly after release of the vascular clamps, the pancreas begins to secrete pancreatic juice into the closed duodenal segment. The duodenal segment should not be allowed to distend, because increased pressure may promote reflux into the pancreatic duct, resulting in pancreatitis. If bladder drainage is chosen, a pancreatic duodenocystostomy is performed in a side-to-side anastomosis to the dome of the bladder with two layers of running absorbable sutures (see Fig. 10-7). If enteric drainage is performed, the duodenal segment is sutured to the ileum in a side-to-side anastomosis with two layers of interrupted silk sutures (Fig. 10-9).

During the procedure, generous intravenous doses of mannitol and albumin are administered to limit reperfusion injury and tissue edema. The abdominal cavity is irrigated copiously with antibiotic solution, including amphotericin B, miconazole, and cephalothin, to minimize the risk of peripancreatic infection, abscesses, and mycotic aneurysms. The use of drains is not recommended. The patient is routinely euglycemic in less than 12 h postoperatively, without the need for postoperative insulin therapy.

Postoperative Management

Rejection occurs with greater frequency after pancreas and simultaneous pancreas-kidney transplantation than after isolated renal transplantation. This difference requires management that balances aggressive immunosuppression against the risks of infection. A quadruple drug regimen, including ATGAM or OKT3, mycophenolate mofetil, cyclosporine, or tacrolimus and steroids, has been the backbone of immunosuppressive therapy. Antibody therapy generally is administered intraoperatively and for 10 to 14 days postoperatively. Analysis of the *United Network for Organ Sharing (UNOS)* registry data failed to demonstrate an improvement in graft survival at one year for SPK recipients whose immunosuppression was induced with antibody therapy versus no anti-T-cell agents, but use of anti-T-cell therapy is standard at most institutions.

Acute rejection is the rule rather than the exception after pancreas transplantation, occurring in 70 to 80 percent of patients, and primarily occurring in the first year after transplantation. Patients treated for rejection are treated with high-dose steroids for two days; if there is no response, the patient is then treated with OKT3 or antithymocyte globulin (ATGAM). With early treatment, 90 percent of rejection episodes are reversed, although rejection remains the primary cause of graft loss.

Difficulty in achieving early diagnosis of rejection is the principal barrier to successful long-term function. Serial histologic studies of pancreas rejection have shown that lymphocytic infiltrates initially involve the exocrine portion of the gland and that islet cell tissue becomes involved later. As a result, a decrease in exocrine secretions occurs early and can be detected by a decrease in urinary amylase. A consistent drop in urinary amylase levels of more than 25 percent strongly suggests the possibility of rejection. Evaluation of pancreas biopsies obtained because of hypoamylasuria in pancreas transplant recipients revealed that only 55 percent of patients exhibiting hypoamylasuria had biopsy evidence of rejection. Although a stable urinary amylase level essentially rules out rejection, a decreasing urinary amylase measurement is not a reliable indicator of rejection. Because most physicians prefer to treat patients for rejection on the basis of histologic evidence, tissue diagnosis of rejection is considered the gold standard. Commonly used biopsy techniques include transcystoscopic pancreatic and duodenal needle biopsies, and percutaneous pancreatic biopsies performed with ultrasound guidance.

In SPK patients, rejection of the pancreas graft alone is unusual, occurring in less than 15 percent of all rejection episodes. Pancreas rejection more commonly occurs with kidney rejection. Because of this, diagnosis of rejection after SPK transplantation relies almost entirely on serum creatinine, β_2 -microglobulin and renal biopsy. In the setting of PAK or PTA transplantation, however, isolated pancreas rejection does occur. There is no ideal screening test for rejection of the pancreas allograft, as the creatinine level is for kidney rejection. Numerous serum and urinary markers of pancreas graft rejection have been evaluated experimentally, including serum immunoreactive anodal trypsinogen level, pancreas-specific protein level, serum amylase and lipase levels, urinary cytology, pancreatic juice cytology, and neopterin level. Their clinical usefulness remains to be definitively demonstrated. Certain findings, such as decreased endogenous insulin levels and hyperglycemia, occur late enough to preclude their use in monitoring for allograft rejection. The clinical signs of rejection, such as fever, graft tenderness, and malaise, are not specific enough to indicate a diagnosis of rejection, but these findings do alert the physician to initiate proper diagnostic studies. When graft rejection is diagnosed, treatment is influenced by previous rejection therapy, response to initial therapy, and the severity of rejection.

Complications

Despite improvements in technique, preservation, and patient selection, surgical complications are not uncommon and threaten the survival of the graft and the patient. Proper management of complications is critical to a successful outcome. The development of a urologic complication (Table 10-3) is the most frequent cause for readmission to the hospital after SPK transplantation performed with bladder drainage.

Postoperative gross hematuria usually is mild and self-limited, clearing within three days of surgery. If bleeding is not stopped with Foley catheter irrigation, cystoscopy is performed with fulguration of the bleeding point, which usually is located along the duodenocystostomy suture line. Hematuria occurring later in the postoperative course may be a result of a suture granuloma, urinary tract infection, or ulceration of the duodenal segment. If hematuria persists despite appropriate therapy, conversion to enteric drainage is indicated. Enteric drainage generally is not required for early postoperative hematuria, but it is required in approximately 33 percent of patients presenting with late or chronic hematuria.

A postoperative urinary leak can occur from three potential sources: the ureteral implantation site, the duodenocystostomy anastomosis, and the duodenal segment. With the extravesical modified Liche technique and the use of an intraureteral stent during surgery, leaks from the ureteral implantation site are rare. Early urinary leaks are almost always technical and usually are located at the bladder-duodenum suture line, but urinary leaks more than 4 weeks after operation occur at the lateral duodenal staple lines or as the result of an ulcer within the duodenal segment. Patients with urinary leaks present with sudden-onset lower abdominal pain, fever, leukocytosis, and increased serum amylase and creatinine levels. A significantly elevated serum amylase level (over 400 U/L) and the development of pancreatic ascites discernible on CT scan are highly suggestive of a urinary leak. The diagnosis is confirmed by a voiding cystourethrogram (VCUG). Whether a conventional Gastrografin VCUG or a ^{99m}Tc -DTPA VCUG is more sensitive or specific is being studied. For small, early contained leaks, treatment consists of Foley catheter bladder decompression for 2 to 3 weeks. Large leaks or leaks that recur after conservative therapy require exploration and repair of the anastomosis or enteric conversion. If duodenal ulceration is found, excision of the ulcer usually is performed in conjunction with the enteric conversion. Late leaks almost always necessitate enteric conversion.

Urinary tract infections are the most common infectious complication after pancreas transplantation, and are significantly more common after bladder-drained SPK transplantation than after renal transplantation alone. Pancreas transplant recipients may be more predisposed to urinary tract infections than renal transplant recipients because of the additive effect of several contributing factors, including alkalinization of the urine secondary to bicarbonate and exocrine secretions, presence of a diabetic neurogenic bladder with incomplete emptying, mucosal injury at the bladder anastomosis, and prolonged catheter drainage. A cause for therapy-resistant or recurrent infection should be sought with diagnostic cystoscopy. Most are caused by a foreign body (e.g., suture granuloma, bladder stone), which can be treated with cystoscopic removal. Conversion to enteric drainage occasionally is indicated for recalcitrant urinary tract infections.

Persistent urethritis can result in urethral stricture or disruption. Urethritis is most likely caused by the digestive action of pancreatic enzymes on the urothelium. Urethritis usually manifests as perineal pain and discomfort during urination, and it occurs almost exclusively in males. Treatment with Foley catheter drainage for several weeks is recommended. If perforation occurs, it usually is in the membranous portion of the urethra and presents with perineal and scrotal swelling in addition to pain, and occasionally in association with acute bladder outlet obstruction. To avoid complications of urethral stricture and disruption, early enteric conversion is recommended if urethritis fails to respond to a short course of

conservative treatment, but these complications are unusual, occurring in only 5 percent of SPK transplant recipients.

Early postoperative hyperamylasemia, thought to be from preservation injury, is not uncommon, is asymptomatic, and improves rapidly. Persistent or marked elevations indicate possible technical errors, including duct ligation or leak. An infrequent but potentially fatal complication occasionally associated with graft pancreatitis is the rapid development of acute respiratory distress syndrome (ARDS) after pancreas transplantation. Pancreases from obese donors are particularly prone to preservation- induced pancreatitis. If ARDS and sepsis develop, strong consideration should be given to immediate removal of the pancreas transplant. Late graft pancreatitis (sometimes referred to as reflux pancreatitis) is defined by the following elements: (1) sudden-onset lower abdominal pain, (2) elevation of serum amylase level, (3) absence of a leak, (4) evidence on CT scan of gland edema and retroperitoneal inflammation and peripancreatic inflammation, without evidence of abscess or fluid collections, and (5) resolution of symptoms within 24 h of Foley catheter drainage. Treatment with Foley catheter drainage for several days usually is successful.

A metabolic acidosis is present postoperatively in approximately 80 percent of patients after pancreas transplantation with bladder drainage, and usually is a result of excessive urinary loss of bicarbonate-containing exocrine fluids. Oral replacement should be initiated to maintain a serum bicarbonate level of at least 22 to 25 mg/dL. Urinary bicarbonate loss is accompanied by an obligate loss of fluid; therefore, low serum levels are associated with dehydration. This problem usually stabilizes and diminishes over time and only infrequently requires conversion from bladder to enteric drainage.

Peripancreatic fluid collections visualized by ultrasonography and CT scans are common and should be drained if there is a suspicion of infection. The incidence of intraabdominal sepsis after bladder-drained pancreas transplants may be less than in transplants performed with enteric drainage. Postoperative graft thrombosis is unusual after SPK transplantation in the uremic recipient, occurring in less than 1 percent of patients. The incidence of this complication after PAK and PTA transplantation in the nonuremic patient may be somewhat higher. Mycotic pseudoaneurysms of the vascular anastomoses occur on rare occasions and can be devastating.

Exocrine secretions of the pancreas transplant must be converted from bladder drainage to enteric drainage in 10 to 25 percent of pancreas transplant recipients. Common indications for enteric conversion are persistent hematuria and urinary leaks from the duodenal segment. The technique of enteric conversion involves taking down the duodenocystostomy, repairing the bladder with a two-layer closure, and performing a simple two-layer side-to-side duodenoileostomy at the terminal ileum (Fig. 10-10). It is helpful to distend the bladder intraoperatively to assist in identifying the duodenocystostomy suture line. If an ulcer of the duodenal segment is the cause, it should be excised in conjunction with enteric conversion.

Results

Between 1987 and 1995, over 4,500 cadaver donor pancreas transplants were performed in the United States and reported to the IPTR. The 1-year patient survival rate is more than 90 percent, and the 1-year pancreas graft survival rate (as measured by insulin independence) is more than 75 percent. The addition of a pancreas to a kidney transplant does not adversely affect patient or kidney graft survival rates in uremic diabetic patients. Eighty-seven percent of pancreas transplants are performed simultaneously with a kidney transplant in the United States. PAK and PTA transplants account for 7 percent and 5 percent of the reported transplants, respectively. The two most significant factors affecting graft survival are the circumstances in which the pancreas transplant occurs and the duct management technique used. When compared with PAK and PTA, SPK is associated with the highest graft survival rate and the lowest technical failure rate. The 1-year pancreas graft function rates for SPK, PTA, and PAK are 78, 55, and 56 percent, respectively. PAK, PTA, and SPK transplants have equally good pancreatic graft survival (better than 75 percent 1-year graft survival). The IPTR reported that bladder-drained pancreas grafts had a better 1-year graft survival rate than grafts whose ducts were managed by enteric drainage or duct occlusion. By 5 years posttransplant, however, this difference was lost and equivalent graft function could be predicted for both techniques. Another factor identified in the UNOS database that correlated with a worse outcome was retransplantation, which was associated with a risk of graft loss 1.5 times higher. Neither HLA mismatching, nor longer preservation times (up to 30 h), nor recipient age had a significant negative impact on graft survival in the UNOS analysis.

A review of 356 patients who received SPK transplants from 1989 to 1997 revealed that immediate insulin independence was achieved in 99.6 percent of patients. There was one primary nonfunctioning pancreatic graft, and the rate of pancreas graft vascular thrombosis was 0.8 percent. For the simultaneously transplanted kidney, postoperative hemodialysis was necessary in only 2.4 percent of patients. One-year pancreas graft and patient survival were 88 percent and 96.5 percent, respectively. Five-year actuarial survival rates in SPK recipients with bladder-drained grafts were 90 percent for patients, 82 percent for kidney grafts, and 80 percent for pancreas grafts (Fig. 10-11). Kidney graft survival in SPK recipients is comparable to that for isolated cadaveric renal transplant recipients. Long- term kidney function is not negatively affected by a simultaneous pancreas transplant. Because of the excellent results of mycophenolate mofetil in renal transplantation, the authors use it rather than azathioprine the immunosuppressive protocol for pancreas transplantation. Using mycophenolate mofetil–based immunosuppression with cyclosporine (Neoral) reduced the incidence of rejection in 1-year SPK recipients from 73 to 21 percent. OKT3 use for steroid-resistant rejection also fell significantly from 28 to 15 percent, and after 1 year no grafts were lost for immunologic reasons.

Effect of Pancreas Transplantation on Secondary Complications of Diabetes

The major benefit of a pancreas transplant over a kidney transplant alone is enhanced quality of life. Other advantages over kidney transplantation alone include reversal of neuropathies, prevention of diabetic nephropathy, and amelioration or stabilization of diabetic retinopathy. Studies evaluating the effects of pancreas transplantation on secondary complications have been prospective, comparative, nonrandomized trials of concurrent pancreas and kidney transplants versus kidney transplants alone in diabetic recipients, or pancreas-kidney recipients with a failed pancreas graft. These studies have consistently demonstrated improvement in pancreas recipients regarding most secondary complications and quality of life.

Diabetic nephropathy recurs almost universally in diabetic patients who receive an isolated kidney transplant, although recurrence of disease may not be evident for many years. A functioning pancreas graft prevents the recurrence of diabetic kidney disease in the renal transplant. Histologic studies of renal allografts in patients who have undergone SPK transplantation demonstrated no evidence of recurrent diabetic nephropathy for up to 4 years after transplantation. Serial posttransplant biopsies have shown that glomerulomesangial volumes in diabetic patients with kidney transplants alone were significantly greater than in patients with combined pancreas-kidney transplants.

After SPK transplantation, most patients have subjective improvement of neuropathic symptoms. Studies have demonstrated an improvement in patient symptoms of peripheral neuropathy and significant objective improvement in motor and sensory nerve conduction velocities up to 4 years after pancreas transplantation. The Diabetes Control and Complications Trial Research Group conducted a multicenter, randomized, single-blinded, controlled trial in diabetics to determine what effect intensive glycemic control had on the progression of neuropathy. They found that after 5 years, fewer patients receiving intensive insulin therapy than patients receiving usual care had clinical neuropathy confirmed by nerve conduction studies (10 percent versus 25 percent). The association of neuropathy with previous poor control and the improvement of conduction velocities observed after better glycemic control have persuaded most clinicians to strive for tight control. This is a compelling argument for achieving permanent long-term normoglycemia with pancreas transplantation. Subjective improvement in gastroparesis has been noted in several studies.

Improvement in visual impairment after pancreas transplantation has been suggested in one study. Regression of retinopathy was noted in 43 percent of SPK recipients, compared to only 23 percent of recipients of a kidney transplant alone. Furthermore, 27 percent of isolated kidney transplant recipients showed progression of disease, but vision deteriorated in only 7 percent of SPK recipients.

One of the most important advantages gained from pancreas transplantation is the improvement in overall quality of life. Several well- controlled studies have been carried out at centers in the United States and in Europe experienced at performing pancreas transplantation. Patients with functioning pancreas transplants were nearly unanimous in judging their overall health and quality of life as significantly improved by insulin independence. Isolated pancreas transplantation most likely will be offered with increasing frequency to diabetics with normal renal function to prevent or reverse end-organ complications and to prevent life-threatening hypoglycemic episodes.

INTESTINE

Anthony M. D'Alessandro

Attempts to transplant small intestine in human beings were made between 1960 and 1980. The initial excitement over this challenging mode of therapy for patients with short-bowel syndrome rapidly faded as technical, immunologic, and septic complications resulted in unacceptable morbidity and mortality rates.

During the same era, parenteral nutrition, which could deliver all necessary fluids and nutrients, became the standard treatment for patients with short-bowel syndrome. Total parenteral nutrition (TPN) has been shown to be an excellent treatment modality that has allowed many patients with intestinal failure to survive with a reasonable quality of life.

Most patients remain dependent on TPN indefinitely and develop complications, such as catheter-related sepsis, metabolic derangements, vascular access problems, and liver dysfunction. One year of TPN treatment in the United States is estimated to cost between \$50,000 and \$120,000. If intestinal transplantation could be performed with acceptable morbidity and mortality rates, it would offer a more physiologic and economic approach to the problem of short-bowel syndrome.

Potential Candidates

The majority of potential candidates for intestinal transplantation are patients with short-bowel syndrome. Different disease processes can lead to short-bowel syndrome, but adults and children generally develop this syndrome after extensive intestinal resections. Approximately two patients per million population will start TPN each year. One patient per million population per year develops irreversible intestinal failure and lifelong TPN dependency.

Before the development of TPN, the long-term prognosis for patients with short-bowel syndrome was poor. Studies have shown that 50 percent of TPN patients are children who have high mortality rates primarily associated with the complications of TPN-related cirrhosis and liver failure. The mortality rate in adults from liver failure with short-bowel syndrome is estimated to be 5 to 10 percent. The long-term prognosis of children with short-bowel syndrome also is considered to be worse than in adults. This is related to growth problems, vascular access difficulty, and problems with psychosocial development.

Common indications for intestinal transplantation in adults are: Crohn's disease, mesenteric thrombosis, and trauma. Necrotizing enterocolitis, intestinal pseudoobstruction, gastroschisis, volvulus, and intestinal atresia are indications in children. However, any disease process that leads to short-bowel syndrome and TPN dependence may be considered for intestinal transplantation.

Operative Procedures

There are three methods of intestinal transplantation, the choice depending on the disease and sequelae that occur in combination with the short-bowel syndrome. For patients without liver failure, isolated intestinal grafting is preferred; patients with liver failure receive a combined liver-intestine transplant. In a small number of patients, a multivisceral procedure is performed that includes grafting of the liver, stomach, pancreas, duodenum, and small intestine with or without the colon.

Establishing continuity between the transplanted intestine and the remaining native intestine with early feeding is important. Portal venous drainage of the transplanted intestine appears to offer physiologic advantages over systemic venous drainage. At the time of implantation, a proximal or distal transplanted intestinal stoma is placed for endoscopic monitoring and biopsies.

Donor Issues

There is a shortage of suitable intestinal grafts, especially in the case of combined liver-intestinal transplantation. Most clinicians consider an identical ABO blood type mandatory because of previously reported hemolytic reactions. However, there have been reports of ABO-compatible but not identical liver-intestinal transplants being performed without problems. Because many recipients are children, a limiting factor is the required size match between donor and recipient. The waiting time and deaths while on waiting lists are higher for potential liver-intestinal recipients.

Management and Pretreatment

The management of the multiorgan donor during which the intestine (with or without liver) is retrieved for transplantation is more complicated. Mechanical cleansing of the intestine is performed as well as antibiotic decontamination with erythromycin and neomycin. ALG/ATG or OKT3 has been used to pretreat the donor in order to prevent recipient graft-versus- host disease. Because of the high incidence of lymphoproliferative disorders in patients treated with antilymphocyte preparations, and because of better immunosuppression with tacrolimus (FK506), donor pretreatment is not used routinely.

Preservation

In most multiorgan donor procedures, the University of Wisconsin (UW) solution is used for preservation. Compared to other solid organs, the intestine appears to be more sensitive to ischemia and reperfusion injury. The UW solution has been shown to sustain good results for up to 16 h of cold ischemic time. The addition of oxygen-free radical scavengers to the preservation solution or at time of reperfusion may be beneficial, facilitating an early repair of small-bowel mucosa and improving immediate function of the gut. Core cooling and immersion of the small bowel in the hypothermic preservation solution are believed to be the most important preservation measures, but one study suggests that luminal flushout with preservation solution may be helpful. It is unclear how well the current solutions preserve this organ. Until the appropriate studies are performed, it is recommended that the intestine be transplanted within 12 h of retrieval.

Immunology

The two major immunological problems after intestinal transplantation have been graft-versus-host disease (GVHD) and host-versus-graft disease (rejection). To prevent GVHD, sufficient immunosuppressive therapy must be administered. Graft pretreatment has been effective in experimental models. Studies have demonstrated the beneficial effects of irradiation, T- cell subset depletion, and mesenteric lymphadenectomy. A pretreatment method that has been applied is the administration of ALG/ATG or OKT3 to the donor prior to organ removal. Ablation of donor cells may be detrimental, since these cells may be required for the development of microchimerism and tolerance. After intestinal transplantation, a substantial migration of lymphoid cells has occurred in both directions, from graft to host and vice versa. T lymphocytes are essential in graft-versus-host reaction. Reducing the number of lymphocytes by using shorter segments, irradiation, or poly/monoclonal antibodies affects the severity of GVHD and possibly reduces its incidence. The introduction of cyclosporine has reduced, but not totally abolished, the incidence of GVHD. Reports on the use of tacrolimus for intestinal transplantation suggest that this drug is more effective against GVHD. Clinically, there has been a low incidence of GVHD even without donor pretreatment.

The prevention of rejection after intestinal transplantation is more difficult. Treatment with prednisone and azathioprine is insufficient to prevent or overcome rejection in intestinal transplantation. The addition of cyclosporine demonstrated improved results only in experimental models. The introduction of tacrolimus significantly improved the outcome after intestinal transplantation. Other new and more specific immunosuppressive agents that are currently being tested in combination with prednisone and cyclosporine or tacrolimus are mycophenolate mofetil (CellCept), rapamycin, and deoxyspergualin.

Most clinicians do not use ATG or OKT3 induction therapy. Oral tacrolimus usually is begun soon after surgery (0.15 mg/kg twice daily) and increased to achieve levels between 10 and 15 ng/mL. Intravenous tacrolimus may be given (0.02 to 0.07 mg/kg/day) if oral absorption is inadequate. Because of increased nephrotoxicity with intravenous tacrolimus, prostaglandin E (0.2 to 0.6 mg/kg/h) should be given to decrease renal dysfunction. Methylprednisolone usually is given in a 6-day taper, the initial dose being 250 mg in infants, 500 mg in children, and 1 g in adults. Azathioprine, cyclophosphamide, and mycophenolate mofetil have been added in an attempt to decrease rejection episodes and prolong graft survival.

Diagnosis of Rejection

The earlier rejection is detected, the more effective might treatment be in reversing the rejection process and minimizing damage to the grafted organ. Rejection is detected primarily by clinical symptoms and graft histology. These symptoms include fever, abdominal pain, elevated white blood cell count, ileus, increased stomal output, gastrointestinal bleeding, and positive blood cultures. Intestinal biopsies may show evidence of cryptitis, shortening of villi, mononuclear infiltrate, or even mucosal sloughing. Histologic changes may confirm the diagnosis of rejection, but usually only after significant progression of the rejection process. Immunopathology may be a more sensitive method of detecting rejection, because it can characterize increased class II or IL-2 receptor expression as well as the phenotypes of cells infiltrating the transplanted intestine before histologic changes have occurred. These techniques are labor intensive and operator dependent. Loss of mucosal integrity is an early indicator of rejection, so measures of intestinal permeability may be of value and are simpler to perform. The isotope ^{99m}Tc-DTPA, which has a short half-life, has been used successfully to diagnose clinical intestinal transplant rejection. The isotope is given orally and the urinary excretion is measured 6 and 24 h after administration. A value greater than 5 percent indicates increased permeability and possibly an acute rejection episode.

Results

The International Intestinal Transplant Registry reported results on 178 intestinal transplants performed worldwide since 1985. The results of tacrolimus-based immunosuppression were superior to cyclosporine-based immunosuppression. The 1-year and 3-year actuarial graft survivals were 65 percent and 29 percent for isolated intestine, 64 percent and 38 percent for liver/intestine, and 51 percent and 37 percent for multivisceral transplants. Although the results have not reached the levels of kidney, pancreas, liver, and heart transplantation, they are comparable to the results achieved with lung transplantation.

LIVER

Stuart J. Knechtle and Munci Kalayoglu

Historical Background

Liver transplantation was first performed successfully in an experimental dog model by Welch in 1955. The first human liver transplants were performed by Starzl in 1963. The decades of the 1960s and 1970s included pioneering work by Starzl in Denver and Calne in England to develop the technique of liver transplantation. Survival rates generally were less than 50 percent at 1 year. In the 1980s, with the advent of cyclosporine-based immunosuppression and technical innovations, such as venovenous bypass and University of Wisconsin preservation solution, the results of liver transplantation improved markedly. The 1990s have been characterized by efforts to expand the donor liver pool using techniques such as living-related liver transplantation, split-liver transplantation, reduced-size liver transplantation, and the use of non-heart-beating donors. Increased knowledge of proper donor and recipient selection, surgical techniques, and immunosuppressive therapy have improved the results of human liver transplantation, with an overall 1-year patient survival rate of 79 percent in the United States.

Indications

Liver transplantation is indicated for the treatment of irreversible liver failure from acute or fulminant disease or, more commonly, chronic liver disease. Fulminant hepatic failure frequently has an unknown cause but may be secondary to viral hepatitis, Wilson's disease, hepatotoxins, or alcoholic hepatitis. Outcomes for liver transplantation in patients with fulminant hepatic failure have a worse prognosis than in patients with chronic liver disease, because the former are generally more unstable and have more comorbid conditions. Survival rates of 70 percent at 1 year are expected for patients transplanted for fulminant disease.

Table [10-4](#) lists the most common causes of liver disease in adults treated with liver transplantation. Posthepatic cirrhosis resulting from hepatitis B or C is associated with the risk of recurrence of viral hepatitis and cirrhosis in the transplanted liver. Strategies to prevent recurrence of hepatitis B have included administration of hepatitis B hyperimmune globulin, interferon, and lamivudine. Hepatitis C also has been associated with frequent recurrence of disease in the transplanted liver and is often unresponsive to interferon therapy. Liver transplantation for primary biliary cirrhosis is associated with a high success rate, but the primary disease may recur and is difficult to distinguish from chronic rejection. Primary sclerosing cholangitis should be treated with liver transplantation before cholangiocarcinoma develops.

Patients with alcoholic liver disease account for 75 percent of liver failure in the United States, comprising the largest group of patients who could potentially benefit from liver transplantation. The ethics of liver transplantation for alcoholic cirrhosis have been debated in view of the behavioral contribution to these patients' liver failure. In support of liver transplantation being applied to this group are the facts that these patients tend to have an excellent outcome (88 percent 1-year patient survival) and a remarkably low recidivism rate (< 10 percent). A period of abstinence and evidence of family and social support are required before the candidate can be eligible for transplantation. Comorbid features, such as alcoholic cardiomyopathy, must also be excluded.

Numerous metabolic defects, such as alpha 1-antitrypsin deficiency and amyloidosis, can be treated successfully with liver transplantation. Many inborn errors of metabolism, with their primary defect in the liver, also can be corrected by liver transplantation. Examples of these conditions are listed in Table [10-5](#) along with other common causes of end-stage liver failure in children.

Liver transplantation for cancer is controversial, and, in general, the results are poor compared to transplantation for benign disease. Five-year patient survival rates around 40 percent are reported, which is better than the outcome without transplantation. Hepatoma and cholangiocarcinoma are associated with poor long-term survival because of the spread of micrometastases outside the liver before liver transplantation. Immunosuppression usually favors the reemergence of underlying malignancies. Fibrolamellar hepatomas, hepatomas found incidentally after hepatectomy but unsuspected by pretransplant CT scan, and small (< 4 cm) hepatomas are associated with favorable outcomes. Low-grade tumors, such as islet cell tumors metastatic to the liver without evidence of extrahepatic disease, may be an exception to the observation that cancer recurs after transplantation. Upper abdominal exenteration with cluster transplantation of the liver, duodenum, and pancreas has not improved long-term survival of patients with upper abdominal malignancies, but may be considered for low-grade malignancies.

Preoperative Evaluation

The signs and symptoms of liver failure (Table [10-6](#)) should be evaluated in detail in patients being considered for liver transplantation. The decision regarding the timing of liver

transplantation depends on an appropriate synthesis of numerous factors, including the practical issues of length of waiting time and donor organ availability. The degree of urgency of liver transplantation dictates the patient's position on a liver transplant waiting list.

Hepatic encephalopathy is determined by clinical examination and is often paralleled by serum ammonium levels. Patients with stage IV coma must be managed aggressively to prevent cerebral edema or hemorrhage, common causes of death in patients with end-stage liver failure. Evaluation may include head CT or MRI scanning to monitor cerebral edema, and continuous intracranial pressure monitoring in an intensive care unit.

Coagulopathy associated with liver failure is associated with an elevated *international normalized ratio (INR)* unresponsive to vitamin K replacement. Thrombocytopenia is common and usually is caused by hypersplenism related to underlying portal hypertension. Patients with bleeding should be treated with administration of fresh frozen plasma and platelet transfusions.

Gastrointestinal bleeding related to underlying portal hypertension, if associated with Child's Class B or C liver failure and liver fibrosis or cirrhosis, constitutes an indication for liver transplantation. Variceal bleeding before transplantation should be controlled with a combination of medical, radiologic, and, if necessary, surgical therapy. Transjugular intrahepatic portosystemic shunts (TIPS) have emerged as an attractive temporizing measure to relieve variceal hemorrhage before liver transplantation. These patients usually have an improvement in their overall liver function and can be more easily managed before liver transplantation.

Ascites resulting from portal hypertension may be severe and require medical treatment with diuretics and paracentesis. TIPS also is an effective means to relieve ascites before liver transplantation. Severe ascites may be associated with umbilical and inguinal hernias, which can be repaired at the time of liver transplantation. Ascites may present as a pleural effusion from transudation of ascitic fluid across the diaphragm. This may require aggressive therapy with paracentesis or TIPS and can be expected to resolve completely after liver transplantation.

Hepatorenal failure, a fatal condition before the advent of liver transplantation, is reversible after liver transplantation. Other correctable causes of underlying kidney dysfunction must be excluded before hepatorenal failure is diagnosed. If hepatorenal failure has caused renal function to deteriorate acutely over a period of days or weeks, renal function can be expected to return to normal after liver transplantation. Patients with concomitant intrinsic renal disease not caused by hepatorenal failure should be considered for combined renal and hepatic transplantation. This is advisable only if all other organ systems have excellent function.

Biliary obstruction in patients with primary biliary cirrhosis is associated with fatigue and severe itching, which can be debilitating psychologically and physically. Severe itching and fatigue should prompt liver transplantation. Patients with primary sclerosing cholangitis who develop recurrent bouts of cholangitis requiring hospitalization should be considered for early liver transplantation to avoid the septic complications of cholangitis. The risk of cholangiocarcinoma developing in these patients should be evaluated with CT scan and *endoscopic retrograde cholangiopancreatography (ERCP)* with brushings for cytology.

Nutritional depletion also may be a reflection of advanced liver failure and must be considered in the timing of liver transplantation. Patients with primary sclerosing cholangitis (PSC) may have associated inflammatory bowel disease and should be evaluated with colonoscopy before transplantation.

Patients with liver failure are immunocompromised. Spontaneous bacterial peritonitis may present with evidence of generalized sepsis and peritonitis and can be treated with antibiotics. Other causes of peritonitis, such as a perforated viscus, must be excluded. Growth of multiple organisms from a paracentesis culture should prompt suspicion of a perforated viscus and requires surgical exploration. Patients with advanced cirrhosis often have a substantial loss of their nonparenchymal liver mass and reticuloendothelial system, rendering them highly susceptible to sepsis, particularly by translocation across the gut.

Patients with liver failure develop a hyperdynamic state, with elevated cardiac output and low systemic vascular resistance. In advanced disease, pulmonary hypertension may develop, associated with clubbing of the digits and right-sided heart failure. Severe pulmonary hypertension is a risk factor for cardiac-related intraoperative death and should be assessed preoperatively when suspected. Patients with left-sided heart failure and liver disease, in addition to being evaluated for atherosclerotic heart disease, should be evaluated for alcoholic cardiomyopathy, hemochromatosis, and amyloidosis.

Contraindications

Contraindications to liver transplantation are summarized in Table 10-7. Liver transplantation in patients with multisystem organ failure (three or more organ systems in failure) is associated with a high mortality rate. Sepsis outside the liver would contraindicate liver transplantation, although patients with a septic focus within the liver may appropriately be treated with liver transplantation (e.g., a liver abscess or cholangitis in a cirrhotic liver would not contraindicate transplantation). Noncompliance with immunosuppressive therapy and medical management is associated with a poor outcome and should exclude patients from consideration. Active alcohol or drug abuse are examples of noncompliance. Severe cardiac or pulmonary disease may be a contraindication to liver transplantation because of the associated high mortality rate. Disseminated cancer would be an absolute contraindication to liver transplantation. A patient with a history of malignancy being considered for liver transplantation must have no evidence of malignant disease at the time of liver transplant evaluation. At the authors' institution, patients with a known malignancy generally are observed for at least 2 years for no evidence of disease before being considered for liver transplantation.

Immunologic Considerations

Graft failure in liver transplantation usually is not because of immunologic rejection in compliant patients, but more frequently is because of primary nonfunction, recurrence of original disease (hepatitis or primary biliary cirrhosis), or biliary and vascular complications. The reason for the liver's relative protection from rejection compared to other organs is poorly understood. Rejection episodes are common during the first 3 months posttransplant (about 50 percent incidence) but usually are reversible with steroids or anti-T-cell agents.

ABO blood group compatibility between donor and recipient is required for liver transplantation, but this is not absolute. ABO-incompatible liver transplants have been performed successfully, but they are associated with a higher risk of graft failure. ABO-compatible but nonidentical liver transplants (e.g., blood type O donor to type A recipient) may be associated with a self-limited graft-versus-host reaction, leading to hemolysis of recipient red blood cells. If necessary, the anemia can be treated with transfusion of donor-type blood. The liver is known to be associated with a much lower risk of hyperacute rejection than other organs transplanted to sensitized recipients, although hyperacute rejection may occur. A T-cell crossmatch usually is not performed preoperatively for liver transplantation, and even a positive cytotoxic crossmatch is of marginal significance, because of the liver's resistance to hyperacute rejection. HLA matching, routine in kidney transplantation, is not a consideration in liver transplantation, because the shorter preservation times required make it a practical impossibility.

Donor Procedure, Procurement, and Preservation

The liver can be procured from a brain-dead donor as part of an en bloc procurement with the pancreas (Fig. 10-12) or as an isolated liver procurement. Care must be taken to preserve anomalous arteries, such as a right hepatic artery arising from the superior mesenteric artery or an accessory left hepatic artery arising from the left gastric artery. All hepatic arteries must be preserved and reconstructed as necessary in order to avoid infarction of the transplanted liver. Various procurement methods are acceptable, but the preferred technique is aortic and portal vein flushing with University of Wisconsin (UW) solution prior to hepatectomy, followed by ex vivo portal vein, celiac axis, and superior mesenteric artery flushing and cold storage on ice. Non-heart-beating donors have been used for liver procurement with good graft function, comparable to that of heart-beating donors. For non-heart-beating donors, an en bloc technique is used to remove the entire abdominal viscera during aortic perfusion. After excision of the abdominal organs, the portal vein, celiac axis, and superior mesenteric artery orifices are flushed with UW solution.

The donor liver may be biopsied and submitted for frozen-section analysis to exclude severe hydropic change or fatty infiltration. The presence of more than 40 percent macrovesicular fat replacing the hepatic parenchyma, severe microvesicular fat, or severe hydropic change have been found to be predictive of primary nonfunction of the liver after transplantation. Biopsy information is useful in excluding these livers from transplantation.

Although the UW solution is capable of preserving livers for up to 24 h using cold storage on ice, an increased incidence of biliary strictures and primary nonfunction can occur with long preservation times. Other factors influencing the length of time the liver can be safely preserved include age of the donor, stability of the donor, injury to the donor liver reflected by elevated liver enzymes, and degree of fatty infiltration of the donor liver.

Recipient Operative Procedure

Orthotopic liver transplantation begins with native hepatectomy (including removal of a segment of the intraabdominal inferior vena cava) followed by implantation of the donor liver (Fig. 10-13). Because this technique requires occlusion of the inferior vena cava and portal vein simultaneously during the entire anhepatic phase, this method was found to result in hemodynamic instability in a significant proportion of adult patients. Venovenous bypass was developed to return blood from the inferior vena caval and portal venous circuits to the superior vena cava during the anhepatic period (Fig. 10-14). Venovenous bypass may be used routinely or selectively in patients who manifest instability after a trial of clamping.

The so-called "piggyback" technique (Fig. 10-15) is a variation of the conventional orthotopic liver transplant technique, in that the native inferior vena cava (IVC) is left intact after the native liver is dissected from it. All hepatic vein branches to the IVC are ligated except the three principal hepatic veins, which are then used to anastomose the donor suprahepatic IVC. The advantages of this technique are that it eliminates one of the vascular anastomoses of the conventional technique, namely, the infrahepatic inferior vena cava, and that inferior vena caval blood flow can be restored after completion of the suprahepatic caval anastomosis of the donor liver. Consequently, venovenous bypass rarely is necessary when the "piggyback" technique is used.

A donor liver procured from an adult may be reduced in size as necessary for transplantation to a pediatric recipient. The largest portion of the liver that would fit into the recipient usually is selected. If necessary, the left lateral segment (segments II and III) can be resected from the donor liver, leaving the hepatic artery, portal vein, and inferior vena cava intact for transplantation into an infant (Fig. 10-16). An additional modification is the split-liver technique, which uses the entire liver for two recipients; the right and left lobes are used for different patients. After reconstruction of the hepatic artery with the donor iliac artery graft and reconstruction of the portal vein with the donor iliac vein graft, segments II and III are transplanted into a child. The donor vessels supplying the liver remain with the right lobe and the medial segment of the left lobe (Fig. 10-17), which are transplanted into an adult. Application of these surgical techniques to the left lateral segment of a live donor has resulted in successful living-related liver transplantation, typically from a parent to a child with liver failure. Segments II and III usually are procured for this purpose. This procedure can be performed with minimal risk to the donor and with results for the recipient similar to those achieved using cadaveric liver donors.

Heterotopic and auxiliary liver transplantation have been used for fulminant hepatic failure, but the most common procedure for fulminant liver failure remains orthotopic liver

transplantation. The advantage of heterotopic or auxiliary liver transplantation (the latter involving a native lobectomy with replacement by the corresponding lobe of the donor liver) is that, should the native liver recover, the patient can be weaned from immunosuppressive drugs, allowing the transplanted liver to atrophy or be removed surgically.

Reconstruction of the common bile duct in liver transplantation involves an end-to-end anastomosis of donor-to-recipient bile ducts. This has been performed over a T tube, although using an internal stent, or no stent at all, has become popular. If the recipient common bile duct is inadequate or unsuitable for any reason, a Roux-en-Y choledochojejunostomy is performed.

If the recipient portal vein is unsuitable for anastomosis to the donor liver, usually because of portal vein thrombosis, a donor iliac vein graft can be interposed between the superior mesenteric vein of the recipient and the donor portal vein. This is an acceptable technique, with results equivalent to an end-to-end anastomosis of the donor and recipient portal vein (Fig. 10-18).

The donor hepatic artery is reconstructed by anastomosing the donor celiac axis to the recipient hepatic artery. If the recipient hepatic artery is compromised (including intrinsic or extrinsic stenosis of the celiac axis), a donor iliac artery graft is placed on the aorta in the supraceliac or infrarenal position and used as a conduit to the donor celiac artery.

Postoperative Management

The immediate postoperative management of liver transplant patients includes the twin goals of optimizing the patient's physiology and optimizing conditions that favor good liver function. This includes avoidance of vasoconstrictive inotropic agents, such as epinephrine, norepinephrine, phenylephrine, or high-dose dopamine, because these agents reduce blood flow to the liver. Every effort is made to avoid sepsis and hypotension, because of their deleterious effect on an already stressed liver as well as their systemic effects. Prostaglandin E₁ infusions may be useful for increasing renal and hepatic perfusion in the immediate postoperative period. Diuretic agents are useful in the first 24 to 48 h postoperatively to help patients mobilize the fluids sequestered as a result of cirrhosis.

Maintenance immunosuppression in liver transplant recipients relies principally on cyclosporine or tacrolimus. Tapering doses of steroids also are used. Induction with ALG or OKT3 is optional but may be useful in patients recovering from hepatorenal failure, because it allows a delay in starting nephrotoxic agents such as cyclosporine or tacrolimus. Azathioprine or mycophenolate mofetil also have been used as maintenance agents.

Acute rejection episodes after liver transplantation are common, but graft loss from rejection is rare. An elevation in liver enzyme levels, particularly canalicular enzymes (gamma-glutamyl transferase [GGT], alkaline phosphatase, and bilirubin), that is not explained by bile duct obstruction or hepatic artery thrombosis should prompt percutaneous liver biopsy, because the diagnosis of rejection is best made histologically. Rejection typically responds to bolus steroid therapy, but steroid-resistant rejection can be treated with antilymphocyte globulin or OKT3. In addition, elevations in liver function tests may occur months or years posttransplant from recurrent disease (e.g., primary biliary cirrhosis or viral hepatitis), de novo hepatitis, or chronic rejection. Chronic rejection may present as pruning of the hepatic arterial tree or as late hepatic artery thrombosis.

Complications

Liver transplant recipients are at risk for the same postoperative complications as any other patient with a major intraabdominal procedure, but several complications are unique to liver transplantation (Table 10-8). Primary nonfunction of the liver is manifested by a high INR, low fibrinogen level, and high ammonia level in the first several days posttransplant. Aspartate transaminase (AST) and alanine transaminase (ALT) may be markedly elevated initially, but they typically fall even in the presence of primary nonfunction simply because the hepatocytes have already died. Patients develop hepatic encephalopathy progressing to coma, and hepatorenal failure. The liver does not make bile or makes a clear-colored bile. Primary nonfunction must be treated with urgent retransplantation, but livers that demonstrate initial poor function typically recover after a period of days.

Portal vein thrombosis is a rare complication of liver transplantation but requires immediate diagnosis and operative intervention. Manifestations may include a marked rise in the serum ammonia level or variceal bleeding as a result of acute portal hypertension. It generally can be diagnosed by Doppler ultrasonography, which shows loss of the portal venous signal. Portal vein thrombosis should prompt immediate return to the operating room for thrombectomy, restoration of portal flow, and evaluation of the cause of thrombosis.

Hepatic artery thrombosis has an incidence of approximately 5 percent in adult liver transplantation and a higher incidence in pediatric liver transplantation. Early hepatic artery thrombosis, occurring within the first month posttransplant, usually is a result of technical problems, whereas later hepatic artery thrombosis is probably related to immunologic injury of the hepatic arterial tree. Because the biliary tree is dependent on hepatic artery blood flow, hepatic artery thrombosis results in ischemic changes of the bile ducts, resulting grossly in sloughing of the biliary epithelium and leading to plugging and obstruction of the bile ducts. If uncorrected, this leads to biloma formation and eventually liver abscess and sepsis. The management of hepatic artery thrombosis includes retransplantation or observation, since some patients tolerate this complication without ill effects. Most patients, however, develop biliary sludging with strictures and biloma formation. Bilomas can be drained percutaneously, and strictures of the common bile duct can be treated with resection and Roux-en-Y drainage.

Bile leaks after liver transplantation must be corrected immediately, because they lead to peritonitis, sepsis, and graft loss. A common complication of T-tube removal in liver transplant recipients is a bile leak from the T-tube exit site, because steroid therapy often prevents formation of a fibrous tract along the T tube. Bile leaks can be managed with operative intervention or by ERCP with papillotomy and temporary stenting of the common bile duct.

Recurrence of original disease may be a problem after liver transplantation in particular groups of patients. This includes primary biliary cirrhosis, which is difficult to differentiate from chronic rejection. Hepatitis B and C are likely to recur after liver transplantation but will not necessarily lead to cirrhosis of the transplanted liver. If alcoholism recurs after liver transplantation, it is considered a contraindication to retransplantation. Hepatoma or cholangiocarcinoma recurring in a transplanted liver contraindicates retransplantation.

Posttransplant lymphoproliferative disorder (PTLD) may arise in liver transplant recipients as a side effect of overimmunosuppression. Particularly at risk are pediatric recipients, especially those treated with high-dose immunosuppression for recalcitrant rejection. The primary treatment of PTLD is reduction of immunosuppressive therapy.

Results

Patient and graft survival with liver transplantation varies with institutional experience as well as comorbid conditions in the recipients before transplantation. Patients in better medical condition at the time of liver transplantation have better outcomes, which has prompted earlier referral of patients with liver failure. Combined patient and graft survival rates for U.S. centers are shown in Fig. 10-19. Pediatric graft and patient survival rates are shown in Fig. 10-20. Certain indications, such as primary biliary cirrhosis in adults and biliary atresia in children, are associated with higher- than-average success rates. The longest-living liver transplant patient is 26 years posttransplant.

Future improvements in liver transplantation include identification of patients who are immunologically tolerant of their livers and do not require long-term immunosuppression. Strategies to develop and monitor tolerance will likely be applied to liver transplantation, particularly since the liver appears to be less susceptible to immunologic rejection than other organs. The means of providing artificial liver support for patients in liver failure are urgently needed as short-term bridges to liver transplantation and as temporary support for patients whose livers might recover. Efforts in this direction include the use of extracorporeal porcine liver perfusion using genetically modified pigs, and the development of cultured human hepatocyte systems. The use of hepatic xenografts has not met with long- term success but remains a challenge with increasing appeal as the discrepancy between donor supply and recipient demand increases.

THORACIC ORGANS

David D. Yuh and Bruce A. Reitz

Transplantation of the human heart and lungs occupies a prominent place in the pantheon of medicine's greatest achievements in the past century. The deliberate, painstaking development of what was once considered a mystical concept into the lifesaving operations performed today stretched the limits to which we would go to renew the lives of those suffering from end-stage cardiopulmonary disease. Over the past 60 years the pioneering efforts of Carrel, Demikhov, Shumway, and others questioned long-standing assumptions and theories upon which conventional, palliative therapeutics were based. Their efforts forced physicians to seriously consider cardiopulmonary transplantation as a potential curative procedure. The decades since the first successful human cardiac transplant, performed by Christiaan Barnard in 1967, have witnessed the development of lung, combined heart-lung, and pediatric transplantation, expanding the range of potential recipients who would benefit from cardiopulmonary replacement. Despite steady experimental and clinical progress, significant problems remain, particularly donor organ shortage, chronic graft rejection, opportunistic infection, and limited organ preservation techniques.

Heart Transplantation

Historical Background

Experimental cardiac transplantation was launched in 1905 at the University of Chicago, where Alexis Carrel and Charles Guthrie, in the process of developing new vascular surgical techniques, transplanted a puppy's heart into the neck of an adult dog. Although the graft displayed normal contractile function for only 2 hours, this experiment established the technical feasibility of heart transplantation. In the mid-1940s the Russian surgeon V. P. Demikhov described a series of innovative anatomic variants of intrathoracic heterotopic cardiac transplantation in the classic monograph *Experimental Transplantation of Vital Organs*. In 1953 Neptune and colleagues described the use of hypothermic circulatory arrest for the purposes of graft preservation and recipient protection by placing donor and recipient dogs in beverage coolers. Using early techniques of "refrigerated" graft preservation and cardiopulmonary bypass, Webb and colleagues reported 12 successful canine orthotopic cardiac transplants in 1959, with the recipients surviving for periods ranging from 30 to 450 min. In these experiments, donor–recipient anastomoses were constructed among the aorta, pulmonary artery, and each of the pulmonary veins, requiring graft ischemic times ranging from 2 to 4 h.

The first successful orthotopic cardiac transplants were performed by Shumway and Lower at Stanford University in 1960. In a landmark *Surgical Forum* presentation, Shumway and Lower described a simple technique of orthotopic heart transplantation in canines through anastomosis of the atrial cuffs, pulmonary artery, and aorta, obviating the need for individual pulmonary venous anastomoses and shortening the graft ischemic time to about 1 h. Coupled with recipient protection using cardiopulmonary bypass and surface cooling as well as graft preservation with profound topical hypothermia, this technique resulted in extended recipient survival times, ranging from 6 to 21 days in five of eight dogs, with full recovery and return to normal activity until death by homograft rejection. In later studies Shumway and colleagues described electrocardiographic voltage drops associated with acute cardiac allograft rejection in dogs, the reversal of such changes with the administration of azathioprine and methylprednisolone, and consequent prolongation of recipient survival. Subsequent experimentation laid the groundwork for clinical cardiac transplantation by defining transplant physiology and by developing new methods of graft monitoring and preservation.

On December 3, 1967, the first successful human heart transplant was performed at the Groote Schuur Hospital in South Africa. Christiaan Barnard and his colleagues transplanted the heart of a 24-year-old automobile accident victim suffering from severe brain injury into a 54- year-old diabetic man with severe coronary artery disease and repeated myocardial

infarctions. The recipient survived for 18 days before expiring from *Pseudomonas pneumonia*. Subsequent heart transplants performed by Kantrowitz, Shumway, and others generated worldwide enthusiasm for the procedure. The number of centers performing human cardiac transplants grew exponentially in the years that followed. Only one year after Barnard's much-heralded operation, over 100 cardiac transplants had been performed in 17 countries.

The celebration of mankind's triumph over end-stage heart disease was short-lived. Numerous postoperative complications stemming from acute rejection and infection led to a 15 percent 1-year survival rate, driving morbidity and mortality rates to nearly unacceptable levels. Significant advances in the diagnosis and management of acute cardiac allograft rejection were made when Caves designed a biotome for safely obtaining serial transvenous endomyocardial biopsies and Billingham developed a clinically relevant histologic scheme for grading rejection in these specimens. Immunosuppressant regimens were refined, including the use of rabbit antithymocyte globulin (RATG) in the treatment of acute rejection. From 1974 to 1981 Stanford's clinical cardiac transplant program, involving 140 patients, achieved 1-year and 5-year postoperative survival rates of 63 percent and 39 percent, respectively. Despite these vital advances, it was the use of cyclosporine in heart transplant recipients at Stanford in 1980 that launched the rebirth of cardiac transplantation as the 1-year and 5-year survival rates surged to over 80 percent and 60 percent, respectively. The total number of cardiac transplants performed worldwide grew to more than 30,200 in over 200 centers by 1995.

Preoperative Considerations for Cardiac Transplantation

Recipient Selection

Rigid adherence to recipient selection criteria is important in achieving the excellent results observed in cardiac transplantation. Advances in the medical management of recipients before and after transplantation have led to a relaxation of certain criteria, including upper age limits, concomitant disease, and level of disability. Although a broader spectrum of patients has benefited from this expansion of eligibility, the problem of donor organ shortage has been exacerbated. Varying patient management philosophies and growing concerns about outcome statistics contribute to the problem. The United Network for Organ Sharing (UNOS) heart transplant waiting list contains over 2,800 patients, with about 300 new patients added to the list each month. Donor organ availability permits only 150 to 160 cardiac transplants each month. The average length of time on the waiting list has increased to over 300 days for outpatients, contributing to the 15 to 20 percent mortality rate among patients on the waiting list. Recent efforts have been directed toward refining and standardizing recipient evaluation methods and selection criteria to channel heart grafts to those patients who are in most immediate need while maintaining good short-term and long-term outcomes. There are several published reviews and recommendations regarding patient selection criteria, from the American Heart Association (AHA, 1992), the American College of Cardiology (ACC, 1993), and the International Society of Heart and Lung Transplantation (ISHLT, 1993). The selection criteria discussed in this section have been adopted by the majority of transplant centers in the United States.

Indications

Generally accepted indications for cardiac transplant evaluation are listed in Table 10-9. Patients who suffer from severe cardiac disability despite maximal medical therapy but who are otherwise healthy are considered for cardiac transplantation. Most cardiac transplant recipients suffer from end-stage, inoperable coronary artery disease or idiopathic cardiomyopathy (Fig. 10-21) and often require multiple hospitalizations. Other diagnoses include defined cardiomyopathy (e.g., viral, postpartum, familial), congenital anomalies, and valvular disease. Disabling symptoms typically include those associated with congestive heart failure (e.g., dyspnea, orthopnea, generalized edema, weakness), although recurrent symptomatic ventricular arrhythmias and severe ischemic symptoms (i.e., unstable angina) are frequently observed. Cardiac transplant candidates generally fall into the New York Heart Association's (NYHA) functional classes III and IV.

Formerly, a left ventricular ejection fraction (LVEF) of less than 20 percent was relied upon as a key indicator of severe cardiac dysfunction requiring transplantation, but refinements in medical management, particularly aggressive vasodilator therapy, have rendered this parameter less representative of severe patient disability or predictive of imminent death.

The NYHA functional classification scheme is widely used in the assessment of transplant candidates, but studies have shown that a more accurate and objective measure of functional capacity is provided by peak oxygen consumption during metabolic stress testing. Peak oxygen consumption, a function of peak cardiac output and peripheral oxygen extraction, correlates well with functional class and is an independent predictor of outcome in heart failure patients. Prospective studies have shown that patients with severely reduced peak oxygen consumption (< 15 mL/kg/min, approximately 50 percent of normal) have a 1-year mortality rate exceeding 50 percent and derive tremendous benefit in survival and functional capacity from transplantation. Oxygen consumption treadmill exercise testing is performed routinely or on an individualized basis for transplant evaluation in the majority of transplant programs in the United States.

Contraindications

There are several well-established contraindications to cardiac transplantation, which are based on significant risks of recurrent cardiac dysfunction, severe symptomatic limitations, and limited rehabilitation or survival after transplantation (Table 10-10). Contraindications include advanced age, irreversible hepatic, renal, or pulmonary dysfunction, active or extrapulmonary infection, severe peripheral or cerebral vascular disease (e.g., diabetes mellitus), unresolved malignancy, severe pulmonary hypertension, psychiatric illness or a history of medical noncompliance, drug or alcohol abuse, current tobacco use, and cachexia or morbid obesity. Some of these conditions represent relative or temporary contraindications, and exceptions can be made on a selective basis.

The original upper age limit of 50 years for cardiac transplant recipients has been raised to 60 or 65 years because of several reports that short-term survival rates and improvements in quality of life for recipients older than 60 years of age are comparable to those of younger recipients. Few transplant centers would disqualify an otherwise favorable candidate solely on the basis of advanced age, but it is generally agreed that older transplant candidates should undergo a more rigorous evaluation. Bull and associates demonstrated that short-term and long-term survival after transplantation was significantly lower for patients who were older than 60 years of age at transplantation than for younger patients. Their studies also showed that older patients were more likely to die of late infectious complications or malignant disease after transplantation. The authors hypothesized that this may be from decreased immune responsiveness and T-cell function, leading to relative overimmunosuppression. Older transplant recipients also are at higher risk for developing steroid-induced diabetes and clinically significant osteoporosis.

Patients presenting with significant irreversible hepatic, renal, or pulmonary dysfunction should not be considered for transplantation because of the high perioperative risks associated with these conditions. Severe chronic obstructive pulmonary disease usually is defined as a forced expiratory volume in 1 second (FEV₁) of less than 1 L or 50 percent of the predicted normal value in the absence of pulmonary edema. In the case of hepatic and renal insufficiency, it is important for the evaluating clinician to determine whether the dysfunction is the result of severe chronic heart failure (i.e., venous congestion, hypoperfusion) or disease intrinsic to the affected organ system. When renal and hepatic function markedly improves or even normalizes with the optimization of cardiac output through the administration of fluids, diuretics, vasodilators, or inotropes, cardiac transplantation may be a therapeutic option. Some transplant centers perform renal or hepatic transplantation in selected patients with irreversible disease in conjunction with cardiac transplantation. Diabetes mellitus, a controversial contraindication to transplantation because of the potential for peripheral vascular and end-organ disease, susceptibility to infection, and exacerbation by steroids, does not appear to have adverse effects on outcome of cardiac transplantation in carefully selected cases.

Active infection and malignancy are absolute contraindications to transplantation in view of the lifelong immunosuppression required. Acute transient infections must be thoroughly cleared before transplantation; chronic infective agents, including chronic hepatitis B, hepatitis C, and human immunodeficiency virus (HIV) preclude transplantation. Chronic conditions predisposing to serious infection, including symptomatic cholelithiasis, severe diverticulitis, active peptic ulcer disease, and cerebral/pulmonary embolization, should be evaluated and treated before transplantation. With the exception of fully resected squamous cell carcinoma of the skin, patients with previous malignancies should not be listed for cardiac transplantation less than 5 years after the malignancy has been considered cured. An exception to this rule is young patients suffering from doxorubicin cardiotoxicity.

Severe, fixed pulmonary hypertension has been confirmed recently as a significant independent risk factor for early mortality after orthotopic cardiac transplantation because of a heightened incidence of acute posttransplant right ventricular failure. Retrospective studies showed that early posttransplant mortality rates were significantly higher in cardiac transplant recipients with a preoperative pulmonary vascular resistance of 5 Wood units or more or a transpulmonary pressure gradient (mean pulmonary artery pressure minus mean pulmonary capillary wedge pressure) of 15 mmHg or more. In some cases pulmonary vascular resistance can be lowered to acceptable levels with the aggressive administration of supplemental oxygen, vasodilators, and inotropic agents. Patients with pulmonary hypertension refractory to these measures should not be considered for orthotopic cardiac transplantation. The use of proportionally large donor heart grafts with short ischemic times to overcome significant pulmonary hypertension in marginal recipients has not been an effective approach and is not recommended. Patients with severe, refractory pulmonary hypertension should be considered for combined heart-lung transplantation.

The psychosocial assessment of a potential recipient's emotional stability, compliance, and socioeconomic environment is vital, because substantial emotional, medical, and financial demands are placed upon the recipient and his or her family. Active substance abuse, including drugs, alcohol, and tobacco, is an absolute contraindication to transplantation; abstinence for no less than 6 months must be documented before transplantation is performed. Morbid obesity (> 150 percent of ideal body weight) is often considered an indication of noncompliance. Obesity increases the risks for posttransplant complications, including diabetes, fractures, and hypertension. Cachexia suggests inadequate nutrition, predisposes to infection, and should be treated with nutritional supplementation preoperatively.

Evaluation and Management of Patients Awaiting Cardiac Transplantation

Much of the health-care cost incurred by cardiac transplantation is generated by the preoperative evaluation and, in many cases, the hospitalization of patients awaiting transplantation. Consequently, much attention has been paid to refining transplant evaluation protocols and medical management strategies to reduce these costs while maximizing the therapeutic benefits of cardiac transplantation.

Candidate Evaluation and Listing

An outline of screening and diagnostic tests used in the formal evaluations of potential heart transplant candidates is presented in Table 10-11. Most of these tests are mandatory and are conducted on an outpatient basis. Some patients in severe heart failure require hospitalization (i.e., monitoring, intensive medical therapy, mechanical circulatory support) and in-hospital evaluation.

Patients suitable for cardiac transplantation are categorized and listed on the basis of clinical status, time on the waiting list, body size, and ABO blood group. Clinical status is composed of two broad status classifications developed by UNOS. The status I designation is applied to patients who (1) require an intensive-care-unit setting receiving parenteral inotropic drugs or mechanical device support (e.g., intraaortic balloon pump, ventricular assist device, ventilator) to maintain adequate circulatory or ventilatory function or (2) are less than 6 months old. Status II comprises all other waiting patients. Depending upon the patient's condition, selected tests are performed, particularly peak oxygen consumption and hemodynamic measurements, and are repeated about every 6 months. Most listed patients continue to deteriorate over time, but a significant number stabilize or even improve with medical therapy, prompting their inactivation on or removal from the list.

Medical Management

The management of listed patients has two primary goals: (1) to relieve debilitating symptoms stemming from end-stage congestive heart failure, and (2) to preserve organ function and optimize the patient's condition for transplantation. Clinically stable transplant candidates with end-stage heart failure usually are treated with a combination of digoxin, diuretics, vasodilators, and angiotensin-converting enzyme inhibitors on an outpatient basis. Oral anticoagulation therapy is used by many transplant programs to reduce the risk of pulmonary and systemic emboli, which may occur asymptotically in as many as 60 percent of patients with idiopathic cardiomyopathy. Anticoagulation therapy is used selectively in patients predisposed to bleeding complications, including those in right ventricular failure with hepatic congestion. The use of antiarrhythmic agents as prophylaxis for ventricular arrhythmias and sudden cardiac death is controversial in view of their proarrhythmic potential and negative inotropic effects, particularly in patients with depressed left ventricular function. Amiodarone has improved survival in patients with end-stage heart failure without increasing perioperative risk, leading to its use in cardiac transplantation candidates. Patients with recurrent sustained episodes of ventricular arrhythmias refractory to antiarrhythmic medications are considered for implantation of an automatic implantable cardioverter/defibrillator (AICD) or a pacemaker cardioverter defibrillator (PCD).

When conventional therapies prove inadequate, more aggressive measures in the form of parenteral vasodilators (e.g., nitroprusside, nitroglycerin) and inotropic agents (e.g., dopamine, dobutamine) are used. Amrinone, a phosphodiesterase inhibitor with inotropic and vasodilator properties, has been used successfully in patients with ischemic cardiomyopathy. Its use is tempered by side effects, however, particularly its potential to produce or worsen ventricular arrhythmias.

Mechanical Support

Heart failure and clinical deterioration refractory to parenteral support necessitates mechanical intervention in the form of intraaortic balloon pump (IABP) counterpulsation or ventricular assist system (VAS) placement. IABP counterpulsation improves coronary perfusion and reduces afterload and mitral regurgitation; however, its overextended use is associated with significant rates of vascular complications. IABP is most appropriately used to deliver moderate, short-term support to patients with acute circulatory failure and end-organ dysfunction or patients with ischemia-related arrhythmias awaiting transplantation.

VASs are implanted mechanical pumping devices designed to assume a significant portion of the systolic work load from the left ventricle (LVAS), right ventricle (RVAS), or both ventricles (BiVAS) and are intended for intermediate to long-term circulatory support for patients in severe ventricular failure. Mechanical support is considered when the cardiac index is less than 2.0 L/min per m² body surface area, the ventricular filling pressures are greater than 20 mmHg, the urinary output is less than 20 mL/h (adults), and the systemic vascular resistance is greater than 2,100 dynes sec/cm⁻⁵ despite maximal parenteral inotropic and vasodilator therapy. Since the initial use of an LVAS as a bridge to transplant, LVASs have yielded encouraging results in transplant candidates. The three most widely used ventricular assist systems include the Novacor LVAS, the Thoratec VAS, and the HeartMate IP LVAS. All three systems have been used successfully to reduce the mortality rates among end-stage cardiac patients awaiting transplantation. The Novacor VAS and Thoratec VAS, used since 1976 and 1984, respectively, have bridged 60 to 65 percent of implanted patients to successful transplantation with 1-year posttransplant survival rates in excess of 80 percent; these survival rates are comparable to those of the general cardiac transplant population. A 55 percent decrease in pretransplant mortality rate and a 23 percent increase in 1-year survival rate was reported among patients placed on HeartMate IP LVAS support as a bridge to transplant. Complications observed in LVAS- supported patients include bleeding (40 percent), infection (20 to 75 percent), and right ventricular failure (10 to 30 percent). Bleeding in these patients often has been attributed to coagulopathy secondary to chronic heart failure and is ameliorated with the use of aprotinin during implantation. It is difficult to determine the actual incidence of significant device-related infections, because many postbypass patients have transient fevers and positive cultures, but LVAS patients have a low incidence of fatal infections. There are no reliable predictive factors for the development of right ventricular failure after LVAS implantation. The high mortality rates associated with this complication (30 to 50 percent) have prompted investigators to identify preimplantation criteria to determine which patients would benefit from biventricular versus left ventricular devices. Parameters such as ventricular size, filling pressures, and ejection fractions in addition to standard hemodynamic measurements are not reliably predictive of postimplantation right ventricular decompensation. It is suggested that a transplant candidate's clinical status rather than right ventricular function defines the need for additional right ventricular support. Patients requiring BiVAS are more compromised, with higher incidences of mental impairment, pulmonary edema, elevated creatinine levels, and lower mixed venous oxygen saturations than patients requiring LVAS. Perioperative blood transfusion requirements for diffuse bleeding are higher in patients requiring BiVAS.

Donor Selection and Management

Criteria

First and foremost, donors must have sustained irreversible brain death, usually as a result of blunt or penetrating head trauma or intracranial hemorrhage. Despite the development of a relatively streamlined organ procurement system in the United States, there remains confusion about brain death and declaration of brain death, because of variations in legislative language, procedures, and standards from state to state. However, absolute and optional criteria for brain death were established by a presidential commission in 1981.

Suggested criteria for cardiac donors and guidelines for recipient matching developed by the American Heart Association in 1992 are listed in Table [10-12](#). Donor evaluation consists of a directed history and physical examination, chest roentgenogram, 12-lead electrocardiogram, arterial blood gas determinations, echocardiogram, serologic screening (e.g., HIV, hepatitis B surface antigen [HB sAg], hepatitis C antibodies, herpes simplex virus, cytomegalovirus, and *Toxoplasma*), and pancultures. Normal cardiac function and the absence of a significant cardiac history and significant coronary atherosclerosis must be established. A donor age of less than 50 years is preferred, although potential donors aged 55 years and older are considered at most centers with a more detailed evaluation, sometimes including coronary angiography, to rule out significant cardiac disease. A higher risk of late graft atherosclerosis is associated with hearts obtained from donors over 40 years of age, and higher rates of early death have been noted in pediatric recipients of such hearts.

Absolute contraindications for donation include severe coronary or structural disease, prolonged cardiac arrest, prior myocardial infarction, a carbon monoxide–hemoglobin level greater than 20 percent, arterial oxygen saturation less than 80 percent, metastatic malignancy (sometimes excluding primary brain and skin cancers), and positive HIV status. Relative contraindications include thoracic trauma, sepsis, prolonged severe hypotension (i.e., mean arterial pressure less than 60 mmHg for more than 6 h), noncritical coronary artery stenosis, HB sAg or hepatitis C antibodies, multiple resuscitations, severe left ventricular hypertrophy, and a prolonged high inotropic requirement (e.g., dopamine in excess of 20 mg/kg/min for 24 h). It is important to rule out reversible metabolic or physiologic causes of impaired cardiac function, rhythm disturbances, and electrocardiographic anomalies (e.g., brain herniation, hypothermia, hypokalemia).

Despite these guidelines, the donor organ shortage coupled with critical clinical situations has prompted several groups to use cardiac grafts from “high-risk” donors with satisfactory short-term results. These include grafts that are potentially compromised by advanced age, high-dose inotropic support, donor-to-recipient undersizing, potential infection, echocardiographic abnormalities, or prolonged ischemic transport time.

Management

The primary goal of managing the cardiac donor is the maintenance of hemodynamic stability. Patients suffering from acute brain injury often are hemodynamically unstable, as a result of neurogenic shock/pulmonary edema, excessive fluid losses, and bradycardia. Continuous arterial and central venous pressure monitoring, aggressive fluid resuscitation, vasopressors, and inotropes usually are required. Judicious fluid management prevents intraoperative hemodynamic instability and minimizes the need for inotropes and vasopressors, which are myocardial stressors. Intravascular fluids should be given to maintain the central venous pressure between 5 and 12 mmHg. Diabetes insipidus is common in organ donors and requires the use of intravenous vasopressin (0.8 to 1.0 U/h) to reduce excessive urine losses. To maintain adequate perfusion pressures, dopamine is the standard inotropic agent, but alpha agonists (e.g., phenylephrine) often are appropriate. Blood transfusions should be used sparingly to maintain the hemoglobin concentration around 10 g/dL to ensure adequate myocardial oxygen delivery. Hypothermia should be avoided, because it predisposes to ventricular arrhythmias and metabolic acidosis.

Donor–Recipient Matching

Donor–recipient matching parameters include ABO compatibility and body size. ABO compatibilities are strictly adhered to because isolated episodes of hyperacute rejection have been observed in cardiac transplants performed across this barrier. Chan and colleagues report that body weight does not correlate well with heart size in adults weighing 50 to 100 kg, suggesting that strict size limits are not warranted. Although fairly wide limits are acceptable, size matching and graft ischemic time is particularly important for recipients with an elevated pulmonary vascular resistance (greater than 6 to 8 Wood units). Grafts from donors whose weight is less than 80 percent of that of the recipient or those with ischemic times longer than 2 h are avoided for adult and pediatric recipients with pulmonary hypertension. There is no upper donor size limit in adults, given the typically enlarged recipient pericardial space resulting from chronic congestive cardiac enlargement.

Once an appropriate donor–recipient pairing is made, the recipient is screened for preformed antibodies against a standardized panel of random donors. A panel-reactive antibodies (PRA) level greater than 5 percent prompts a prospective specific crossmatch between the donor and the recipient. Several retrospective studies demonstrated that the degree of donor–recipient HLA mismatching influences rejection rates and survival after cardiac transplantation, but HLA matching is not feasible on a prospective basis in thoracic transplantation.

Operative Procedures

Procurement

After the chest is entered through a median sternotomy, a retractor is placed and the pericardium is opened and tethered, creating a pericardial well (Fig. [10-22](#)). The heart is inspected and palpated for contusions, perforations, thrills, and coronary atherosclerosis. If the heart is deemed satisfactory, its acceptance is immediately communicated to the recipient team. The aorta and pulmonary artery are dissected superiorly to the level of the arch and bifurcation, respectively, to ensure adequate length for implantation. The superior vena cava is mobilized superiorly to the origin of the azygous vein and encircled with two ligatures, taking care to preserve the sinoatrial node. An adequate length of the inferior vena cava is dissected free from its pericardial reflection and is surrounded with an umbilical tape. The aorta is encircled with an umbilical tape, and a 14-gauge cardioplegia perfusion cannula is inserted into its ascending segment. Intravenous heparin is administered at a dose of 300 U/kg and allowed to circulate for 3 to 5 min.

Removal of the heart begins with ligation and division of the superior vena cava. The inferior vena cava is clamped and partially divided just above the diaphragm, and the heart is allowed to beat several times, until it is empty. The ascending aorta is then clamped distal to the perfusion cannula at the level of the innominate artery, and a hyperkalemic cardioplegic solution at 2 to 4°C is rapidly infused into the aortic root at a pressure of 150 mmHg, arresting the heart in diastole. Concurrently, topical cold saline at 4°C is poured onto the heart and into the pericardial well. When the heart is fully arrested, cooled, and perfused with cardioplegic solution, it is elevated from the pericardial well, and each of the pulmonary veins is divided at its pericardial reflection. The pulmonary artery is divided at the level of the bifurcation, and the aorta is divided at the level of the innominate artery. The explanted heart is placed into two sterile plastic bags with a cold saline interface. This in turn is placed in an airtight container filled with ice-cold saline solution and transported in a standard ice-filled cooler.

Orthotopic Transplantation

After anesthesia is induced in the transplant recipient and arterial and venous lines are placed, the supine patient's chest and groin areas are prepared and draped. Central venous access via the left internal jugular vein usually is obtained, sparing the right side for future endomyocardial biopsies. After a median sternotomy is performed and the pericardium is opened and suspended, the patient undergoes routine cannulation of the aorta and both venae cavae (Fig. [10-23](#)). The arterial cannula is inserted in the most distal aspect of the ascending aorta, just below the innominate artery. The venous cannulae are placed laterally in the right atrium and positioned in the superior and inferior venae cavae; caval tapes

are applied. After institution of cardiopulmonary bypass with moderate hypothermia (28 to 30°C) and snugging of caval snares, the ascending aorta is cross-clamped, and 50 to 100 mL of cardioplegic solution is infused rapidly into the aortic root, causing diastolic arrest. The atria are transected at the level of their atrioventricular grooves, excluding the atrial appendages, leaving two recipient atrial cuffs. The aorta and the pulmonary artery are then separated and divided at the level of the semilunar valve commissures.

The donor heart is placed in a bowl of cold saline solution, and the left atrium is opened by connecting the pulmonary vein orifices, fashioning the donor atrial cuff. The aorta and the pulmonary artery are completely separated from each other. Under continuous application of cold saline solution into the pericardial well, implantation begins with the direct anastomosis of the donor and recipient left atrial cuffs with a continuous 3-0 Prolene suture. On completion of this anastomosis, a cold saline infusion line is placed into the left atrium through the atrial appendage for continuous endocardial cooling and evacuation of air from the left heart.

The donor right atrium then is opened with an incision extending from the inferior vena cava orifice superiorly in a curvilinear fashion along the lateral atrial wall up to the base of the atrial appendage. Through this incision, an intact tricuspid valve and fossa ovalis are inspected and assured. It is important to close a patent foramen ovale because, in the setting of elevated right heart pressures, a significant right-to-left shunt might result in early posttransplant hypoxemia. The right atrial cuff anastomosis is performed with a continuous 3-0 Prolene suture. Systemic rewarming is initiated at this time.

The pulmonary artery and aortic anastomoses are then completed in an end-to-end fashion using continuous 4-0 Prolene sutures. The caval snares are released, the head of the bed is lowered, and topical cooling is halted, permitting blood to enter the heart and lungs; any air trapped in the left-sided chambers is displaced through an aortic needle vent. Lidocaine is infused into the bypass circuit, the aortic cross-clamp is removed, and de-airing procedures are continued to vent any residual air from the heart. The left atrial line is removed, and the hole is oversewn. Spontaneous fibrillation usually occurs at this time, and electrical defibrillation is effected as necessary. All suture lines are inspected for hemostasis before cardiopulmonary bypass is weaned. The superior vena cava cannula is drawn back into the right atrium, and the inferior vena cava cannula is removed just before bypass is discontinued. An isoproterenol infusion (10 to 75 ng/kg/min) is titrated to achieve a heart rate of 90 to 110 beats/min to maximize cardiac output chronotropically and inotropically and to lower pulmonary vascular resistance. Temporary ventricular and atrial pacing wires (donor right atrium) are placed. The pericardium is left open, and mediastinal chest tubes are placed. Pleural space violations or effusions are treated with chest tubes. The sternum and overlying fascia and skin are closed in the usual fashion.

The authors have used a bicaval anastomotic technique, using separate end-to-end superior vena cava and inferior vena cava anastomoses instead of the right atrial cuff anastomosis (Fig. 10-24). In addition to the functional benefits, separate caval anastomoses are particularly useful in cases of significant donor–recipient size discrepancies and transplantation for congenital heart disease.

Heterotopic Transplantation

Originally described by Demikhov, heterotopic transplants account for about 2.5 percent of the cardiac transplants currently performed. The operative technique bypasses the left heart and involves anastomoses between the left atria, aorta, pulmonary arteries, and donor superior vena cava to recipient right atrium (Fig. 10-25). The major indication for this operation is the presence of irreversible severe pulmonary hypertension, whereby the native right heart continues to work against the elevated pulmonary vascular resistance while the graft bypasses the left heart. Other indications include cases in which diminished donor heart function is anticipated because of size mismatch or prolonged ischemic time, and the graft is to serve as a temporary support in the setting of reversible cardiac failure.

Postoperative Management

Early Postoperative Period

On completion of the cardiac transplant procedure, the intubated transplant patient is transported to the intensive care unit (ICU) where cardiac rhythm and arterial and central venous pressures are monitored. The use of Swan-Ganz pulmonary artery balloon catheters usually is reserved for recipients with significant pulmonary hypertension. Strict isolation precautions, previously enforced to reduce the incidence of infection in these immunocompromised patients, are no longer required; simple handwashing and face masks are considered sufficient. Precautions are taken to minimize patient contact with objects or persons harboring active infectious agents.

A primary objective in the immediate postoperative period is to maintain adequate perfusion in the recipient while minimizing cardiac work. Approximately 10 to 20 percent of transplant recipients have some degree of transient sinus node dysfunction, often manifested as sinus bradycardia that usually resolves within a week. Because cardiac output is primarily rate-dependent after transplantation, the heart rate should be maintained between 90 to 110 beats/min during the first few postoperative days, using temporary pacing or isoproterenol. Although uncommon (< 5 percent), persistent sinus node dysfunction and bradycardia may require a permanent transvenous pacemaker. The systolic blood pressure should be maintained between 90 to 110 mmHg, using afterload reduction in the form of nitroglycerin or nitroprusside if necessary. Renal dose dopamine (3 to 5 mg/kg/min) frequently is used to augment renal blood flow and urine output. The adequacy of cardiac output is indicated by warm extremities and a urine output greater than 0.5 mL/kg/h without diuretics. Cardiac function generally normalizes within 3 to 4 days during which parenteral inotropes and vasodilators can be weaned. Hypovolemia, cardiac tamponade, sepsis, and bradycardia should be considered and treated expeditiously in the event of reduced cardiac output and hypotension.

Several factors may contribute to some form of depressed global myocardial performance in the acute postoperative setting. The myocardium is subject to prolonged ischemia, inadequate preservation, or catecholamine depletion prior to implantation. Right heart failure is not uncommon in the early postoperative period. Although considered multifactorial in origin, an elevated pulmonary vascular resistance often is the principal cause. Inotropes with pulmonary vasodilatory effects, including isoproterenol, dobutamine, and amrinone, are effective in treating early right heart failure. The more selective pulmonary vasodilator prostaglandin E₁, combined with standard inotropes, is effective in more severe cases. Rarely, mechanical right ventricular support is necessary.

Optimizing pulmonary function is another critical objective in the acute postoperative period. When the patient arrives in the ICU, an anteroposterior chest radiograph is obtained and the ventilator is set to a 100% fractional inspired oxygen content (FiO₂), tidal volume of 10 to 15 mL/kg, an assist-control rate of 10 to 14 breaths/min, and positive end-expiratory pressure (PEEP) of 3 to 5 cmH₂O. These settings are adjusted every 30 min to achieve an arterial oxygen pressure (PaO₂) above 75 mmHg with an FiO₂ of 40 percent, arterial carbon dioxide pressure (PaCO₂) from 30 to 40 mmHg, and pH from 7.35 to 7.45. Ventilatory weaning is initiated after the patient is deemed stable, awake, and alert. Weaning usually is accomplished through successive decrements in intermittent mandatory ventilation rate, followed by a trial of continuous positive airway pressure. When acceptable ventilatory mechanics and arterial blood gas levels are achieved, the patient is extubated, usually within the first 24 h postoperatively. After extubation, pulmonary care consists of supplemental oxygen for several days, aggressive pulmonary toilet, and serial chest radiographs.

Expedient removal of vascular lines has been shown to reduce the incidence of line-related infections. Pleural and mediastinal thoracostomy tubes are removed when drainage has fallen off to less than 25 mL/h, and pacing wires are removed after 7 to 10 days if pacing is not required. After several days, the patient is transferred from the ICU to a standard cardiac surgery ward, where patient instruction, immunosuppressant dosage adjustment, an initial endomyocardial biopsy, and early rehabilitation occur. A second endomyocardial biopsy and baseline coronary arteriogram are obtained approximately 2 weeks postoperatively, just before discharge.

Graft Physiology

The grafted heart presents several unique physiologic characteristics (Table 10-13). During procurement, the heart graft is separated from the sympathetic and parasympathetic cardiac plexus of nerves located between the tracheal bifurcation and the aortic arch. Normally this plexus autonomically regulates the heart rate, contractility, and coronary arterial tone. The denervated heart graft is isolated from normal autonomic regulatory mechanisms. The resting heart rate is higher, because vagal tone, sinus arrhythmia, and carotid reflex bradycardia are absent. The denervated heart graft develops an increased sensitivity to catecholamines, possibly from an increase in beta-adrenergic receptor density and a loss of norepinephrine uptake in postganglionic sympathetic neurons. This augmented sensitivity has an important role in maintaining an adequate cardiac response to exercise and stress.

The output of cardiac allografts is at the low end of the normal range, and the measured cardiac response to exercise or stress is below normal, but the response of the cardiac allograft is adequate for most activities. During exercise the cardiac transplant recipient experiences a steady but delayed increase in heart rate, primarily from a rise in circulating catecholamines. This initial rise in heart rate is accompanied by an immediate increase in filling pressures resulting from augmented venous return. These changes result in an augmentation of stroke volume and cardiac output sufficient to sustain the increase in activity. The ability of the coronary circulation to dilate and increase blood flow in response to increased myocardial oxygen demand is normal. Conversely, graft coronary vasodilator reserve is abnormal in the presence of rejection, hypertrophy, or regional wall motion abnormalities.

The atrial cuff anastomoses also result in abnormal cardiac physiology. The normal atrial contribution to ventricular end-diastolic filling is impaired by the dissociation between recipient and donor atrial contractions. The atrial anastomoses may partially deform the atrioventricular annuli, leading to mitral and tricuspid regurgitation. Studies show a significant reduction in atrioventricular valve regurgitation and lower atrial volumes and pressures in patients with vena caval anastomoses.

Immunosuppression

Conventional immunosuppression in cardiac transplant recipients consists of the “triple-drug” combination of cyclosporine, azathioprine, and glucocorticoids. High doses of these drugs are given initially, with eventual tapering for chronic administration. Cyclosporine potently inhibits T lymphocyte activation, presumably by blocking the release of interleukin 2 from helper T cells. Cyclosporine is titrated to maintain a trough serum concentration of 150 to 250 ng/mL during the first few weeks after transplantation and 50 to 150 ng/mL thereafter. Azathioprine, a cytotoxic agent and bone marrow suppressant, is dosed to maintain the white blood cell count above 5,000/mm³. Glucocorticoids exhibit potent immunosuppressive effects by inhibiting leukocyte elaboration of inflammatory mediators (e.g., lymphokines, colony-stimulating factors). A methylprednisolone bolus (500 mg intravenously) is given intraoperatively after cardiopulmonary bypass is discontinued and during the first 24 h postoperatively (125 mg intravenously every 8 h). Prednisone is then dosed at 1.0 mg/kg/day during the first posttransplant week and, in the absence of acute rejection, tapered to a maintenance dose of 0.2 mg/kg/day.

The addition of prophylactic induction therapy with OKT3 monoclonal antibodies to its standard triple-drug regimen given over the first 10 postoperative days has delayed the time to first rejection and has reduced early rejection rates, but has not resulted in significant differences in the total number of rejection episodes, rates of infection, graft coronary artery disease, renal function, or overall recipient survival.

Judicious doses of these drugs usually are well tolerated by patients, but each drug is associated with side effects. Cyclosporine is associated with nephrotoxicity, hypertension, hepatotoxicity, hirsutism, and an increased incidence of lymphoma. The primary toxicity of azathioprine is generalized bone marrow depression manifested as leukopenia, anemia, and thrombocytopenia. Steroids are associated with a myriad of side effects, including the appearance of cushingoid features, hypertension, diabetes, osteoporosis, and peptic ulcer disease. Initial doses of OKT3 may cause significant hypotension, bronchospasm, or fever, presumably due to T-cell-mediated release of lymphokines. Patients receiving OKT3 are closely monitored and premedicated with acetaminophen, antihistamines, and corticosteroids. Most of these adverse effects are manageable or reversible with dosage reduction; however, the prevalence of adverse effects emphasizes the inadequacies of pharmacologic immunosuppression. Research is being directed toward the development of more potent, less toxic immunosuppressive agents and modes of tolerance induction. Tacrolimus (FK506), rapamycin, and mycophenolate mofetil are several promising drugs

recently approved for use in organ transplantation by the Food and Drug Administration. Like cyclosporine, tacrolimus selectively inhibits T-cell proliferation by blocking cytokine synthesis. Rapamycin inhibits the actions of cytokines and growth factors on T and B cells. By depleting nucleotides, mycophenolate mofetil interrupts DNA synthesis and glycosylation of adhesion molecules in immune cells.

Postoperative Complications (Fig. 10-26)

Acute Rejection

Acute graft rejection is a major cause of death after cardiac transplantation. The incidence of acute graft rejection is highest during the first 3 months after transplantation. At Stanford Hospital, 84 percent of cardiac transplant recipients receiving triple-drug therapy without OKT3 induction had acute rejection during this period. The addition of OKT3 therapy reduces this figure to about 75 percent. After this initial 3 month period, the incidence of acute rejection averages about one episode per patient a year.

Despite attempts at developing noninvasive means to detect acute rejection in a timely manner, the endomyocardial biopsy remains the gold standard for the diagnosis of acute rejection. Surveillance endomyocardial biopsies allow rejection to be diagnosed before significant organ damage and dysfunction occurs. The technique, performed under local anesthesia, involves the percutaneous introduction of a Caves-Schultz bioptome into the right ventricle, usually via the right internal jugular vein, under fluoroscopic guidance. Alternative access sites include the right subclavian and both femoral veins. Multiple biopsy specimens are taken from the interventricular septum per session. Safe, simple, and relatively well tolerated by the patient, endomyocardial biopsies of the cardiac allograft begin 7 to 10 days after transplantation and are repeated at progressively longer intervals.

Acute rejection is characterized histologically by lymphocytic infiltration and myocytic necrosis (Fig. 10-27). Many grading systems have evolved from different transplant groups, culminating in the uniform criteria developed by the International Society for Heart and Lung Transplantation in 1990. The Stanford classification and International Grading Systems are outlined in Table 10-14.

The timing and severity of rejection episodes dictate therapy. Severe or moderate rejection episodes occurring in the early posttransplant period are treated with pulsed steroid dosing. Methylprednisolone is given intravenously at a dose of 1,000 mg/day for 3 consecutive days. Subsequent severe or moderate rejection episodes detected on routine surveillance biopsy are treated with steroid pulsing or by increasing the oral prednisone dosage to 100 to 200 mg/day for 3 consecutive days then tapering it back to baseline dosages over 2 weeks. Mild rejection usually is not treated with augmented immunosuppression unless it is persistent.

Acute rejection refractory to steroid therapy is treated with antilymphocyte preparations in the form of ATG or OKT3 monoclonal antibody. Potent second-line therapies that are used in especially difficult, persistent cases of rejection include methotrexate and total lymphoid irradiation (TLI). Endomyocardial biopsies are repeated 10 to 14 days after antirejection therapy to assess efficacy.

Chronic Rejection

Accelerated graft coronary artery disease (CAD) or atherosclerosis is a major limiting factor for long-term survival in cardiac transplant recipients. Significant graft CAD resulting in diminished coronary blood flow may lead to arrhythmias, myocardial infarction, sudden death, or impaired left ventricular function with congestive graft failure. Typical angina from myocardial ischemia usually is not noted in transplant patients because the cardiac graft essentially is denervated. In a retrospective analysis of cardiac transplants from 1980 through 1993, the actuarial freedom from graft CAD at 1, 5, and 10 years was 95, 73, and 65 percent, respectively. Risk factors for developing this condition have included older donor age, older recipient age, incompatibility at the HLA-A1, A2, and DR loci, hypertriglyceridemia (serum concentration > 280 mg/dL), frequent acute rejection episodes, and documented recipient cytomegalovirus infection.

Multiple causes for graft CAD have been proposed primarily focusing on chronic, immunologically mediated damage to the coronary vascular endothelium. Elevated levels of antiendothelial antibodies have been correlated with graft CAD. Unlike coronary occlusive disease in the native heart, which is focal in nature, transplant atherosclerosis represents a more diffuse vascular narrowing extending symmetrically into distal branches. Histologically, transplant arteriopathy is characterized by concentric intimal proliferation with smooth muscle hyperplasia (Fig. 10-28). Coronary angiograms are performed on a yearly basis to identify recipients with accelerated CAD. Because graft CAD is manifested as diffuse coronary intimal thickening, intracoronary ultrasonography has been advanced as a more sensitive means to detect graft atherosclerosis because of its ability to assess vascular wall morphology in addition to luminal diameter.

Percutaneous transluminal coronary angioplasty and coronary artery bypass grafting have been used to treat discrete proximal lesions in some cases of graft CAD, but the definitive therapy for diffuse disease is retransplantation. Effective prevention of graft CAD relies on developments in improved immunosuppression, recipient tolerance induction, improved CMV prophylaxis, and inhibition of vascular intimal proliferation.

Infection

Infection is the leading cause of morbidity and mortality in post–cardiac transplantation patients. The risk of infection and infection-related death peaks during the first few months after transplantation and rapidly declines to a low persistent rate. In a retrospective study the actuarial freedom from any infection at 3 months, 1 year, and 5 years was 40, 27, and 15 percent, respectively. The actuarial freedom from infection-related death at 3 months, 1 year, and 5 years was 95, 93, and 85 percent, respectively. The most frequent agents of infections after cardiac transplantation are listed in Table 10-15.

Postoperative infections can be broadly classified into those that occur early and those that occur late after transplantation. Early infections, those occurring during the first month after transplantation, are commonly bacterial (especially gram-negative bacilli) and are manifested as pneumonia, mediastinitis, catheter sepsis, and urinary tract and skin infections. Treatment of these infections involves identification of the infective agent (e.g., cultures, antibiotic sensitivity tests), source control (e.g., catheter removal, debridement), and appropriate antibiotic regimens. In the late posttransplant period, opportunistic viral, fungal, and protozoan pathogens are more prevalent. The lungs, CNS, gastrointestinal tract, and skin are the usual sites of invasion.

Cytomegalovirus infection is widely recognized as the most common and important viral infection in transplant patients, with an incidence of 73 to 100 percent in cardiac transplant recipients. It presents as a primary infection or reactivation of a latent infection, most commonly 1 to 4 months after transplantation. By definition, primary infection results when a previously seronegative recipient is infected via contact with tissue or blood from a seropositive individual. The donor organ is thought to be the most common vector of primary CMV infections. Reactivation infection occurs when a recipient who is seropositive before transplantation develops clinical CMV infection during immunosuppressive therapy. Seropositive recipients also are subject to infection by new strains of CMV. CMV infection has protean manifestations, including leukopenia with fever, pneumonia, gastroenteritis, hepatitis, and retinitis. CMV pneumonitis is the most lethal of these, with a 13 percent mortality rate, while retinitis is the most refractory to treatment, requiring indefinite treatment. CMV is significant as an infective agent because of its role as a trigger for accelerated graft CAD and as an inhibitor of cell-mediated immunity.

Diagnosis of CMV infection is made by direct culture of the virus from blood, urine, or tissue specimens, by a fourfold increase in antibody titers from baseline, or by characteristic histologic changes (i.e., markedly enlarged cells and nuclei containing basophilic inclusion bodies). Most cases respond to ganciclovir (DHPG) and hyperimmune globulin. Both of these agents have been used prophylactically, especially in seronegative patients receiving a graft from a seropositive donor. Prophylaxis has been shown to decrease the incidence of clinical CMV infections in recipients who were seropositive before transplantation, but not in seronegative patients.

Fungal infections are less common than bacterial or viral infections. Early recognition is important because these infections are more refractory to therapy and are more lethal. Therapy consists of antifungal agents, including amphotericin B, fluconazole, and flucytosine.

Infection prophylaxis in cardiac transplant patients consists of vaccinations, perioperative broad-spectrum antibiotics, and long-term prophylactic antibiotics. Pretransplant inoculation with pneumococcal and hepatitis B vaccines and diphtheria-pertussis-tetanus (DPT) boosters are recommended. In pediatric patients immunization with live measles-mumps-rubella (MMR) and polio vaccines should be performed before transplantation. All cardiac transplant recipients should receive annual influenza vaccinations. Perioperative antibiotic regimens vary widely, but first-generation cephalosporins (e.g., cefazolin) or vancomycin are used commonly. Long-term prophylaxis typically includes nystatin mouthwash for thrush, sulfamethoxazole-trimethoprim for opportunistic bacterial and *Pneumocystis carinii* infections, and antiviral agents such as acyclovir or ganciclovir.

Neoplasms

Organ transplant recipients are at significantly higher risk for developing cancer, undoubtedly because of chronic immunosuppression. Recipients are predisposed to skin cancer, B-cell lymphoproliferative disorders, carcinoma in situ of the cervix, carcinoma of the vulva and anus, and Kaposi's sarcoma. Conversely, neoplasms of the breast, lung, prostate, and colon are not increased in these patients. On average, tumors appear approximately 5 years after transplantation.

The incidence of B-cell lymphoproliferative disorders in transplant patients is many times greater than in the normal age-matched population. Diagnosis is established by lymph node biopsy. Lymphomas frequently are observed in recipients younger than 20 years of age within a year after transplantation. Recipients older than 45 years of age diagnosed with lymphoma tend to present several years after transplant, with an average survival of 9 months after diagnosis. Thought to be caused by unchecked Epstein-Barr virus infection in the setting of T-cell suppression, B-cell lymphoproliferative disorders are treated with a reduction in immunosuppression and administration of an antiviral agent, such as acyclovir or ganciclovir. A response rate of 30 to 40 percent can be expected, and recurrence is uncommon. Chemotherapy and radiotherapy have been used successfully in some cases. Close monitoring of the graft with echocardiography combined with clinical assessment of tumor status is important during therapy.

Retransplantation

The primary indications for cardiac retransplantation are graft failure from accelerated graft atherosclerosis or recurrent acute rejection. Patients in need of retransplantation are held to the same standard criteria as initial candidates. Survival rates after retransplantation are significantly less than those achieved in primary transplant patients. At the Stanford University Medical Center, the 1-year survival rate was 55 percent after cardiac retransplantation.

Results

According to the Registry of the International Society for Heart Transplantation, 2,500 to 3,500 heart transplantations per year were performed at more than 200 transplant centers worldwide from 1990 to 1995. A review of 496 heart transplants performed at Stanford between 1980 and 1993 placed 1-year, 5-year, and 10-year actuarial survival estimates at 82, 61, and 41 percent, respectively (Fig. 10-29). Most cardiac transplant patients are fully rehabilitated to the New York Heart Association's functional class I status. The benefits of this procedure are apparent when compared to the prognosis of these patients without transplantation.

Pediatric Cardiac Transplantation

Cardiac transplantation is now an accepted therapeutic option for infants and children with end-stage heart disease. The number of children undergoing heart transplantation has increased rapidly from about 200 performed between 1978 and 1986 to well over 200 performed in 1995, with over 80 recipients less than 1 year old. The leading indications for

cardiac transplantation in children are acquired dilated cardiomyopathy and congenital heart disease (Fig. [10-30](#)). Contraindications for transplantation in this group are similar to those in adults, with the addition of some complex venous drainage anomalies.

Blood type and donor size are the most important considerations in donor–recipient matching. The paucity of pediatric heart donors, wider ranges of patient size, and severe disability of many pediatric patients on the waiting list has resulted in an expanded range of accepted donor-to- recipient weight ratios. The average ratio is 1.4 ± 0.45 , ranging from 0.75 to 3.54. Moderately oversized heart grafts are preferred for recipients with an elevated pulmonary vascular resistance.

The operative technique of orthotopic heart transplantation in the pediatric population is similar to that used in adults, with some exceptions in certain cases of congenital heart disease (e.g., aortic arch or venous drainage reconstruction). Immunosuppressive regimens are similar to those used in adults, namely, triple-drug therapy with or without antilymphocytic induction. Steroids are tapered more quickly in pediatric recipients to minimize growth retardation and infectious complications.

Acute rejection rates in children and adolescents do not appear to differ significantly from those encountered in adults, although rejection may be less frequent in neonatal patients because of immune system immaturity. Acute rejection in children is often suspected from a spectrum of signs, including fever, tachycardia, anorexia, and restlessness, coupled with echocardiographic abnormalities (e.g., left ventricular free wall thickening, decreased function). Routine endomyocardial biopsies for rejection surveillance are performed less frequently in neonates and small children, but they are used to confirm the diagnosis of rejection. Acute rejection in children is initially treated with pulsed steroids, and antilymphocyte preparations, total lymphoid irradiation, and methotrexate are reserved for refractory cases. Graft CAD in children occurs at a frequency comparable to that of adult patients and is a frequently encountered cause of death.

Actuarial 1-year, 5-year, and 10-year survival estimates are 75, 60, and 50 percent, respectively, with most survivors achieving the New York Heart Association's functional class I. Normal somatic growth rate can be maintained in these patients, and normal cardiac chamber dimensional growth also occurs.

Lung and Heart-Lung Transplantation

Historical Background

The first experimental attempt at single-lung transplantation was described in dogs by Demikhov in 1947. Although the right lower lobe transplants met with limited success in terms of long-term survival, Demikhov established the technical feasibility of lung transplantation and demonstrated that preservation of the bronchial arteries and nerves was not necessary to maintain graft viability. In 1950 Metras in France reported the first successful whole left lung allotransplantation in canines using a left atrial cuff technique, analogous to that developed by Shumway and Lower, to avoid separate pulmonary venous anastomoses. In the course of studying the pulmonary autonomic reflexes in asthmatics, Juvenelle and colleagues were the first to achieve long-term survival after whole lung reimplantation in dogs. Subsequent physiologic studies demonstrated varying degrees of deterioration in the reimplanted lung's function (e.g., pulmonary artery pressure, lung compliance, gas exchange) and abnormal respiratory patterns, possibly from lung denervation. For some time, it was feared that the decline in lung function after reimplantation would preclude isolated lung transplantation in human beings. Then, in 1963, Haglin and associates demonstrated normal and adequate pulmonary function in reimplanted lungs before and after contralateral pneumonectomy in baboons leading to long-term survival.

The initial experiences with single human lung transplantation were discouraging. Hardy performed the first human lung transplant in 1963. The patient received a single left lung transplant but survived for only 18 days, dying of renal failure. Over the next 18 years, 37 more single-lung transplants were attempted, with the longest survivor dying after 10 months. The majority of these patients died from early graft dysfunction caused by inadequate graft preservation, rejection, or bronchial anastomotic complications (e.g., dehiscence and leakage). Successful efforts in human single-lung transplantation began in 1983, when Cooper performed a right lung transplant in a patient suffering from pulmonary fibrosis.

En bloc double-lung transplantation was successfully applied clinically in the middle to late 1980s. This technique, originally intended for patients with chronic obstructive pulmonary disease, arose from concern about preferential hyperinflation and herniation of the emphysematous native lung and mediastinum that might occur with single-lung transplantation. Since 1989 chronic obstructive lung disease has been effectively treated with single-lung transplantation and currently constitutes a major indication for this procedure. When en bloc double-lung transplantation was extended to patients with septic lung disease, particularly cystic fibrosis, high mortality rates resulted from bleeding complications in the poorly visualized mediastinum. As a result, bilateral sequential lung transplantation was developed, bringing marked reductions in bleeding and airway complications and supplanting the en bloc technique. Since 1983 over 2,700 lung transplants have been reported to the International Lung Transplant Registry.

The first successful experimental heart-lung transplants also are credited to Demikhov, who developed the method of orthotopic en bloc heart-lung transplantation, establishing the technical feasibility of these operations. Cardiopulmonary transplantation using mechanical cardiopulmonary bypass was first attempted in experimental animals by Webb and Howard in 1957, and then by Lower and Shumway in 1961. The survival rates after these early attempts at heart-lung transplantation were quite low, with no animal surviving for more than 5 days postoperatively. The main problem in the canine model was respiratory paralysis and insufficiency from denervation of the heart-lung bloc, casting doubt on the feasibility of these operations. Studies by Haglin and associates showed that denervation of both lungs did not prevent a return of adequate spontaneous respiration in nonhuman primates. Reports by Castaneda and associates revealed that pulmonary ventilation and perfusion and systemic circulatory hemodynamics were preserved after heart-lung autotransplants in baboons. These studies demonstrated the ability of primates to withstand complete cardiopulmonary denervation and assume normal pulmonary function after transplantation.

Reitz described a major modification to the standard technique in the use of a retained portion of the right atrium for a single inflow anastomosis instead of separate caval anastomoses. This technique preserved the donor sinoatrial node, eliminated potential caval anastomotic stenosis, and would facilitate the operation in human beings, where the intrathoracic inferior vena cava is short. The use of cyclosporine also proved to be a key factor. The early results in monkeys suggested that the degrees of rejection in the heart and lung were similar and that the endomyocardial biopsy used successfully in cardiac transplantation might be sufficient for monitoring graft rejection in heart-lung patients. Experience later proved that heart and lung rejection often occurs asynchronously.

In 1981 Reitz performed the first successful human heart-lung transplantation in a 45-year-old woman with end-stage primary pulmonary hypertension. This success rejuvenated worldwide efforts in lung and heart- lung transplantation. Improvements in recipient selection criteria, surgical technique, perioperative management, immunosuppression, infection prophylaxis, and posttransplant surveillance accompanied the expansion of clinical heart-lung transplantation after 1981. The number of heart-lung transplants performed grew rapidly from the five operations performed in 1981 to 241 performed worldwide in 1989. Nearly 1,100 heart-lung transplants have been reported to the Registry of the International Society for Heart and Lung Transplantation (ISHLT).

Preoperative Considerations

Recipient Selection Criteria

The overriding objective in recipient evaluation is to select individuals with progressively disabling cardiopulmonary or pulmonary disease with the greatest potential for full rehabilitation after transplantation. Most recipients suffering from end-stage lung disease with or without concomitant cardiac dysfunction are considered for three broad categories of transplant operations: single-lung, bilateral single-lung, and heart-lung transplantation. The indications for each operation are listed in Table [10- 16](#).

Single-lung transplantation is most ideally suited for patients with fibrotic lung disease because the low compliance and increased vascular resistance of the native lung ensure that ventilation and perfusion are preferentially diverted to the transplanted lung. Emphysematous lung disease also is a major indication for this procedure.

Bilateral single-lung transplantation is intended for patients suffering from septic lung disease, including cystic fibrosis and bronchiectasis, or from chronic obstructive pulmonary disease. Single-lung transplantation is avoided in patients with septic lung disease because the associated chronic bilateral pulmonary infections would place the recipients at high risk for infection from the retained native lung. The extra pulmonary reserve and potential survival advantage afforded by bilateral single-lung as compared to single-lung transplantation must be weighed against the benefits of providing lungs to two patients versus one.

Heart-lung transplantation was developed initially for patients suffering from severe pulmonary vascular disease—specifically, pulmonary hypertension and Eisenmenger's syndrome stemming from congenital cardiac anomalies. The indications for this procedure have broadened to include patients with end-stage lung disease, intercurrent cardiac dysfunction, cystic fibrosis, and bronchiectasis.

Other permutations of cardiopulmonary replacement have been developed. In the so-called domino transplant procedure, the explanted heart of a heart-lung recipient is transplanted into another patient in need of a cardiac transplant. In patients who have had prior thoracic surgery, heart–single-lung transplantation has been successfully used to avoid extensive adhesions in a previously treated pleural cavity. Lobar lung transplantation (from living-related and cadaveric donors) has been developed to expand the effective donor pool, readily providing healthy lung tissue to patients with a severely limited life expectancy, and accommodating the small chest dimensions of pediatric patients.

Contraindications to lung and heart-lung transplantation are similar to those in cardiac transplantation. Among most lung and heart-lung transplant programs, upper age limits typically range from 50 to 60 years, with projected life expectancies limited to less than 12 to 18 months with the use of appropriate medical or alternative surgical strategies. With some pulmonary diseases, survival without transplantation has been estimated using certain parameters, e.g., in patients with primary pulmonary hypertension, elevated right atrial pressure, diminished cardiac output, and elevated pulmonary artery pressure are correlated with diminished survival. Mortality rates of patients on the waiting list are considerable, ranging from 10 to 30 percent.

Patients suffering from systemic disease with significant renal or hepatic dysfunction, acute illness, unresolved malignancy, or psychiatric illness are not offered transplantation. Relative contraindications include cachexia or obesity and a recent history of active peptic ulcer disease. Patients requiring systemic corticosteroids are tapered to the lowest tolerable level before transplantation. Cigarette smokers must quit smoking and be completely abstinent for several months before transplantation. During the early years of heart-lung transplantation, previous cardiothoracic surgery or pleurodesis were considered absolute contraindications to heart-lung transplantation because of bleeding from chest wall adhesions and difficulty in preserving the vagus, recurrent laryngeal, and phrenic nerves; with improved surgical technique over the years, however, these cases are now being considered. A stable, supportive socioeconomic environment is an important consideration.

All prospective cystic fibrosis recipients should have an otolaryngologic evaluation before being placed on the active waiting list. Most of these patients require endoscopic maxillary anastomies for sinus access and periodic antibiotic irrigation to decrease bacterial colonization in the upper respiratory tract. This measure has decreased the incidence of serious posttransplant bacterial infections. Previous smokers must undergo extensive screening to exclude smoking-related illnesses, specifically peripheral vascular disease and malignancy. A negative sputum cytology, thoracic CT scan, bronchoscopy, otolaryngologic evaluation, carotid duplex scan, and flexible sigmoidoscopy/colonoscopy usually are required prior to listing.

Evaluation and Management of Patients Awaiting Lung or Heart- Lung Transplantation

Because adequate cardiac function is the most important determinant of whether single-lung or bilateral single-lung transplantation will be tolerated, the first phase of the preoperative evaluation includes echocardiography with Doppler and saline-contrast flow studies, radionuclide angiography, and, possibly, Holter monitoring and cardiac catheterization.

Patients suitable for heart-lung transplantation enter a second phase of testing and are categorized and listed on the basis of clinical status, time on the waiting list, ABO blood group, and thoracic cage dimensions. The average time from listing to transplantation ranges from 1 to 2 years. Listed candidates are seen in clinic every 3 to 6 months before transplantation to maintain them in optimal medical condition.

Some patients awaiting lung or heart-lung transplantation are in some degree of heart failure. All standard therapeutic measures are applied in these cases, namely, dietary restrictions, diuretics, and vasodilators. Intravascular fluid management for patients in heart failure consists of dietary water and salt restriction and diuretic therapy. Afterload reduction in the form of vasodilators, particularly nitrates, hydralazine, and angiotensin-converting enzyme inhibitors effectively improve functional capacity and prolong survival in patients suffering from severe cardiac failure.

Despite the clinical heterogeneity among patients with primary pulmonary hypertension, conventional medical therapy targets the sequelae of the pulmonary vascular derangements associated with this disease process. Supplemental oxygen therapy is recommended for any patient exhibiting arterial hypoxemia, defined as either an arterial oxygen saturation (SaO_2) less than 90 percent or a PaO_2 less than 60 mmHg at rest, during exertion, or while asleep. Oxygen supplementation is intended to eliminate the stimuli for hypoxic pulmonary vasoconstriction and secondary erythropoiesis, lessening the burden on the right heart and diminishing the contribution of myocardial hypoxia to cardiac arrhythmogenesis. The use of pulmonary vasodilator therapy is based on the notion that pulmonary vasoconstriction is an important component of primary pulmonary hypertension (PPH). Reeves and coworkers have shown that patients in whom a reduction in pulmonary vascular resistance of more than 30 percent is achieved during an acute vasodilator trial are more likely to demonstrate sustained clinical improvement. Severely ill patients with PPH awaiting heart-lung or lung transplantation who were unresponsive to oral vasodilator therapy have received continuous infusions of prostacyclin or prostacyclin analogs, incurring reductions in pulmonary arterial pressure and pulmonary vascular resistance. Despite the favorable response to prostacyclin seen in most of these patients, many eventually become refractory to the drug's effects. Patients placed on hold for lung or heart-lung transplantation after dramatic improvements with prostacyclin therapy should be monitored closely and reactivated early if deterioration is noted.

Interstitial lung disease in patients awaiting transplantation can result from a wide variety of diffuse inflammatory processes, such as sarcoidosis, asbestosis, and collagen-vascular diseases. Increases in pulmonary vascular resistance leading to right heart failure are thought to result from interstitial inflammatory infiltrates that entrap and eventually destroy septal arterioles, reducing the distensibility of the remaining pulmonary vessels. This process, coupled with closure of peripheral bronchioles, results in arterial hypoxemia, further aggravating pulmonary hypertension. Corticosteroids are used in treating this class of diseases, but the adverse effects of steroids on tracheal healing mandate significant dose reductions (prednisone to less than 0.1 mg/kg/day) in anticipation of heart-lung transplantation. Patients with dilated cardiomyopathy, congestive heart failure, and primary pulmonary hypertension are predisposed to pulmonary and systemic thrombosis and embolization. Most centers routinely use prophylactic anticoagulation agents (e.g., heparin or warfarin) or antiplatelet agents in these patients.

The multisystem manifestations of cystic fibrosis, particularly chronic bronchopulmonary infection, malabsorption, and diabetes mellitus, pose difficult management problems in potential lung and heart-lung recipients. Aggressive chest physiotherapy, antibiotics, enteral or parenteral nutritional supplementation, and tight serum glucose control require meticulous management.

Donor Selection and Management

Potential donors must have sustained irreversible brain death, usually as a result of blunt and penetrating head trauma or intracranial hemorrhage. Because of the susceptibility of the lungs to infection and edema, particularly in the settings of brain death and trauma, suitable heart-lung blocs are more difficult to obtain than other organs. Less than 20 percent of non-thoracic organ donors possess lungs suitable for donation.

Donor evaluation consists of a directed history and physical examination, chest radiograph, 12-lead electrocardiogram, arterial blood gas determinations, echocardiogram, serologic screening (i.e., HIV, hepatitis B surface antigen [HB sAg], hepatitis C antibodies, herpes simplex virus, cytomegalovirus, and *Toxoplasma*), and direct inspection and palpation of the heart and lungs at explantation. Echocardiographic evidence of normal cardiac function and the absence of a significant cardiac history and significant coronary atherosclerosis must be established. A donor age of less than 40 years is preferred, but potential donors aged 40 to 50 years are considered with a more detailed evaluation, including coronary angiography to rule out significant coronary artery disease. A donor chest film must be entirely clear and the PaO_2 should exceed 100 mmHg on an FiO_2 of 30 percent and 400 mmHg on an FiO_2 of 100 percent. Lung compliance can be estimated by measuring peak inspiratory pressures, which should be less than 30 cmH₂O. Bronchoscopy should assure the absence of purulent secretions or signs of aspiration. Donors should receive broad-spectrum antibiotics for infection prophylaxis before explantation.

Absolute contraindications for donation include severe coronary or structural heart disease, prolonged cardiac arrest, prior myocardial infarction, a carbon monoxide-hemoglobin level greater than 20 percent, arterial oxygen saturation less than 100 percent, active malignancy (sometimes excluding primary brain and skin cancers), a significant smoking history (more than 5 pack-years or 1 pack per day over the past year), and HIV seropositivity. Relative contraindications include thoracic trauma, sepsis, prolonged severe hypotension (i.e., less than 60 mmHg for more than 6 h), noncritical coronary artery stenosis, HB sAg or hepatitis C antibodies, multiple resuscitations, severe left ventricular hypertrophy, and a prolonged high inotropic requirement (e.g., dopamine in excess of 15 mg/kg/min for 24 h). Correctable metabolic or physiologic causes of cardiac rhythm disturbances and electrocardiographic anomalies (e.g., brain herniation, hypothermia, hypokalemia) should be ruled out.

The primary objective in managing the lung or heart-lung donor is the maintenance of hemodynamic stability and pulmonary function. Patients suffering from acute brain injury often are hemodynamically unstable from neurogenic shock, excessive fluid losses, and bradycardia. Donor lungs are subject to neurogenic pulmonary edema, aspiration, nosocomial infection, and contusion. Continuous arterial and central venous pressure monitoring, judicious fluid resuscitation, vasopressors, and inotropes usually are required.

Meticulous fluid management prevents intraoperative blood pressure instability and minimizes the need for inotropes and vasopressors, which are myocardial stressors. Intravascular fluids should be given to maintain the central venous pressure at 5 to 8 mmHg, being careful not to administer fluids at rates far in excess of the hourly urine output. Crystalloid fluid boluses should be avoided. Diabetes insipidus is common in organ donors, requiring the use of intravenous vasopressin (0.8 to 1.0 U/h) to reduce excessive urine losses.

To maintain adequate perfusion pressures, dopamine is the standard inotropic agent used, although alpha agonists (e.g., phenylephrine) often are appropriate. Blood transfusions should be used sparingly to maintain the hemoglobin concentration about 10 g/dL to ensure adequate myocardial oxygen delivery. The use of CMV-negative and leukocyte-filtered blood should be used whenever possible. Hypothermia should be avoided because it predisposes to ventricular arrhythmias and metabolic acidosis.

In mechanical ventilation, FiO_2 levels in excess of 40 percent, especially 100% oxygen "challenges," should be avoided, because these oxygen levels may be toxic to the denervated lung. Ventilator settings should include positive end-expiratory pressures (PEEP) between 3 and 5 cmH₂O to prevent atelectasis.

Donor-Recipient Matching

Donor-recipient matching parameters include ABO compatibility and body size. Donor-to-recipient lung volume matching is based on the vertical (apex to diaphragm along the midclavicular line) and transverse (level of the diaphragmatic dome) dimensions on chest roentgenogram as well as body weight, height, and chest circumference. Matching donor and recipient height seems to be the most reproducible method of selecting the appropriate donor lung size. The dimensions of the donor lungs should not be greater than 4 cm over similar measurements in the potential recipient and, preferably, should be smaller than those of the recipient. In a series of 82 heart-lung transplants, Tamm and colleagues recorded recipient lung volumes after transplantation and compared them to preoperative and predicted volumes to evaluate the influence of donor lung size and recipient underlying lung disease. The investigators showed that 1 year after transplantation total lung capacity (TLC) and dynamic lung volume returned to values predicted by the patients' sex, age, and height. They proposed that the simplest method of matching donor lung size to the recipient is to use their respective predicted TLC values.

When an appropriate donor-recipient pairing has been made, the recipient is screened for preformed antibodies against a panel of random donors. A panel-reactive antibodies (PRA) level greater than 50 percent prompts a prospective specific crossmatch between the donor and recipient. The relationship of heart-lung allograft-related death and chronic rejection to HLA matching was studied in 40 consecutive heart-lung transplant recipients. This study revealed a significant increase in graded obliterative bronchiolitis (OB) with total mismatch at the HLA-A locus. There was a slightly increased frequency, severity, and mortality with respect to OB in these mismatched patients. The incidence of OB is correlated with the number or severity and persistence of acute rejection episodes after lung and heart-lung transplantation, but prospective HLA matching currently is not feasible in cardiopulmonary transplantation.

Operative Techniques

Donor Lung and Heart-Lung Bloc Procurement

In single-lung and bilateral single-lung procurement operations, the chest is entered via a median sternotomy. Donor lungs usually are procured in conjunction with the heart (Fig. 10-31). First, the aorta and venae cavae are encircled with umbilical tapes in preparation for inflow occlusion and cardioplegic arrest. For each lung to be procured, the pleura is opened longitudinally, and the entire pericardium is excised from the diaphragm to the pleural apex, extending posteriorly through the phrenic nerve to the hilum. The inferior pulmonary ligament is incised to the inferior pulmonary vein. The right or left pulmonary artery is dissected free from the pulmonary artery bifurcation to the pulmonary hilum. Pulmonoplegia solution, most commonly Euro-Collins solution at 2 to 4°C, is rapidly flushed into the main pulmonary artery simultaneously with cardioplegic infusion into the ascending aorta. The tip of the left atrium is cut to vent the pulmonoplegia solution. Iced saline solution is then poured over the heart and lungs. The superior and inferior venae cavae, the ascending aorta, and common the pulmonary artery are divided, leaving the heart attached only to the pulmonary veins. In single-lung procurement, the contralateral pulmonary veins are divided and a left atrial cuff surrounding the two remaining pulmonary veins is created. In bilateral single-lung procurement, all four pulmonary veins are left intact with the atrial cuff. After removal of the heart, the trachea is clamped or stapled at its midpoint and divided.

In combined heart-lung transplantation, the heart and lungs are excised en bloc through a median sternotomy that is extended into the abdomen. Both pleural spaces are entered anteriorly, and the pericardium is trimmed, leaving both phrenic nerves on pedicles. After cannulating the heart, cardiopulmonary bypass is instituted and ventilation is stopped to facilitate the remainder of the operation. After inflow occlusion and cardioplegic/pulmonoplegic arrest, the ascending aorta, pulmonary artery, and venae cavae are dissected free. The innominate artery and vein are ligated and divided, facilitating exposure of the underlying trachea. The trachea is divided about 6 to 8 cm above the carina. The heart-lung bloc is dissected free from the esophagus, both pulmonary ligaments are divided inferiorly, and the posterior hilar attachments are divided. The respective organ or organs are removed from the chest after stapling of the trachea and immersed in a bag containing ice-cold saline solution at 2 to 4°C which in turn is transported in an ice-filled cooler.

Heart-Lung Transplantation

Heart-lung transplantation is performed with the recipient on cardiopulmonary bypass. After the chest is entered through a median sternotomy, both pleural spaces are opened anteriorly. The anterior portion of the pericardium is excised, preserving the lateral segments to support the heart and to protect the phrenic nerves (Fig. 10-32). The ascending aorta and both venae cavae are cannulated, and hypothermic cardiopulmonary bypass is instituted, cooling the patient to about 30°C.

The native heart and lungs are excised from the thorax. The native heart is arrested and excised at the ascending aorta just above the aortic valve, through the main pulmonary artery, and both atria. After the left and right phrenic nerves are dissected free, the pulmonary ligaments are divided inferiorly and the pulmonary artery and vein are divided in each hilum. The left and right bronchi are stapled and divided, and the lungs are removed. A portion of the pulmonary artery is left intact adjacent to the underside of the aorta near the ligamentum arteriosus to preserve the recurrent laryngeal nerve.

The donor heart-lung graft is then prepared by irrigating, aspirating, and culturing the tracheobronchial tree and by trimming the trachea to leave one cartilaginous ring above the carina. The heart-lung graft is lowered into the chest, positioning the right lung below the right atrial cuff and right phrenic nerve. After opening the recipient trachea just above the carina, the tracheal anastomosis is performed with a running polypropylene suture. Next, a donor right atrial cuff is constructed and anastomosed to the retained recipient right atrial cuff. At this point the patient is rewarmed toward 37°C and the aortic anastomosis is performed, completing the implantation. After the ascending aorta and pulmonary artery are cleared of air, the aortic cross clamp and caval tapes are removed. Isoproterenol usually is administered on graft reperfusion to increase the heart rate and to lower pulmonary vascular resistance. Ventilation is resumed, starting with an FiO_2 of 50 percent. Cardiopulmonary bypass is discontinued and decannulation performed after normothermia and satisfactory cardiopulmonary function have been achieved. Temporary right atrial and ventricular pacing wires are placed. Right and left pleural chest tubes are placed, and the chest is closed.

Single-Lung and Bilateral Single-Lung Transplantation

The single-lung transplant procedure (Fig. 10-33) is performed under single-lung anesthesia with the aid of a bronchial blocking device or double-lumen endotracheal tube. The side receiving the lung graft is not ventilated during implantation. Cardiopulmonary bypass usually is not required, but it is often indicated in cases of significant pulmonary hypertension. A posterolateral thoracotomy is used for single-lung grafting. Sequential bilateral single-lung transplants are performed through a bilateral anterotheracosternotomy or “clamshell” incision.

Excision of the recipient's diseased lung begins with the temporary occlusion of the ipsilateral pulmonary artery. If the patient's blood pressure, contralateral pulmonary artery pressure, and arterial blood gases are satisfactorily stable during this maneuver, the case can proceed. If significant perturbation of one or more of these parameters occurs, cardiopulmonary bypass will be required. The pulmonary veins are isolated lateral to the pericardium and the mainstem bronchus is mobilized just proximal to the upper lobe bronchus. After complete hilar mobilization, cardiopulmonary bypass is instituted if needed; otherwise, the pulmonary artery is clamped as proximally as possible and divided just distal to the first upper lobe branch. The two pulmonary veins are ligated extrapericardially. The mainstem bronchus is then divided just proximal to the upper lobe bronchus, and the recipient's diseased lung is removed from the chest.

Lung graft implantation begins with clamping the recipient's left atrium to isolate the ligated pulmonary vein stumps. The recipient left atrial cuff is then fashioned and, after the graft is placed in the chest, anastomosed to the donor cuff. The bronchial anastomosis is completed end-to-end with a running polypropylene suture. Omental wrapping of the bronchial anastomosis usually is not performed. The pulmonary artery anastomosis is constructed end-to-end, leaving the running suture untied to permit flushing and backbleeding. The left atrial clamp is released gradually, permitting backbleeding through the untied pulmonary artery suture line. After the pulmonary artery clamp is momentarily flashed to flush the pulmonary artery, the suture line is tied and the clamp is removed, restoring circulation to the lung graft, which is then ventilated. A second sequential lung transplant is performed in the same manner. After all anastomoses are completed and hemostasis is assured, the patient is weaned from bypass if it was required. Thoracostomy tubes are placed and the chest is closed. All bronchial anastomoses are checked endoscopically before the patient leaves the operating suite.

The “Domino” Operation

Because of the severe shortage of thoracic organs for transplantation, the so-called domino operation has been developed. This approach involves the use of the explanted heart from a patient undergoing heart-lung transplantation for primary lung disease for a second recipient in need of a heart transplant. More than 50 percent of heart-lung transplant recipients possess hearts with normal or near-normal left ventricular function and some degree of right ventricular hypertrophy resulting from elevated pulmonary artery pressures. The use of hearts explanted from patients with pulmonary hypertension has theoretic appeal in that the right ventricle already is adapted to elevated pulmonary vascular resistances, decreasing the likelihood of acute donor right heart failure in recipients with preexisting pulmonary hypertension. Starnes and Kells reported preservation or improvement of right ventricular function among seven domino heart transplant recipients. Oaks and associates described their experience with 32 domino heart transplants. No difference was observed in the 3-month and 1-year survival rates between domino and nondomino cardiac transplant recipients. Yacoub and colleagues reported an actuarial 1-year survival rate of 75 percent among recipients of domino hearts from cystic fibrosis patients who underwent heart-lung transplantation.

The domino donor cardiectomy differs slightly from the standard technique described by Lower and Shumway. In preparing the heart-lung recipient for cardiopulmonary bypass, the venous cannulas are placed into the inferior vena cava extrapericardially close to the diaphragm, and into the high superior vena cava at least 4 or 5 cm above the sinoatrial node. This modification enables the excision of the domino heart high on the superior vena cava to preserve the sinoatrial node for the domino recipient.

Living-Related Lobar Transplantation

In 1994 Starnes and associates described bilateral lobar transplantation with donor lobes from related donors. The technique is most commonly performed in recipients with end-stage cystic fibrosis. Early survival rates are equivalent to those in cadaveric bilateral single-lung transplants. This procedure provides an alternative for patients who may otherwise not survive before a suitable organ donor is identified.

Management of Recipients

Early Postoperative Period

Ischemia and reperfusion injury in the transplanted lung result in increased vascular permeability and impaired mucociliary clearance mechanisms, prompting special considerations in postoperative management. After the operation the patient remains intubated and is transported immediately to the intensive care unit. Monitoring of the recipient includes a pulmonary artery catheter, a peripheral arterial catheter, a pulse oximeter, and an end-tidal CO_2 monitor.

Proper ventilator management is of paramount importance during this period, as barotrauma and high airway pressures that might compromise bronchial mucosal flow must be avoided. Lower tidal volumes and flow rates may be necessary to limit peak airway pressures to less than 40 cmH₂O. Pulmonary toilet with endotracheal suctioning is an effective means of reducing mucus plugging and atelectasis. After the patient is fully awake, he or she is weaned from the ventilator and is extubated when blood gases and ventilatory mechanics are satisfactory. Most patients are extubated within 3 days and weaned from supplemental oxygen by 10 days after transplantation. Early ambulation is encouraged.

A diffuse interstitial infiltrate often is found on early postoperative chest x-rays. Previously referred to as a “reimplantation response,” this finding is better defined as graft edema due to inadequate preservation, reperfusion injury, or early rejection. It appears that the degree of edema is inversely related to the quality of preservation. Judicious administration of fluids and loop diuretics is required to maintain fluid balance and minimize this edema. Early lung graft dysfunction manifested by persistent marginal gas exchange without evidence of infection or rejection occurs in less than 10 percent of transplants. Histologic analysis revealing diffuse alveolar damage suggests that this phenomenon is a result of ischemia or reperfusion injury. Technical causes of graft failure should be considered.

Immunosuppression

Immunosuppression protocols for lung and heart-lung transplant recipients are similar to those used in cardiac transplantation. Triple-drug therapy begins immediately after operation and is tapered according to standard protocols. Induction therapy with antilymphocytic antibody preparations can be used.

Postoperative Complications

Early morbidity and mortality after lung and heart-lung transplantation are most commonly caused by infection, graft failure, and heart failure. Mortality after 1 year is caused most commonly by obliterative bronchiolitis, infection, and malignancy.

The majority of acute rejection episodes in lung transplant recipients occur during the first 3 months after transplant. Biopsy-proved rejection occurs in 60 to 70 percent of patients in the first month. In the early posttransplant period, the diagnosis of acute rejection usually is based on clinical parameters. Signs of rejection include fever, dyspnea, impaired gas exchange manifested by a decrease in PaO_2 , a diminished forced expiratory volume in 1 second (FEV_1 , a measure of airway flow), and the development of an interstitial infiltrate on chest x-ray. Fiberoptic bronchoscopy with transbronchial parenchymal lung biopsy and bronchoalveolar lavage is used routinely to diagnose acute rejection or rule out infection. Acute lung rejection is characterized histologically by lymphocytic perivascular infiltrates.

In the case of heart-lung transplantation, the results of simultaneous endomyocardial and transbronchial surveillance biopsies have been compared. Pulmonary and cardiac rejection presented asynchronously in most cases. Transbronchial biopsy had a sensitivity of 89 percent in predicting cardiac rejection, and endomyocardial biopsy had a sensitivity of 34 percent in predicting lung rejection.

Episodes of acute rejection are treated with a short course of intravenous steroid boluses. After steroid therapy improvement often is rapid and dramatic and is considered confirmatory of rejection. Persistent rejection is treated with ATG or OKT3 monoclonal antibodies.

Chronic lung allograft rejection is the greatest limitation to the long-term benefits of lung and heart-lung transplantation. Chronic lung rejection most commonly presents as obliterative bronchiolitis (OB), a pulmonary corollary to cardiac graft atherosclerosis. Diagnosed in 20 to 50 percent of long-term lung transplant survivors, OB is histologically characterized by dense eosinophilic submucosal scar tissue that partially or totally obliterates the lumina of small airways (Fig. 10-34). Physiologically, OB is manifested as decreases in PaO_2 and FEV_1 . Experimental and clinical evidence points to injury of the bronchial epithelium by several mechanisms as the cause of OB, including infection, particularly by CMV, toxic fume inhalation, chronic foreign body exposure stemming from impaired mucociliary clearance, and immunologic mechanisms. A working formulation for the clinical staging of chronic lung graft dysfunction based on the ratio of the current FEV_1 to the best posttransplant FEV_1 has been proposed by Cooper and associates (Table 10-17). Often there is no correlation between the histologic and the physiologic manifestations of OB.

There is no effective treatment for OB. Augmentation of immunosuppression constitutes current therapy. Pulmonary function can be stabilized in most patients, but significant improvement occurs infrequently. Relapse rates exceed 50 percent, and progressive pulmonary failure or infection resulting from increased immunosuppression are the most common causes of death after the second year. Efforts at preventing the development of OB focus on improved immunosuppression, aggressive treatment of acute rejection episodes, and infection prophylaxis.

Infection

Bacterial, viral, and fungal infections are the leading causes of morbidity and mortality in lung and heart-lung transplant recipients. Bacterial infections, particularly those caused by gram-negative bacteria, predominate during the early postoperative period, but the risk of bacterial pneumonia persists throughout the recipient's life. Most common are pulmonary bacterial infections involving the allograft. The absence of the cough reflex in the denervated lung, abnormal mucociliary clearance mechanisms, and deficiencies in lymphatic drainage predispose grafted lungs to infection. Between 75 and 97 percent of bronchial washings obtained from donor lungs before organ retrieval will culture at least one organism. Posttransplant invasive infections frequently are caused by organisms cultured from the donor. Conversely, bacterial infections developing in patients with septic lung disease, particularly cystic fibrosis, most commonly originate from the recipient's airways and sinuses. Therapy consists of identifying the offending organism and instituting antibiotic therapy on the basis of sensitivity studies.

CMV is the most common and most clinically significant viral pathogen. Infections occur most frequently between 2 weeks and 100 days after transplantation. Primary and reactivation CMV infections in lung transplant recipients encompass a wide range of severities and variegated clinical presentations. Primary infection in previously seronegative recipients is more serious than reactivation or reinfection in seropositive patients. The diagnosis of CMV pneumonitis, usually the most severe manifestation of CMV infection, is made from a positive viral culture or cytologic evidence obtained from bronchoalveolar lavage or transbronchial biopsy, respectively. Ganciclovir is the treatment of choice. Herpes simplex pneumonia, which presents similarly to CMV pneumonitis, is treated with acyclovir.

In one series, CMV infections developed in approximately 90 percent of seronegative recipients who received lungs from seropositive donors, compared to about 10 percent who received lungs from seronegative donors. CMV infections occur in the majority of patients who are seropositive before lung transplantation. Because of organ scarcity, most transplant centers perform transplants across CMV serologic barriers. CMV prophylaxis includes ganciclovir, acyclovir, and polyvalent immune globulin. Lung and heart-lung transplant patients also are at a higher risk for developing lymphoproliferative disease, particularly in association with Epstein-Barr virus infection. Treatment consists of lowering immunosuppression and administering acyclovir. Certain lymphomas have been treated successfully with chemotherapy and radiotherapy.

Fungal infections, the most infrequent but most deadly of infectious complications in transplant patients, peak in frequency between 10 days and 2 months posttransplant. Fungal species encountered in these patients include *Candida albicans* and *Aspergillus*. Treatment consists of fluconazole, itraconazole, or amphotericin B. Prophylaxis with inhaled amphotericin B has greatly reduced early fungal infections. *Pneumocystis carinii* pneumonia has been effectively prevented in lung transplant patients since the institution of prophylaxis in the form of oral trimethoprim-sulfa or, for sulfa-allergic patients, inhalational pentamidine. The highest risk of infection with *Pneumocystis* occurs during the first transplant year; infections do occur late after transplant, however, prompting the continued use of prophylaxis indefinitely.

Improvements in surgical technique and posttransplant management have resulted in a relatively low incidence of airway complications after lung and heart-lung transplantation. The rates of lethal airway complications and late stricture have been reported at 3 percent and 10 percent, respectively. The most common airway complications are partial anastomotic dehiscence and stricture. Such complications usually are diagnosed during bronchoscopic examination. Airway dehiscence is treated by reoperation or close observation and supportive care. Strictures are treated with laser ablation or dilation with rigid bronchoscopy or balloon and bougie dilators. Most strictures are stented after dilation. Cystic fibrosis patients are at a higher risk for developing airway complications after transplantation.

Retransplantation in patients who have developed end-stage obliterative bronchiolitis has yielded poor results. In a collected series of pulmonary retransplantation performed at centers in North America and Europe, actuarial 1-year and 2-year survival rates are 41 percent and 33 percent, respectively. The most common causes of death after retransplantation were infection and OB.

Results

According to the International Heart-Lung Registry, the 6-year actuarial survival rate for single-lung and bilateral single-lung transplants performed worldwide from 1982 to 1995 is about 40 percent (Fig. 10-35). At Stanford, the 1-year, 5-year, and 10-year actuarial survival rates after adult heart- lung transplantation performed from 1981 to 1994 are 68, 43, and 23 percent, respectively (Fig. 10-36). Most recipients are able to resume an active lifestyle without supplemental oxygen. Pulmonary function measured by spirometry and arterial blood gases is improved significantly in patients after transplantation, with a normalization of ventilation and gas exchange after 1 to 2 years. The improvements in these parameters are greater in bilateral versus single-lung transplant recipients, but significant differences in exercise testing parameters have not been observed.

Pediatric Lung and Heart-Lung Transplantation

More than 350 children have undergone single-lung, double single-lung, or heart-lung transplantation worldwide since 1982. Heart-lung transplantation is indicated for children suffering from end-stage pulmonary vascular or parenchymal diseases, particularly pulmonary hypertension (primary and associated with congenital heart defects) and cystic fibrosis. Single or bilateral single-lung transplantation has become an established option with preserved right ventricular function.

The selection criteria for pediatric recipients are similar to those used in adult lung and heart-lung transplantation. Potential recipients generally have a life expectancy of less than 1 year, have normal hepatic and renal function, are free of active systemic infection, and are in a stable psychosocial environment. Blood group and lung size are the primary criteria used for donor–recipient matching. Size matching is particularly difficult in the pediatric population because of wider size disparities. The recent development of lobar transplantation has permitted the grafting of lung lobes from living-related or cadaveric adult donors into children. Early experience with lobar transplantation has been encouraging. The operative techniques for lung and heart-lung transplantation in children are essentially the same as those used in adult thoracic transplantation. Triple- drug therapy is the mainstay of postoperative immunosuppression. Some also use OKT3 induction therapy.

The incidence of acute rejection in pediatric lung and heart-lung patients does not differ significantly from that seen in adults. Surveillance for rejection in children and adolescents consists of serial transbronchial biopsies with lavage and pulmonary function studies. Routine invasive bronchoscopy is not performed in neonates for technical reasons, shifting emphasis to clinical and radiographic signs in diagnosing pulmonary rejection. The diagnosis of acute pulmonary rejection in the pediatric population is based on a combination of clinical signs, laboratory tests, and histologic analysis. Pyrexia, fatigue, dyspnea or oxygen desaturation, an interstitial or perihilar infiltrate on chest radiography, and a decreasing FEV₁ are suggestive of rejection.

Long-term survival in pediatric lung and heart-lung recipients is limited by infection, obliterative bronchiolitis, and accelerated graft coronary artery disease. For pediatric lung and heart-lung transplants performed at Stanford from 1986 to 1993, actuarial 3-month, 12-month, and 24-month survival rates of 78, 55, and 41 percent, respectively, have been observed. Somatic growth in the pediatric transplant population generally is improved after transplantation despite the harmful effects of steroids and cyclosporine on bone metabolism. This is most likely because of the marked improvement in cardiopulmonary status after transplantation. Corresponding growth of the transplanted organs also occurs.

KIDNEY

Anthony M. D'Alessandro and Mark H. Deierhoi

Preoperative Management

Renal transplants are the most common solid organ allografts performed, and transplantation has become the preferred treatment of chronic renal failure for many patients. Graft and patient survival have steadily improved, and relative and absolute contraindications to transplantation have consequently receded.

Transplant Recipient Evaluation

Not all patients with end-stage renal disease are transplant candidates. A thorough evaluation is essential for all potential candidates in order to assess their general health, to identify any comorbid conditions that might increase operative risk and the risk in long-term immunosuppression, and to determine the benefits that might be derived from transplantation. Evaluation also provides an opportunity to educate the potential candidate about the risks of immunosuppression and the importance of compliance with medication regimens.

Patients should be referred for transplant evaluation as they are approaching end stage but before dialysis therapy is initiated. The timing is not critical, because long-term dialysis usually is well tolerated and transplantation is not an immediately life-saving procedure. If the patient presents at end stage with the need for immediate dialysis, it is better to delay evaluation to avoid an overwhelming flood of information and subsequent anxiety.

The pretransplant evaluation can be performed in an outpatient setting with occasional admission for specific diagnostic procedures such as cardiac catheterization. Evaluation should include a careful and complete history and physical examination, with attention directed to the history of renal disease, prior surgery, and comorbid conditions, such as heart disease, peripheral vascular disease, and diabetes. Any history of cancer or recent infection should be documented. Laboratory studies should include standard chemistries, complete blood counts, urinalysis, serologic studies for hepatitis B and C, cytomegalovirus, and HIV. A chest x-ray and electrocardiogram also are included for adult candidates.

Evidence of risk factors for surgery should prompt more thorough investigations. Specific tests for associated conditions may include noninvasive cardiac studies such as an echocardiogram or a stress test, evaluation for peripheral vascular disease with noninvasive vascular studies, and pulmonary function tests for patients with a significant history of chronic pulmonary disease. Cardiac catheterization may be required for assessment of coronary disease. Urine cultures should be obtained and a urologic evaluation performed if there is evidence of urologic anatomic abnormalities or prior urologic surgeries, such as ureteral reimplantation or creation of an ileal conduit in a patient with a neurogenic bladder.

Evaluation for long-term risks posttransplant focuses on conditions that could be exacerbated by immunosuppression, such as malignancy and infections. A metastatic work-up is appropriate for a patient with a remote history of malignancy, and any sign of recent malignancy should prompt aggressive evaluation with studies such as endoscopy or even biopsy. The chief infectious risks are tuberculosis, chronic fungal infections, and viral infections such as hepatitis and HIV. Other conditions requiring in-depth evaluation include peptic ulcer disease, cholelithiasis, and diverticulosis.

The majority of patients require immunosuppressive medication for the life of their allograft. Some attempt should be made to assess the patient's risk for noncompliance. This should include evaluation by social workers and psychiatrists and a record of any history of substance abuse. If the potential for substance abuse is unclear, a drug screen is indicated. Compliance on dialysis also should be documented.

Patients should understand the long-term commitment to follow-up and immunosuppressive therapy. Patients should be given comprehensive information regarding the risks and benefits of transplantation and should be educated regarding immunosuppressive medications and their side effects.

In certain situations, surgery may be required in preparation for transplantation. Cholecystectomy should be performed on any patient with symptomatic gall bladder disease and in any diabetic with documented gallstones. Some patients are evaluated for diverticulosis and have elective colon section. Bilateral nephrectomy may be required in specific situations, including recurrent urinary tract infections with reflux, uncontrollable hypertension, and polycystic kidney disease with recurrent bleeding and infections. There is an occasional patient with polycystic disease whose kidneys are so massive that they fill the abdomen and must be removed to allow room for the transplant. Occasionally, patients with urinary tract abnormalities require bladder augmentation or creation of an ileal conduit.

Indications and Contraindications

There are a number of absolute and relative contraindications to transplantation (Table [10-18](#)). Absolute contraindications include recently treated cancer other than squamous and basal cell carcinoma of the skin, HIV infection, hepatitis with evidence of cirrhosis or of chronic hepatitis on biopsy, and severe ischemic cardiac disease not amenable to bypass surgery or angioplasty. Patients with severe peripheral vascular disease requiring bypass surgery usually are not candidates. Relative contraindications include obesity, history of noncompliance with other medical therapy, a history of tuberculosis, and several renal diseases, including oxalosis and sickle cell disease, which have a high incidence of recurrence.

Management on Dialysis

Pretransplant management also may include the provision of access for dialysis therapy. Dialysis techniques include hemodialysis through some sort of vascular conduit and peritoneal dialysis. The choice of technique depends to a large extent on patient preference and the patient's ability to follow medical regimens. Peritoneal dialysis and home hemodialysis require significant patient involvement in care but are more compatible with an active lifestyle. Both forms are compatible with transplantation. The choice of dialysis access usually is not based on the patient's eligibility for transplant.

Hemodialysis access may be gained through various intravenous catheters, the creation of native arteriovenous fistulas, or the placement of a prosthetic conduit, such as polytetrafluoroethylene (PTFE) grafts, to create an arteriovenous fistula. A long-term plan should be worked out with the patient before the institution of hemodialysis therapy. This is important if grafts are required, because the grafts have a limited life span. When a patient is identified with renal failure, it is imperative that he or she be instructed regarding the preservation of extremity vasculature, which may be needed for future access surgery. The most critical vessels are the cephalic vein in the forearm and upper arm and the veins of the antecubital space.

A number of catheters are available for temporary or long-term access. These are placed in a large central venous site, preferably the internal jugular vein, followed by subclavian, femoral, and inferior caval placement. Temporary catheters may be required if the patient presents with end-stage disease and the urgent need for dialysis. They also may be required if a long-term conduit, such as a fistula or graft, fails. In this situation, catheters can be used for dialysis while a new long-term access site matures. Repeated cannulation or long-term placement of a catheter in the subclavian vein frequently leads to stenosis, which can preclude the use of that extremity for long-term dialysis. All attempts should be made to place temporary catheters in the internal jugular veins. In the potential transplant recipient who is nearing end stage and has an identifiable living donor, a catheter might be suitable for short-term use until transplantation. Long-term use of catheters may be necessary in patients with severe peripheral vascular disease and calcification. In some instances, success with dialysis via catheters has been maintained for periods exceeding 12 months.

The usual choices for long-term dialysis include an arteriovenous fistula and PTFE graft placement. If a fistula site and a suitable vein are available, fistula creation is the overwhelming choice for longevity and suitability for dialysis. The classic Brescia-Cimino fistula is created by anastomosing the cephalic vein and the radial artery at the wrist side-to-side or end-to-side. If a large cephalic vein is present in the upper arm, a brachiocephalic fistula created in the antecubital space also is appropriate. Some surgeons use transposed basilic vein as a fistula conduit. This fistula is created by dissecting the basilic vein in the upper arm and positioning it in a subcutaneous tunnel with anastomosis end-to-side to the brachial artery just proximal to the antecubital space.

PTFE grafts are now the standard conduit used in patients who do not have suitable anatomy for a fistula. These may be placed in upper and lower extremities. The initial site usually is the forearm of the nondominant arm; a graft can be fashioned from the radial artery at the wrist to an antecubital vein, or in a loop fashion from the brachial artery to the antecubital space. Subsequent grafts are placed in the upper arm in a straight or looped configuration with subsequent progression to the use of the femoral vessels if all upper extremity sites have been used.

Short-term complications include infection and distal limb ischemia from preferential flow in the conduit. This occurs most often with fistulas or large-diameter grafts placed in the upper arm. Infection generally requires graft removal, and distal limb ischemia is treated by revising the graft to provide a smaller inflow diameter. Venous hypertension may occur as a result of placement of a graft distal to subclavian venous stenosis. Relief often can be obtained with angioplasty and stenting of the vein. Severe untreatable venous hypertension necessitates ligation of the graft.

Long-term patency is significantly greater for fistulas than for grafts. The mean survival of fistulas is 10 years, whereas for PTFE grafts it is 18 to 24 months. Fistulas fail secondary to proximal stenosis or pseudoaneurysm formation from repeated cannulation in the same sites. The most common cause of failure of the PTFE graft is intimal hyperplasia of the venous anastomosis. Development of pseudoaneurysms from repeated cannulation, venous stenosis proximal to the anastomosis, and infection all can lead to graft thrombosis and loss. Intimal hyperplasia is best treated by following patients with pressure measurements on dialysis, identifying stenoses with Doppler ultrasonography, and revising the anastomoses with patch angioplasties or jump grafts. Interventional radiology has had some success in dilating stenoses with occasional stent placement. Interventional techniques have been developed for clot lysis and angiography, even when a graft has thrombosed.

Peritoneal dialysis is another important option for renal failure therapy. A catheter is placed intraperitoneally and a dialysis fluid of variable composition is instilled. This technique uses the filtering capabilities of the peritoneal lining and osmotic gradients to draw the by-products of uremia into the peritoneal fluid. Patients must perform six to eight exchanges per day to maintain homeostasis. Peritoneal dialysis has the advantage of freeing the patient from scheduled time on a dialysis machine. In addition, it is continuous and easily portable therapy. Catheters are placed in an outpatient setting and require 2 to 3 weeks to mature before the institution of therapy. Complications of peritoneal dialysis include tract infections and peritonitis. This can be a devastating problem, particularly if the infection is fungal. These infections tend to cause exudative peritonitis with formation of multiple adhesions. Patients undergoing peritoneal dialysis require significant involvement in their own care and an understanding of sterile technique.

Histocompatibility Testing

The work-up of a potential transplant recipient begins with blood group typing and HLA typing. Blood group typing is essential because renal endothelial cells express major blood group antigens and the preformed natural antibodies to these antigens can result in hyperacute rejection. In cadaveric transplantation, absolute blood type matching is required, but type O donors are universal donors. For living donor transplants, there is an exception in that the ABO subtype A2 appears less antigenic. With plasmapheresis and preoperative preparation of the patients, an A2 donor kidney can be transplanted into a non-A2 recipient.

HLA typing is performed routinely on all cadaveric and living donor recipients. The standard typing procedure is a lymphocytotoxic serologic test in which the potential recipient's cells are tested against a battery of sera, most obtained from multiparous women or as monoclonal antibody preparations. These sera have been selected because of reactivity against specific HLA antigens. All cadaveric and potential living donors are HLA typed as well. Standard tests include typing for the class I antigens HLA-A, B, and C, and the class II antigens HLA-DR, DP, and DQ. In many centers, typing for the class II antigens is being performed with molecular techniques. With polymerase chain reaction technology and multiple probes for the DNA genetic sequences of specific antigens, more accurate typing can be obtained.

HLA typing can be of significant importance in living donor transplantation by allowing identification of the best match from multiple potential donors. Because all of the HLA antigens are found together on the major histocompatibility complex on chromosome 6, it is not necessary to type for every antigen in a living-related donor situation; typing for HLA-A, B, and DR is sufficient. Typing from multiple family members identifies the segregation of antigens and allows the selection of the best donor. Because the antigens segregate together, this set of antigens known to be on the same chromosome is identified as a haplotype. HLA matching for living donors is defined by haplotypes. A living-related transplant between siblings could have zero-, one-, or two-haplotype match. Transplants between parents and children will always be a one-haplotype match, but one-haplotype matches also can exist between more distant relatives.

For cadaver donor allocation, matching also has an important role. It is possible to type for at least six antigens, but in practical terms, typing for organ allocation involves HLA-A, B, and DR. Typing for the other antigens increases the complexity of finding a good match without significant benefit in terms of graft survival.

In living donor transplantation, an HLA match is correlated with short-term and long-term graft survival. Recipients of two-haplotype-matched living donor kidneys have near 100 percent graft survival with an extremely low rate of acute rejection. A two-haplotype match transplant should be used whenever possible, but excellent results have been achieved with completely mismatched living-related donors. If a healthy donor is available that is blood-group-compatible, the patient should be considered even with a poor match. In cadaveric transplantation, matching is used to facilitate organ allocation. There is a correlation between the degree of HLA matching and the assurance of long-term graft survival, particularly when no mismatches for HLA are found between the donor and the recipient. In the United States there is a national system for organ retrieval and allocation with mandatory sharing of six-antigen matches on a national basis. Improvement in graft survival can be seen with progressively better HLA matching in cadaveric transplantation, but with improvements in immunosuppression, the survival differences have been decreasing each year. The 1-year survival advantage for six-antigen matches over completely mismatched cadaveric transplants is about 5 percent.

Serum screening is another important histocompatibility test in renal transplantation. As a result of sensitizing events, such as blood transfusions, pregnancies, and previously failed transplants, patients may produce anti-HLA antibodies. The risk of sensitization varies with the nature of the sensitizing event. Multiparous women who have lost a transplant are at highest risk for sensitization. The consequences of sensitization are the production of antibodies against specific HLA antigens. Serum screening is performed by testing a patient's serum against a panel of lymphocytes selected to represent the known HLA antigens. Activity is defined on the basis of the percentage of cells in the panel against which detectable antibodies can be found. Sensitization is designated by the patient's panel reactive antibodies (PRA) level, which is a reflection of the percentage of cells on the panel against which the sera react. Serum screening is an important factor in predicting the likelihood of finding a suitable donor for a given recipient. A patient whose PRA level is more than 90 percent will likely have reactivity against 90 percent of potential donors. This significantly lessens the likelihood of finding a suitable donor and prolongs the patient's time on the waiting list. This is important information for counseling cadaveric transplant recipients about the likelihood of their receiving a transplant. A greater knowledge of sensitization can lessen the likelihood of patients being admitted for transplants unnecessarily.

The most important histocompatibility test in renal transplantation is the final crossmatch. This is very similar to the crossmatch test performed for blood transfusions. Cells from a potential donor and serum from a recipient are incubated together. Antibody binding is then detected using a cytotoxic technique or another procedure, such as flow cytometry or the antihemophilic globulin (AHG) test. Both of these tests depend on binding of a second antibody to anti-HLA antibodies that have already bound to the target cell.

Crossmatching is performed just before proceeding with transplantation. The use of sensitive crossmatch techniques has essentially eliminated hyperacute rejection as a problem in

renal transplantation. There are several confounding factors that must be recognized. Autoantibodies directed against lymphocytes can affect interpretation. These generally are not relevant but will generate a positive crossmatch. T and B cells are tested separately to identify class I and class II antigens, because class II antigens are found only on B cells. Positive T-cell crossmatches are responsible for the classic anti-HLA class I antibodies that are correlated with hyperacute rejection. The importance of anti-B cell antibodies is unclear, but there are isolated reports of hyperacute rejection when specific anti-class II antibodies, such as an anti-DR antibody, are identified. A positive T-cell crossmatch is considered an absolute contraindication to transplantation, and in some settings a positive anti-B-cell crossmatch may be as well. Along with blood typing, the final crossmatch constitutes the most important histocompatibility test and is mandatory before all renal transplants.

Renal Donor

Evaluation of the Living Donor

According to recent statistics from the United Network of Organ Sharing (UNOS), there are over 30,000 patients awaiting renal transplants in the United States. The cadaveric donor pool has remained static over the past 5 years, with only 4,000 to 5,000 donors realized each year. Because of the shortage of cadaveric donors, live donors, related and unrelated, have a larger role in many renal transplant programs. Although live donors account for approximately 25 percent of kidneys transplanted in the United States, in some programs the figure is as high as 50 percent.

Because of the widening gap between the cadaveric renal supply and potential recipients, definitions for who can be a live donor have expanded. Initially, only HLA-identical or one-haplotype matches were considered for live donation, and more distantly related or unrelated transplants were discouraged. Recent results with zero-haplotype-matched siblings and unrelated donors are similar to six-antigen-matched cadaver donors and approach results of one-haplotype-matched donors. What is important in live donation is not only matching but also the excellent condition of live- donor kidneys. Live-donor kidneys result in less warm ischemia and allow for longer preservation times compared to cadaveric kidneys. Living-related donors can be grandparents, aunts, uncles, siblings, and cousins. Unrelated donors may be adopted siblings, friends, spouses, or anyone with a strong emotional relationship to the recipient. The importance of this emotional relationship cannot be overstated, because this discourages financial incentives to donate kidneys. Countries that permit and even encourage financial incentives for emotionally unrelated transplants have significantly inferior results.

When patients with end-stage renal disease present for evaluation, it is important to discuss in detail the advantages and disadvantages of living- related, living-unrelated, and cadaveric renal transplantation. The advantages of live donation are excellent immediate graft function and avoidance of posttransplant dialysis, better short-term and long-term results, preemptive transplantation (i.e., avoidance of dialytic support), avoidance of waiting time for a cadaveric kidney, and in the case of HLA- identical transplants, a reduction in immunosuppressive therapy. The risks to the donor are relatively low, but there is a 1/10,000 risk of death and a 10 percent or less risk of morbidity. No definitive long-term morbidity has been demonstrated for live donors, but minor concerns have been raised about the development of mild hypertension, proteinuria, and potential trauma to the remaining kidney.

During the initial evaluation, a careful family history will eliminate obviously inappropriate donors. The presence of diabetes, hypertension, malignancy, significant cardiopulmonary disease, a history of renal disease, and age over 65 years are the primary reasons not to proceed with live donation. Potential donors can be excluded during the work-up phase of live donation as a result of proteinuria > 250 mg/24 h, a creatinine clearance of less than 80 mL/min, and significant urologic abnormalities.

The donor and the recipient should not feel pressured to donate or accept a kidney. A well-conducted evaluation of the potential donor and recipient and an in-depth explanation of live donation will help to alleviate fears.

If more than one potential donor is available, the initial determination usually is made by selecting the best match (i.e., two-haplotype is preferred to one-haplotype, which is preferred to zero-haplotype). Related donors are chosen over unrelated donors, and older donors chosen over younger donors, when the haplotype match is the same.

Once a potential donor has been initially screened and evaluated, surgical evaluation is necessary. Intravenous pyelogram and arteriograms formerly were a necessary part of the surgical evaluation, but newer CT technology can produce images of the collecting system, ureters, and bladder and enumerate the renal arteries and veins in a noninvasive manner. If minor urologic abnormalities are detected in one kidney, this would be the kidney chosen for donation. Multiple renal arteries (usually fewer than three) are not a contraindication to donation, because ex vivo techniques are available to manage multiple renal arteries. The kidney with the fewest number of renal arteries usually is chosen but the left kidney is preferable because the renal vein is longer, making the kidney easier to transplant. Table [10-19](#) summarizes the evaluation of the living donor.

Evaluation of the Cadaver Donor

There has been a significant increase in the number of live renal donations, but cadaveric donations account for approximately 75 percent of all kidneys transplanted. Cadaveric donation, however, has not kept pace with the number of recipients added to the waiting list. Identification of potential cadaveric donors by health care professionals, a sensitive approach to requesting donation by trained procurement professionals, and education are paramount to maintaining and increasing organ donation levels. Any patient who has been declared brain dead or is to be withdrawn from support, usually for a severe neurologic injury, should be considered a potential donor. Inappropriate donors may then be ruled out by the Organ Procurement Organization (OPO) on the basis of known absolute and relative contraindications (Table [10-20](#)).

Because of the extreme shortage of organs, many centers have used kidneys from every category of listed relative contraindications. Before the waiting list grew so long, only ideal donors were used. Ideal donors are young, normotensive, brain dead, free of any disease, with minimal warm ischemic time. However, increasing experience is being obtained with marginal donors. Marginal donors usually are older and may have other diseases not affecting the kidneys, including low-grade brain tumors. They also may have mild renal dysfunction or acute tubular necrosis and may have longer periods of warm ischemia as a result of cardiac arrest or prolonged hypotension. There has been renewed interest in non-heart- beating donors (NHBDs), donors whose organs are removed after cardiopulmonary arrest. Brain death cannot be declared in these patients, usually because of the presence of brain stem reflexes. Before the establishment of brain death criteria in the late 1970s, this was the sole method of organ retrieval. Because NHBDs must have cardiopulmonary cessation prior to organ retrieval, there is a longer warm ischemic period. Some centers have reported similar organ function (kidney, pancreas, and liver) from NHBDs to that of organs retrieved from heart-beating donors (HBDs). It is estimated that an increase in transplanted organs of nearly 20 percent could be expected with widespread use of NHBDs.

Donor Nephrectomy

The donor operation is carried out through a flank incision and retroperitoneal approach. The ureter is dissected as far distally as possible and up to the level of the renal vein. The renal vein, and then the renal artery, are dissected. The importance of renal angiography or spiral CT cannot be overstated in identifying multiple renal arteries and veins. Care is taken not to manipulate the kidney and renal artery excessively in order to prevent vasospasm and resultant acute tubular necrosis. Before removal of the kidney, mannitol and furosemide usually are given, and in selected cases (e.g., multiple renal arteries and a difficult dissection) systemic heparin is used, with subsequent protamine reversal. The renal artery and vein are clamped and the kidney is removed and flushed with a cold solution; any necessary ex vivo work (such as renal artery reconstruction) is completed, and the kidney is transplanted.

The most common complications after live donation include urinary tract infections, wound infections, and pneumothorax. More serious complications are rare. Long-term follow-up of live donors has not demonstrated any adverse health effects.

In the case of cadaveric donation, with the exception of older donors, whose liver and pancreas may not be used for transplantation, retrieval of kidneys usually is part of a multiorgan procurement that includes the heart, lung, liver, pancreas, and, most recently, the intestine. Fig. [10-37](#) illustrates intraabdominal multiorgan procurement of the liver, pancreas, and kidneys.

During a multiorgan procurement, the kidneys are dissected minimally to avoid arterial vasospasm. Donors usually are given 10 to 20 mg of the alpha-adrenergic blocker phentolamine to prevent renal and hepatic vasospasm, and 10,000 to 20,000 units of heparin. Many donors are given mannitol or furosemide to assure adequate diuresis before the kidneys are removed. The usual sequence of organ removal after dissection of intrathoracic and intraabdominal organs is the heart and lungs, then the liver and pancreas, and en bloc removal of the kidneys.

During in vivo flushout, and while waiting to remove the intraabdominal organs, topical sterile ice is used to cool organs more rapidly, thereby helping to minimize warm ischemic times.

Retrieval of kidneys without retrieval of other intraabdominal organs can be performed in a manner similar to multiorgan procurements, with the placement of an aortic, and occasionally a vena cava, cannula for removal of venous effluent. Alternatively, kidneys can be removed individually or en bloc without cannula placement or in vivo flushout after systemic heparinization. Ex vivo flushout is performed after the kidneys have been removed.

With brain-dead donors there is adequate time for complete dissection of the intrathoracic and intraabdominal organs. However, in donors who have had cardiac arrest just before organ retrieval, or in non-heart-beating donors, adequate time is not available for organ dissections. In these situations rapid en bloc techniques of retrieval have been developed that minimize warm ischemia and operative time while maintaining the viability of transplantable organs (Fig. [10-38](#)). All intraabdominal organs are removed en bloc after division of proximal aorta and vena cava; this is followed by retroperitoneal dissection and then division of the sigmoid colon, distal ureters, and distal aorta and vena cava. This technique is rapid, but it does require extensive ex vivo dissection for separation of the individual organs. This can be safely done because the organs have been flushed with preservation solution and have been cooled to 4°C.

Organ Preservation

The success of organ transplantation is the result of a number of factors, including the ability to preserve organs from the time of removal from the donor until transplantation into the recipient. The goal of hypothermic organ preservation is to maintain organ viability long enough that organ transport, tissue typing, and recipient preparation can be performed. The length of preservation also must be brief enough that no organ is ever discarded as a result of having exceeded the safe period of preservation.

Although the mechanisms of injury during organ preservation involve multiple factors that are complex and not fully understood, several have been shown to be important. Hypothermia is the cornerstone of organ preservation, but during hypothermia metabolic functions, such as the sodium-potassium pump, are slowed. This allows hypothermia-induced cell swelling to occur, which is injurious to preserved cells. Preservation solutions were designed to prevent this swelling and have improved the quality of preservation. Of other important factors, four have received the most attention as probable causes of injury. One is the loss of energy- generating capabilities resulting from mitochondrial damage because of a loss of precursors for adenosine triphosphate (ATP) regeneration. If the organ cannot rapidly restore a near-normal energy balance, irreversible injury with cell and tissue death results. The role of oxygen free radicals in reperfusion injury to cold-stored organs also has gained much attention. Some studies have shown that these cytotoxic agents might be involved in preservation injury, but suppression of oxygen free radical generation with various pharmacologic agents has met with limited success.

Another mechanism of preservation injury might be the activation of catabolic enzymes, such as those in lysosomes, phospholipases, and proteases. The activation of the arachidonic acid cascade and production of cytotoxic products, such as thromboxane and leukotrienes, may have a role in reperfusion injury of preserved and transplanted organs.

There are two basic methods of kidney preservation: cold storage and machine perfusion, developed in the mid-1960s by Collins and Belzer, respectively. Until 1987 the majority of renal transplant centers cold-stored kidneys with Collins' solution while a smaller number of centers machine- perfused kidneys with Belzer's machine perfusate. In 1987 Belzer developed the University of Wisconsin (UW) solution primarily to extend the preservation times of extrarenal organs such as the liver and pancreas. UW solution has essentially replaced Collins' solution for cold storage preservation of kidneys, because 85 percent of organ retrievals are multiorgan, and the UW solution also is effective in renal preservation (Table 10-21).

Citing simplicity and reduced expense, the majority of kidney transplant centers still cold-store kidneys. Machine perfusion results in a reduced incidence of delayed graft function and reduced expense for posttransplant dialytic support. The need for dialytic support after transplantation of cold- stored kidneys is 25 to 30 percent, but only 5 to 10 percent with machine- perfused kidneys, even with longer preservation times. Evidence also indicates that good immediate function is as important as matching in predicting improved long-term results. With the increasing use of kidneys from non-heart-beating donors, which have longer warm ischemic times, machine perfusion has been shown to result in dialysis requirements equal to those for cold-stored kidneys from heart-beating donors.

Surgical Procedures

Preoperative Preparation

Patients with long-standing diabetes, those with a history of cardiac disease, and those over 50 years of age should undergo cardiac stress testing or cardiac catheterization. Thallium, adenosine, and dobutamine stress tests may be used as screening tools before proceeding with cardiac catheterization. In patients with long-standing diabetes, there is evidence that cardiac catheterization should be the primary modality used to eliminate false-negative stress tests. Patients with a history of peripheral or cerebrovascular disease should have carotid Doppler ultrasonography, and pulse-volume recordings (PVRs) of their lower extremities, and a vascular surgery consultation.

Patients with poorly controlled hypertension or with chronic pyelonephritis and hydronephrosis, most likely will require a bilateral native nephrectomy before transplantation. Patients with polycystic kidney disease may require bilateral native nephrectomy if pain, hematuria, and urinary tract infections accompany their disease.

Nephrectomy poses a special problem for patients who are not yet on dialysis and who do not have a live-donor kidney that can be transplanted 10 to 14 days after nephrectomy. These patients are relegated to dialysis and must wait an average of 18 to 24 months for a cadaveric kidney.

In patients with a urologic history, a cystoscopy, cystometrogram, or voiding-cystourethrograms (VCUG) may be required to assess the bladder and upper urinary tract before transplantation. Myelomeningocele patients with an ileal loop or other urinary reservoir should undergo a loop contrast study because the transplanted ureter will be implanted into the urinary reservoir.

A history for gastrointestinal disease should be elucidated, and patients with symptoms of active peptic ulcer disease should undergo upper endoscopy. Patients with a history of pancreatitis will need a pancreatic ultrasound or CT scan and a work-up for hyperparathyroidism. Patients with symptomatic cholelithiasis should undergo laparoscopic cholecystectomy before transplantation. Some renal transplant centers also recommend cholecystectomy in asymptomatic diabetic patients with cholelithiasis because posttransplant cholecystitis is difficult to diagnose and may be associated with a higher complication rate.

Potential transplant candidates with a history of diverticulitis, especially patients with polycystic kidney disease, who have a higher incidence of diverticulosis and diverticulitis, may require a pretransplant sigmoid colectomy. Patients with a previous history of colonic polyps should have yearly colonoscopy while awaiting transplantation. Patients with active hepatitis or chronic liver disease and cirrhosis are not renal transplant candidates, but patients with hepatitis B who are Hb eAg-negative and those with chronic persistent hepatitis C may be considered.

Waiting times for cadaveric kidneys can be long, so it is important to work with local nephrologists in maintaining patient health before transplantation. A patient's transplant may need to be canceled because of the presence of active sepsis. Sepsis can occur as a result of infected dialysis grafts or catheters, urinary tract infections, dental caries, sinus infections, and pneumonias. Because of the powerful effects of immunosuppression, it is unwise to proceed with transplantation when there is active sepsis or infection.

Anesthesia

Intraoperative management of the anephric patient depends on the cause of renal failure and the presence or absence of cardiac disease. Older patients and patients with significant hypertension, diabetes, or previous coronary bypass procedures or angioplasty will require additional monitoring. This usually is accomplished with the placement of a Swan- Ganz catheter and an arterial line. Significant hypertension should be avoided, and care must be taken to prevent lower blood pressures and underperfusion of the transplanted kidney. Renally excreted anesthetics or muscle relaxants, such as gallium triethiodide, are rarely used and should be avoided. The muscle relaxant succinylcholine may be used if the patient is normokalemic. If the patient is hyperkalemic, succinylcholine should be avoided, because hyperkalemia can be exacerbated and result in cardiac arrest. The muscle relaxant atracurium is preferred. Most inhalational anesthetics are suitable for use in the anephric patient.

Intraoperative fluids should be administered under the assumption of delayed graft function even if dialysis rarely is required. This approach prevents fluid overload and the need for urgent postoperative dialysis. Renal-dose dopamine (3 to 5 mg/kg/min) may be used to enhance renal transplant perfusion. Patients usually are given methylprednisolone, mannitol, and furosemide before removal of vascular clamps and reperfusion of the kidney. If blood is needed intraoperatively, leukopoor cells should be administered to prevent sensitization of the recipient.

Surgical Technique

Usually a right curvilinear incision is made, extending from the pubic tubercle to a point just medial to the iliac crest, to the tip of the eleventh rib. The right side is preferable because the right iliac artery and vein take a slightly more superficial course than the left. The right side affords another technique of minimal dissection of the distal vena cava and common iliac artery. This technique is more rapid than dissection of the external, internal, and common iliac arteries and veins. Once the retroperitoneum is entered, the lymphatics overlying the iliac artery are ligated in an attempt to prevent postoperative lymphocele formation. Most centers transplant either kidney on the right side, but some prefer to transplant a left kidney on the right side and a right kidney on the left side, allowing the renal pelvis to be anterior in the event that a ureteral repair become necessary in the future. Others prefer to transplant a right kidney on the right side and a left kidney on the left. In the event of a second transplant, the opposite side is used. If three or more transplants are necessary, a transabdominal approach is used. In small children who receive an adult kidney, the transabdominal approach is used. The standard renal transplant is depicted in Fig. 10-39.

The donor renal artery may be anastomosed end-to-side to the common or external iliac artery or end-to-end to the hypogastric artery with 5-0 or 6-0 polypropylene sutures. Occasionally in children the donor renal artery is sewn to the distal aorta. When the kidney is from a cadaveric donor, a patch of aorta (Carrel patch) with the donor renal artery attached is sewn to the iliac artery. Multiple renal arteries are handled with a variety of ex vivo techniques, including placing all the arteries on a single Carrel patch, attaching the arteries side-to-side or end-to-side, using donor iliac artery grafts, and implanting each artery on the recipient iliac artery separately.

The renal vein is sutured with 5-0 or 6-0 polypropylene sutures to the common or external iliac vein or the distal vena cava. The presence of multiple renal veins is managed differently than with arteries. Small renal veins can be ligated because of intrarenal venous connections, but if two renal veins of equivalent size are present, some prefer to suture them together to prevent venous hypertension.

The ureteral anastomosis is performed most commonly on the recipient's bladder. A few centers routinely perform native-to-donor ureteroureterostomy or native-to-donor ureteropyelostomy. The ureter also may be anastomosed to an existing ileal or colonic loop. Two antireflux techniques commonly are used for the ureteral anastomosis. The Liche extravescical technique is preferred over the Leadbetter-Politano intravesical technique, because of simplicity. In the Liche technique, a submucosal antireflux tunnel is fashioned from the outside, whereas in the Leadbetter- Politano technique it is from an intravesical approach. Usually 4-0 or 5-0 absorbable sutures, such as chromic catgut or polydioxanone, are used in two layers (mucosa and muscle) with both techniques. A 6-mm by 17-cm double J stent may be used and can be sutured to the tip of the Foley catheter. The Foley catheter usually is removed after 1 to 3 days with the Liche technique and 5 to 7 days after a Leadbetter-Politano neoureterocystostomy.

Postoperative Care

Immediate Care

The early postoperative management of the renal transplant recipient does not differ significantly from the management of nontransplant surgical patients. Vital signs are monitored frequently along with hourly urinary output. Central venous pressure is monitored via a central line usually placed at the time of transplantation. Because of the obligatory diuresis that occurs with a well-functioning renal transplant, urine output usually is replaced with half-normal saline solution with 5% dextrose without potassium. A good rule of thumb is to replace urine output plus 30 mL/h (insensible losses) not exceeding 200 mL/h. Placing a limit on replacement of urinary losses helps to prevent excessive fluid administration. Once the early obligatory diuresis is over, fluid management is similar to that in other surgical patients. Because early delayed graft function occurs in approximately 25 percent of cadaveric transplants, fluid replacement linked to urinary output will help to prevent fluid overload and the need for urgent hemodialysis.

Most renal transplant patients can be cared for adequately on the ward, but older patients, those with a cardiac history, or those with Swan-Ganz catheters or arterial lines placed intraoperatively need a short stay in an intensive care or a step-down unit. Diabetic patients need to have blood glucose level monitored closely, and insulin may be given via a sliding scale or an insulin drip. Patients also need to have their blood pressure monitored closely, because moderate hypertension is common in the postoperative period. This results from preexisting hypertension from long-standing renal disease, postoperative pain, fluid administration, and medications that are known to cause hypertension, such as prednisone, cyclosporine, and tacrolimus (FK506). Hypertension can be treated with sublingual nifedipine, intravenous labetalol, esmolol, enalapril, or, in severe cases, nitroprusside.

Most cadaveric and nearly all related and unrelated renal transplants will function immediately, but delayed graft function can result in anuria or oliguria. The presence of anuria or oliguria in a live donor kidney should alert the surgeon to a serious problem, because these kidneys are expected to function immediately. If the patient is a first-time renal transplant recipient without high levels of circulating antibodies, emergent exploration might be indicated. Assessment of anuria or oliguria must also be evaluated with knowledge of the patient's pretransplant urine output. A normal urine output may be inadequate after transplantation if the patient produced a significant amount of urine before transplantation. The first step in assessing low urinary output is to be sure that the Foley catheter is not obstructed with blood clots. The catheter can be irrigated free of clots, or a larger catheter or one with irrigating capacity may be placed. The volume status of the patient is assessed by measuring the central venous pressure (CVP) or the pulmonary capillary wedge pressure. If the CVP is less than 12 mmHg, a 500-mL fluid bolus of isotonic saline solution should be given and repeated as necessary until the CVP rises. If this fails to increase urinary output or if the initial CVP is less than 12 mmHg, an intravenous dose of furosemide (up to 200 mg) is administered. If urine output increases, then replacement is resumed.

If urine output does not increase, especially in a kidney that was expected to function well, then a Doppler ultrasound scan to assess blood flow is indicated. If blood flow is adequate, then obstruction or a urine leak, usually at the ureterovesical junction, should be assessed with ultrasonography or nuclear scintigraphy.

If the evaluation is unremarkable, then the diagnosis is delayed graft function, which may need to be managed with hemodialysis, particularly in the presence of hyperkalemia and volume overload. Peritoneal dialysis may be used, but this usually is avoided in the immediate postoperative period because of the possibility of infection and peritoneal leakage around the newly transplanted kidney.

Technical Complications

Refinements in surgical technique and in pre- and postoperative management have significantly decreased technical complications. Early technical complications include graft thrombosis, urine leaks, bleeding, and wound infections; late complications include lymphoceles, ureteral strictures, and renal artery stenosis.

Graft thrombosis is from an arterial or venous thrombosis and in the early postoperative period is technical in origin. Unless diagnosed and operated on rapidly, there is little hope of salvaging the kidney. Diagnosis is made by abrupt cessation of urine output (which may be difficult to diagnose in a patient with good pretransplant urine output), and by Doppler ultrasonography. Graft thrombosis occurring more than 2 weeks after surgery usually is the result of severe acute rejection.

Early postoperative bleeding is unusual but may occur as a result of small vessels that were in spasm at the time of surgery. Uremic patients also have dysfunctional platelets that may contribute to bleeding. Patients should be well dialyzed before transplantation. At the time of operation, if the operative field is not dry, desmopressin acetate (DDAVP) may be administered to help with platelet function. Patients with significant bleeding manifested by hypotension, severe graft tenderness, and swelling, even in the presence of a normal hematocrit level, should undergo surgical reexploration immediately. Less significant bleeding may be present with fewer symptoms. The diagnosis usually can be made with ultrasonography, but clots may have the same echogenicity of surrounding tissues, making the diagnosis more difficult. A CT scan will delineate the hematoma and is indicated if there is a high index of suspicion. Late bleeding may occur because of a mycotic aneurysm, which is rare, or a ruptured kidney from severe acute rejection, an uncommon presentation.

Urine leaks occur most commonly at the ureterovesical junction but may occur anywhere along the length of the ureter or from the renal pelvis. Technical failure results from a ureteral anastomosis that is too loose or too tight, or from a bladder closure that is less than watertight. Urine leaks also occur because of distal ureteral slough from inadequate blood supply. This complication usually can be prevented by ensuring that ureteral length is not excessive and by preserving all lower polar renal arteries. The diagnosis can be made from decreasing urine output, lower abdominal pain, scrotal or labial edema, and a rising serum creatinine level. Ultrasonography may have determined a fluid collection that should be aspirated and analyzed for creatinine. A renal scan may demonstrate extravasation of the radioisotope beyond the confines of the collecting system and bladder. Early exploration and repair usually is required for urine leaks. Ureterovesical leaks can be repaired by reimplanting the transplant ureter, and bladder leaks repaired by primary closure. A distal transplant ureteral slough requires a ureteroureterostomy or a ureteropyelostomy from the native ipsilateral or contralateral ureter to the transplant ureter or pelvis. This usually is done over a double J stent that is left in place for 4 to 6 weeks.

Lymphoceles are perinephric collections of fluid that occur in approximately 5 percent of renal transplants, usually as a result of excessive iliac dissection and failure to ligate the lymphatics overlying the iliac artery. Less commonly, lymphoceles occur as a result of lymphatic leakage from the transplanted kidney. Lymphoceles may present with swelling over the transplant, unilateral leg edema caused by iliac vein compression, and an increased creatinine level as a result of ureteral compression. Even small inferomedial lymphoceles can cause ureteral obstruction from compression. Ultrasonography will show a homogeneous perirenal fluid collection, often with hydronephrosis. Venous Doppler ultrasonography may demonstrate compression of the iliac vein by the lymphocele (Fig. [10-40](#)). The diagnosis can be confirmed by simple aspiration, because a lymphocele has a high protein content and a creatinine concentration equal to that of serum.

Small asymptomatic lymphoceles do not require treatment, but lymphoceles that cause obstruction or venous compression must be drained. When lymphoceles are aspirated, a small catheter often is left attached to a closed drainage system. This catheter may be used to sclerose the lymphocele with povidone-iodine or tetracycline, but in most centers the preference is to create a peritoneal window surgically to drain the lymphocele intraabdominally. Lymphoceles that develop early within 2 to 4 weeks or are multiloculated may be approached via the transplant incision. Unilocular and inferomedial lymphoceles usually are approached laparoscopically. This approach results in decreased morbidity and shorter hospitalizations. If a drainage catheter was previously placed, it can be filled with saline or methylene blue to better delineate the lymphocele cavity.

Renal artery stenosis occurs in approximately 10 percent of renal transplants, usually within the first 6 months. The presentation of renal artery stenosis is sometimes subtle and the diagnosis difficult to make. Renal artery stenosis primarily presents with hypertension, which is difficult to control. Fluid retention manifested by peripheral edema, sensitivity to cyclosporine or tacrolimus, renal dysfunction, and a bruit over the transplanted kidney also may be present, but many renal transplant recipients with a normal renal artery have these manifestations. Another observation leading to the diagnosis of renal artery stenosis is worsening renal function when angiotensin-converting enzyme inhibitors, such as enalapril or captopril, are used to treat hypertension. The definitive diagnosis of renal artery stenosis usually is made angiographically (Fig. [10-41](#)). Ultrasonography and MRI angiography may be used as screening tools.

Renal artery stenosis may occur distal to the anastomosis as a result of rejection, atherosclerosis, and clamp or perfusion cannula injury. Renal artery stenosis at the anastomosis occurs more frequently when end-to-end, rather than end-to-side, anastomoses are performed. This usually is a result of faulty surgical technique or an intense fibrotic reaction to the suture used. The majority of renal artery stenoses (> 80 percent) can be corrected with balloon angioplasty at the time of angiography. Those that cannot be approached angiographically and those that have failed balloon angioplasty are repaired surgically.

Ureteral obstruction, whether early or late, usually presents with a rising creatinine level and evidence of hydronephrosis on ultrasonography. In the absence of a lymphocele, the cause of obstruction usually is a stricture in the distal ureter. Mild early hydronephrosis may be a result of postoperative edema and can resolve without treatment. Distal ureteral strictures usually are the result of ischemia and sometimes of rejection. The most effective way to diagnose a ureteral stricture is with an antegrade pyelogram (Fig. [10-42](#)).

A nephrostomy tube usually is left in place at the time of antegrade pyelography for access to the distal ureter when the hydronephrosis is relieved. Most short structures (< 2 cm) are amenable to balloon dilation and endourologic incision, but longer strictures typically require surgical repair. Strictures tend to be less amenable to ureteral reimplantation, as opposed to surgical treatment of early leaks. Native ureteroureterostomy or ureteropyelostomy are the most commonly used surgical procedures for repairing a transplant ureteral stricture.

Immunosuppression

New immunosuppressive agents have permitted immunosuppression to be tailored to the type of transplant and according to specific recipient needs. The immunosuppressive agents in use include antithymocyte globulin (ATG), OKT3, cyclosporine, tacrolimus (FK506), azathioprine, mycophenolate mofetil, and prednisone.

The number of agents available has resulted in numerous immunosuppressive protocols. After cadaveric renal transplantation, triple therapy is used for maintenance immunosuppression. Previously, triple therapy included azathioprine, cyclosporine, and prednisone, but mycophenolate mofetil, with its ability to prevent rejection, has replaced azathioprine in most protocols. Cyclosporine is administered orally (5 to 8 mg/kg) twice daily, and trough levels are monitored. Monitoring of cyclosporine trough levels can be done by high-performance liquid chromatography (HPLC), radioimmunoassay, or by fluorescent polarization immunoassay (FPIA). Mycophenolate mofetil also is given orally in doses of 1,000 or 1,500 mg twice daily, but gastrointestinal symptoms or leukopenia should prompt a lowering of the dosage. Prednisone may be given in large doses (500 to 1,000 mg) intravenously followed by a rapid taper to 5 to 10 mg daily after 3 to 6 months. Protocols that withdraw steroids completely or those that avoid steroids altogether are uncommon in cadaveric renal transplantation. The risk of acute rejection and long-term deterioration of graft function are too great to offset any gains from avoiding steroid-related side effects.

In addition to triple immunosuppressive therapy, quadruple therapy protocols use a powerful anti-T-cell agent, such as ATG or OKT3, after transplantation. The rationale for using this induction therapy is to avoid the nephrotoxic effects of cyclosporine and tacrolimus. In these protocols, ATG and OKT3 usually are used in a sequential manner. Cyclosporine or tacrolimus is not given until the creatinine level falls below 3.0 mg/dL, and it is overlapped with the administration of ATG or OKT3 until target cyclosporine levels are achieved. Although 1-year graft survival rates with the use of induction therapy exceed 90 percent, a higher incidence of viral infections, particularly cytomegalovirus and Epstein-Barr virus, has been documented.

Most two-haplotype living-related protocols have changed to a cyclosporine-based protocol with azathioprine or mycophenolate mofetil. Many two-haplotype identical protocols also include early steroid withdrawal with excellent results. One-haplotype and living-unrelated renal transplant protocols are similar to cadaveric renal transplant protocols. Some centers also have eliminated induction with OKT3 or ATG in one-haplotype living-related transplants since the introduction of mycophenolate mofetil.

Treatment of Rejection

High-dose steroids, usually methylprednisolone, are the first line of treatment for first rejection episodes. With the exception of HLA-identical transplant recipients, first rejection occurs in 40 to 50 percent of renal transplant recipients. Reports indicate that mycophenolate mofetil combined with cyclosporine and prednisone may reduce the incidence of first rejection episodes by 20 to 25 percent. Methylprednisolone is given in doses of 500 to 1,000 mg for 3 to 5 days. Some centers taper steroid doses over a longer period (7 to 10 days), but there is little evidence that a longer taper is more effective than a shorter course of pulsed high-dose steroids.

When a rejection episode is resistant to high-dose steroids, which usually is evident after 1 or 2 days, OKT3 is effective in reversing 90 percent of these rejection episodes. Rather than recycling high-dose steroids, it is better to use OKT3 earlier when the rejection appears resistant. Although OKT3 is highly effective in first rejection episodes, most centers prefer to use steroids because they can be used in an outpatient setting and are associated with fewer side effects. In patients who have antibodies to OKT3 or who have had serious side effects from OKT3-induced cytokine release, polyclonal preparations such as ATG may be used. In patients who have biopsy-proved evidence of vascular rejection, OKT3 should be considered as first-line treatment, because this form of rejection is difficult to treat. Patients who are being treated for rejection, and who are not on mycophenolate mofetil, should be switched to this agent because, unlike azathioprine and cyclosporine, it has the ability to treat established rejection. Patients with rejection refractory to steroids and OKT3 also may be treated with tacrolimus. Reversal of refractory rejection has been documented in 75 percent of patients treated with tacrolimus.

Chronic rejection, which must be differentiated from other forms of late graft dysfunction, has no specific treatment. Prevention of acute rejection episodes and earlier treatment of acute rejection episodes with OKT3 may reduce the incidence of chronic rejection. The effect of mycophenolate mofetil on the development of chronic rejection remains to be seen. Mycophenolate mofetil not only reduces the incidence of acute rejection but also may prevent the progressive arteriolopathy seen in chronic rejection.

Long-term Complications

The three most common causes of death after renal transplantation are cardiovascular disease, infectious disease, and malignancy, which are known to be increased significantly in

transplant recipients and reflect chronic long-term immunosuppression, particularly the infectious and malignancy-related deaths.

The two most common causes of graft loss are death with a functioning graft and chronic rejection. Improvements in immunosuppression and in the management of cardiovascular disease should result in longer patient graft survivals. When chronic rejection is better understood, strategies directed at its prevention and treatment will yield improved long-term graft survival.

Noncompliance, particularly in adolescent transplant recipients, may be responsible for 10 to 15 percent of late graft losses. The development of a late acute rejection episode, especially in an adolescent, should arouse suspicion. Pre- and posttransplant education and counseling may improve compliance.

Recurrent disease, especially recurrent glomerulonephritis, may result in late graft loss. The incidence of recurrence varies with the type of glomerulonephritis and occurs in 10 to 30 percent of transplants, but graft loss from recurrent disease occurs less frequently and accounts for approximately 2 percent of late graft losses.

Cardiovascular disease after renal transplantation is more common in patients with diabetes, hypertension, elevated serum cholesterol and triglyceride levels, and in patients who smoke. The risk of ischemic heart disease is three or four times greater in posttransplant patients without recognized pretransplant disease or risk factors than in the general population.

Because cardiovascular disease is a common cause of death after transplantation, elimination or reduction of risk factors through preventive measures or treatment strategies is important. Hyperlipidemia in obese patients can be improved by diet and weight loss. Minimized use of steroids, treatment of hypertension (which is known to predispose to cardiovascular disease), and exercise may help in the prevention of cardiovascular disease. If posttransplant hyperlipidemia is unresponsive to dietary modification, pharmacologic treatment with lovastatin, pravastatin, or simvastatin is indicated. These agents are hydroxymethylglutaryl (HMG)–CoA reductase inhibitors that reduce total and low-density lipoprotein (LDL) cholesterol levels. Because hepatotoxicity can occur with these agents, liver enzymes should be monitored.

Hepatic dysfunction is common after renal transplantation and may be secondary to hepatitis B, hepatitis C, or drug toxicity, particularly with azathioprine and cyclosporine. Patients who have hepatitis B before transplantation can be considered for transplantation if they are HBsAg-negative or HBV DNA-negative, because these are markers of active viral replication. Progression to cirrhosis is variable, but patients need to be informed that immunosuppression, particularly with steroids, may enhance viral replication and hasten progression to cirrhosis. Surveillance with liver enzyme determinations and possibly liver biopsies is necessary to monitor disease activity. Long-term hepatitis B immune globulin (HBIG), interferon- α , and lamivudine may be effective in limiting progression of the disease.

Hepatitis C also may lead to cirrhosis and has been more common than hepatitis B since the hepatitis B vaccine was introduced. Chronic liver disease after renal transplantation can be attributed to hepatitis C in the majority of patients, but immunosuppression appears to have a lesser role in the rapid advancement of disease than it does with hepatitis B. Patients known to be hepatitis C-positive should have liver enzymes monitored closely and perhaps a baseline liver biopsy performed before transplantation. Interferon- α has limited effectiveness, and trials with other antiviral agents, such as ribavirin, are under way.

Hepatotoxicity related to azathioprine or cyclosporine is uncommon. When other causes of hepatotoxicity have been ruled out and the pattern is cholestatic, azathioprine may be responsible and should be discontinued. Cyclosporine hepatotoxicity, manifested by increased bilirubin and transaminase levels, usually responds to a reduction in dosage.

Other gastrointestinal diseases that can lead to hepatic dysfunction include pancreatitis, cholelithiasis, and choledocholithiasis. Pancreatitis may be a result of immunosuppression, hyperlipidemia, hyperparathyroidism, stones, and viral infections. The presence of symptomatic cholelithiasis or choledocholithiasis is an indication for cholecystectomy and ERCP with papillotomy.

After transplantation, bone and mineral metabolism can be adversely affected. Early manifestations include hypophosphatemia and hypercalcemia, which may be from persistent secondary hyperparathyroidism. Asymptomatic patients with serum calcium levels in the range of 10.5 to 12.5 mg/dL should not undergo subtotal parathyroidectomy within the first year because the majority of patients will have resolution of their hypercalcemia. Patients with severe mental status changes, renal stones, fractures, pancreatitis, and those who have a serum calcium of 12.5 mg/dL or greater after 1 year should undergo a subtotal parathyroidectomy.

Long-term bone disease usually is manifested as severe osteopenia or osteonecrosis. Osteopenia with significantly decreased bone mineral density is a result of long-standing renal disease and is worse in postmenopausal women, in patients with persistent hyperparathyroidism, and in patients who have had large cumulative doses of prednisone. Minimizing steroid dosage and correcting hyperparathyroidism decreases bone loss; pediatric patients may require vitamin D supplementation, and postmenopausal women should be treated with estrogen and calcium supplements. In more severe cases, calcium-lowering agents, such as etidronate and nasal calcitonin, may be effective.

Osteonecrosis, particularly of the femoral head, is a significant long-term complication after renal transplantation, related primarily to steroid therapy. It may present with hip pain and limitation of movement, and the diagnosis is made with MRI. In severe cases a total hip arthroplasty is required.

Another common problem after renal transplantation is hyperglycemia, which requires treatment with oral hypoglycemic agents or insulin. Although 20 percent of patients develop hyperglycemia, only 5 to 10 percent require treatment, and only 1 to 2 percent require treatment beyond the first year. Steroids increase the production of glucose, impair peripheral use, and increase glucagon levels, while cyclosporine and tacrolimus decrease beta cell excretion and cause peripheral insulin resistance. Tacrolimus has a higher diabetogenic potential than cyclosporine.

Transplant-associated malignancies related to long-term immunosuppression, particularly lymphomas, are a long-term concern for renal transplant recipients. The overall incidence of malignancy is approximately 6 percent; 1 to 2 percent are lymphomas. Carcinoma of the colon, rectum, prostate, breast, and lung do not appear to have an increased incidence in transplant recipients. But skin cancer, particularly squamous cell carcinoma, has an incidence up to 20 times higher in immunosuppressed patients than in immunocompetent individuals and usually is more aggressive and presents more frequently with metastatic disease. Transplant recipients also have a higher incidence of Kaposi's sarcoma and genital neoplasms, such as vulvar, vaginal, and cervical carcinomas. Long-term transplant recipients should have dermatologic surveillance, and female recipients should have annual pelvic examinations and Pap smears.

Posttransplant lymphoproliferative disease (PTLD) is a spectrum of B-cell abnormalities that are driven by the Epstein-Barr virus (EBV). The first stage resembles infectious mononucleosis, with polymorphic B-cell hyperplasia. This is followed by the production of a subpopulation of B cells with nuclear atypia and cytogenic abnormalities, and then by the development of a malignant monoclonal B-cell lymphoma. The development of posttransplant lymphomas occurs more frequently in heavily immunosuppressed patients, especially those who have received antilymphocyte preparations, such as ALG, ATG, and OKT3. Treatment consists of drastically lowering or stopping immunosuppression, thereby restoring host immunity. Polyclonal lymphomas also may respond to antiviral treatment with acyclovir or ganciclovir, but monoclonal lymphomas respond less favorably. The development of rejection is a risk, which in renal transplant patients is not as problematic as in heart or liver transplant patients, in whom graft loss results in death. Once a monoclonal lymphoma develops, chemotherapy and radiotherapy can be used with acyclovir or ganciclovir, but the mortality rate nevertheless exceeds 80 percent.

Results

Figure 10-43 illustrates the current 5-year survival rates of patients receiving living-related, living-unrelated, and cadaveric renal transplants.

Special Problems

Several situations deserve special consideration, because the management can deviate from the standard treatment of renal transplant recipients.

Diabetes

Diabetes, particularly juvenile onset diabetes, is the single most common cause of renal failure in the United States. Diabetics account for as many as 30 percent of the patients awaiting renal transplantation. Diabetics are at significantly higher risk for heart disease and peripheral vascular disease and should be evaluated accordingly. All diabetics should undergo some form of noninvasive cardiac evaluation for ischemia, because many patients have no history of chest pain as a result of autonomic neuropathy that occurs as part of their diabetic neuropathy. The incidence of significant coronary disease is over 25 percent in patients who have had diabetes for more than 25 years, are over the age of 45 years, and have a smoking history. Coronary disease that is correctable with angioplasty or bypass is not a contraindication to transplantation. If the patient has diffuse disease with left ventricular dysfunction and is not a candidate for revascularization, then transplantation is high risk and probably contraindicated. Diabetics may have significant gastrointestinal tract problems, including gastroparesis and diabetic enteropathy. These problems frequently have a uremic component and may improve with transplantation, although severe gastroparesis can cause significant problems in the postoperative period. Diabetic candidates for renal transplantation should be evaluated for combined kidney-pancreas transplantation.

A compelling argument for transplantation in juvenile diabetics concerns the long-term outcome of these patients. Long-term survival rates of juvenile diabetics on hemodialysis are dismal, at approximately 30 percent at 5 years. In contrast, juvenile diabetics with transplants have 5-year survival rates upwards of 80 percent.

Pediatric Patients

Renal failure can be particularly devastating to children, because of its interference with normal growth patterns and, particularly in young children, the impairment of intellectual development. The time requirements for dialysis also impose particular stresses on a child's life. Pediatric patients with functioning kidneys grow more normally, have significant catch-up growth (particularly if growth hormone is administered), and potentially can have more normal lifestyles. All pediatric diabetics should be considered as candidates for transplantation.

The work-up of a pediatric patient is similar to that for an adult, but there is less need for evaluation for comorbid conditions, such as heart disease, pulmonary disease, and malignancies, because they are significantly less likely in this population. Particular attention should be given to patients' urologic status, because reflux from posterior urethra valves or neurologic problems, such as spina bifida, are the cause of many pediatric patients' renal failure. Reconstructive procedures on the urinary tract should be completed before transplantation.

Certain technical aspects of transplantation merit consideration in the pediatric patient. With very small children, the use of an adult kidney may require selection of unusual sites of anastomoses. The size of the aorta and iliac vessels also makes the procedure technically more demanding. Very small children have been successfully managed with peritoneal dialysis, intensive nutritional support, and administration of growth hormone. It is recommended that children under the age of 1 year be allowed to grow before consideration for transplantation. An allograft can be considered in patients who do poorly on peritoneal dialysis or have a complication requiring its termination.

Pediatric patients have poorer overall graft survival than adult patients, particularly in cadaveric transplantation. Consideration must be given to the type of immunosuppressive protocol used. Some clinicians recommend the use of an induction protocol with OKT3 or ATG, and triple therapy with cyclosporine, azathioprine, and prednisone. Because the outcome of transplantation for pediatric patients is considerably better with living- donor transplants, every effort should be made to identify a living-related donor.

Pediatric patients also must be monitored very closely. In the very small child, in whom posttransplant creatinine levels may be in the range of 0.3–0.5 mg/dL, biopsies may be required to diagnose rejection, as the serum creatinine level is relatively insensitive in identifying early rejection. Issues such as body image problems from drug side effects and the constraints of a strictly applied medication regimen put all younger patients, but particularly those in the teenage years, at an extremely high risk for noncompliance. Patients and their families should be counseled extensively about the risks of noncompliance.

The results of transplantation in pediatric patients have progressively improved, and 1-year graft survival rates for living donor and cadaveric transplants are in the range of 90 percent and 80 percent, respectively.

Retransplant Patients

Patients who have lost one renal allograft can be considered at high risk for retransplantation, but a significant number of these patients are pediatric and young adults who potentially can benefit from retransplantation. It is difficult to consign these patients to a life of dialysis if an initial transplant has failed. Large transplant registries have demonstrated that retransplant recipients have poor overall graft survival compared to first transplant recipients; therefore, a special approach must be taken for these patients.

Patients with a known history of heart disease or peripheral vascular disease should be reevaluated for these problems at the time they are considered for retransplantation. A number of factors can be identified that increase the risk of graft loss for retransplant candidates. These include extremely short survival of the first transplant with loss because of acute rejection in less than 6 months, and high panel-reactive antibodies levels, indicating generalized sensitization from a previous failed graft.

There is considerable evidence that extremely sensitive crossmatch techniques, such as the antiglobulin technique or flow cytometry, can reduce the risk of delayed function and early graft loss in retransplant recipients. These techniques should be used routinely for anyone receiving a second or third graft. The best outcomes have been demonstrated with induction therapy, and the use of OKT3 or ATGAM can be recommended for these patients.

ORGAN PRESERVATION

Hans W. Sollinger

Because the majority of organs transplanted are from a cadaveric source, the organ inevitably must be stored for some time after removal from the organ donor until the recipient is prepared for the transplant procedure. The organ donor and the recipient often are not in the same location, and time is needed for transport of the donor organ to the hospital where the recipient is being prepared for transplantation. This requires the use of effective, safe, and reliable methods to preserve the organ ex vivo until the transplant procedure can be performed. Acceptable preservation times vary with the organ. Most surgeons prefer to transplant the heart within 5 h after donor cardiectomy; the kidney can be stored safely for 40 to 50 h, but earlier transplantation is preferable. For the pancreas, preservation times up to 30 h have been reported, but most pancreas transplants are performed after 10 to 20 h of preservation time. Liver transplants usually are performed within 6 to 12 h after donor hepatectomy.

Preservation of the organ begins at the time a donor is identified, and the donor must be adequately maintained hemodynamically so that the organ is not injured before procurement and preservation. Injury to the organ could occur because of cardiovascular instability and hypotension. It is the donor surgeon's responsibility to assure the adequacy of organ function. During the operation to remove the organ, warm ischemia time must be reduced to a minimal amount, and the organ must be cooled rapidly, either in situ or by a well-timed back table flushout.

Hypothermia and the composition of the organ preservation solution are key factors in successful organ preservation. In cold storage of organs, the organ is rapidly cooled to approximately 4°C by flushout of the vascular system with an appropriate organ preservation solution. The flushout should remove blood as completely as possible, be delivered at a pressure that is not damaging to the organ (usually 60 to 100 cmH₂O), and in a volume that is not excessive. The volume used for each organ varies, but the liver usually is flushed with approximately 2 to 3 L, the kidney with 200 to 500 mL, and the pancreas with a similar amount. The organ is then placed in a sterile container and kept cold at 4 to 6°C.

Hypothermia is beneficial because it slows metabolism. Organs exposed to normothermic ischemia remain viable for relatively short periods (for most organs, 1 h or less). In warm ischemia, the absence of oxygen leads to a rapid decline in the energy content (adenosine triphosphate [ATP]) of the organ, a redistribution of electrolytes across the cell membrane, and a decrease in biosynthetic reactions. But biodegradable reactions continue, including the accumulation of lactic acid, a decrease in intracellular pH, proteolysis, and lipolysis. These events contribute to changes in the concentration of intracellular metabolites, and structural alterations in cellular membranes contribute to loss of viability on restoration of blood reperfusion of the organ. With hypothermia, the degradative reactions are slowed. A 10°C decrease in temperature slows the metabolic rate approximately by a factor of two. Cooling an organ from 37°C to approximately 0°C slows metabolism by a factor of 12 to 13. Hypothermia alone is not sufficient for adequate preservation for the time necessary for optimal use of cadaveric organs; the organ also must be flushed with an appropriate preservation solution.

Two requirements of any ideal preservation solution are: (1) the presence of impermeant molecules that suppress hypothermically induced cell swelling, and (2) an appropriate biochemical environment. Impermeants are agents that remain outside the cells and are sufficiently active osmotically to retard the accumulation of water by the cell. Under conditions of cold storage (cold ischemia), there is a loss of ATP, which is necessary to drive the ion pumps (sodium, potassium, ATPase) required to maintain normal cell volume. Hypothermia slows the activity of the ion pumps but has little effect on the permeability of cell membrane electrolytes, so there is a relatively rapid accumulation of sodium in exchange for potassium, a loss of the electrical potential across the membrane, and entry of chloride down its chemical gradient. This results in the accumulation of water in the cell, and this cell swelling is a major detriment to successful preservation of organs.

The University of Wisconsin (UW) solution contains lactobionic acid as the primary impermeant. Lactobionic acid has a relatively large molecular mass (358 kD) and is negatively charged. It remains outside most cells and suppresses hypothermically induced cell swelling. The UW solution also contains raffinose, a trisaccharide; hydroxyethyl starch as a colloid; and adenosine, to stimulate ATP synthesis during reperfusion of the organ.

In addition to simple cold storage, organs can be preserved by continuous hypothermic perfusion. This method, developed by Belzer in 1967, uses a machine to pump a perfusion fluid continuously through the organ. In this way, oxygen and substrates are continuously delivered to the organ, which maintains metabolism, including the synthesis of ATP and other molecules, and ion pump activity. Machine perfusion for kidneys gives superior results compared to simple cold storage. With simple cold storage, approximately 25 to 30 percent of transplanted kidneys have delayed graft function, but with machine perfusion, the rate of delayed graft function is less than 10 percent. The perfusate is similar to the UW solution, except for the impermeant. For continuous perfusion, gluconate is used in place of lactobionic acid.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 11 - Surgical Complications

Josef E. Fischer
Elliott Fegelman
Jay Johannigman

OPERATIVE RISK

Operative risk is defined as the sum total of abnormalities of all organ systems and their interactions that determine the outcome of an operation. The approach used in determining operative risk is to identify the patient at risk, identify the organ system(s) at risk, prevent or protect against the complication or the failure of that organ system (especially those aspects of organ instability or inadequacy that are reversible), and improve the outcome. The organ systems that determine outcome include: (1) cardiac, (2) pulmonary, (3) renal, (4) hepatic, (5) hemostatic, (6) nutritional-immunologic, and (7) vascular.

Cardiac Risk

One means of estimating cardiac risk is to use Goldman's cardiac risk index (Table 11-1). The focus of the Goldman classification in patients undergoing general anesthesia is the history of a previous myocardial infarction. If the patient has not had any previous myocardial infarction, the risk of cardiac death is between 1 and 1.2 percent. If more than 6 months has elapsed between the cardiac infarction and the current operation, there is a 6 percent risk. If a transmural infarct has occurred less than 3 months before operation, the risk of cardiac death is between 16 and 37 percent.

Factors that predispose to the occurrence of life-threatening cardiac events in the perioperative period include a history of any of the following: (1) infarction within 6 months, (2) congestive heart failure, (3) arrhythmias, (4) aortic stenosis, (5) emergency or major surgery, (6) age greater than 70 years, and (7) poor medical condition.

Significant peripheral vascular disease should alert the surgeon to consider the cardiac risk. If the patient gives a history of angina pectoris, it should be determined whether there has been new onset of unstable angina.

The electrocardiogram (ECG) and hematocrit level are significant. A cardiac stress test is indicated to identify those patients at coronary risk. A positive cardiac stress test includes any or all of the following: an ST depression of more than 0.2 mV, an inadequate heart rate response to stress, or hypotension. The most sensitive examination of cardiac risk is the inability to perform a bicycle exercise for 2 min and achieve a heart rate higher than 100. Data from rest and exercise radionuclide ventriculography provide little additional information but can provide useful information.

The principal form of treatment for patients with cardiac risk because of an antecedent myocardial infarction is (if possible) to delay operating for approximately 6 months after the myocardial infarction. Some patients should be admitted to the intensive care unit the day before operation, and oxygen consumption, oxygen delivery, mixed venous oxygen saturation, and cardiac output should be optimized. In patients with significant angina, angioplasty or coronary bypass procedure may be necessary before any major surgical procedure is undertaken. In some situations simultaneous cardiovascular revascularization and carotid endarterectomy are performed.

For patients in congestive failure, the use of calcium channel blockers or beta blockers, digitalization with cardiac glycosides, and diuresis are part of the therapeutic armamentarium. Patients with rapid atrial fibrillation should have their cardiac rates controlled. If the cardiac rhythm cannot be returned to normal sinus rhythm with the use of glycosides, quinidine, or procainamide hydrochloride (Pronestyl), cardioversion should be considered. If cardioversion is attempted, the patient should first be given anticoagulants to prevent embolization.

Pulmonary Risk

Pulmonary risk factors include smoking, obesity, advanced age, and industrial exposure. The patient at pulmonary risk can be identified by simple functional tests such as walking up a flight of steps or blowing out a match with unpursed lips from a distance of 8 to 10 inches (20–25 cm). If arterial blood is drawn with the patient inspiring room air, it is not the reduced P_{O₂} that identifies a patient significantly at risk but rather a P_{CO₂} of greater than 45 mmHg, suggesting a serious diffusion defect. Other easily identifiable risk factors include a *maximum breathing capacity (MBC)* of less than 50 percent of that predicted, or a *1-second forced expiratory volume (FEV₁)* of less than 2 liters. MBC is a highly subjective test, depending on the patient and the encouragement of the examiner exhorting the patient to breathe maximally. Patients with pulmonary artery pressure higher than 30 mmHg also are at increased pulmonary risk (Table 11-2).

The most sensitive test for patients undergoing thoracotomy is the exercise oxygen consumption (V_{O₂}). Patients with a maximum V_{O₂} of less than 15 mL/kg/min usually will have great difficulty; if the consumption is greater than 20 mL/kg/min, postoperative pulmonary complications are less likely.

The use of saturated solutions of potassium iodide as an expectorant, incentive spirometry, chest physical therapy, and postural drainage should be considered. Antibiotics are appropriate for patients with bronchiectasis. In the postoperative period humidified gas, chest physical therapy, incentive spirometry, and postural drainage are beneficial. Use of patient- controlled analgesia (PCA), which minimizes narcotic use or postoperative epidural analgesia, decreases the incidence of pulmonary complications.

Because most lung damage is a result of smoking or industrial pollution, cessation of smoking is essential for patients who are to undergo long elective procedures; 8 weeks cessation preoperatively is required for maximal benefit. Pulmonary physical therapy may improve the operative risk. Expectorants, physical therapy, incentive spirometry, humidified air (humidified oxygen is not necessary), and, in the case of bronchiectasis, antibiotics based on cultures of the patient's sputum flora are all appropriate to improve operative risk. With good preparation, thoracotomy can be carried out in patients with an FEV₁ of less than 1.0 L and an MBC of 35 to 40 percent.

Renal Risk

Renal abnormalities are reflected by elevations in blood urea nitrogen (BUN) and creatinine levels. Serum abnormalities are not manifest until more than 75 to 90 percent of the renal reserve is lost (Fig. 11-1). An elevation of BUN or creatinine that is not because of dehydration generally means that renal function is compromised by 75 to 90 percent. Increased levels of BUN or serum creatinine in a hydrated patient should be quantified by measuring creatinine clearance:

Cl_{Cr} = (140 - age) × weight (kg)(mL/min) / 72 × Cr (serum) mg/dL

Reversible causes of renal insufficiency should be identified and corrected. These include infection, uncontrolled hypertension, obstruction, and dehydration. The patient should be admitted preoperatively to a monitored bed the night before surgery, a pulmonary artery catheter placed, and the patient hydrated with intravenous saline solution to the optimal filling pressure in order to protect the kidneys. Aminoglycosides should be avoided whether administered systemically or for bowel preparation.

In the postoperative period, if severe hyperkalemia supervenes (greater than 7.5 mEq/L), with accompanying ECG changes, intravenous calcium should be administered and followed shortly thereafter by 50% dextrose, 10 units of insulin, and intravenous bicarbonate. Sodium polystyrene sulfonate (Kayexalate) can be given by mouth or by enema; 5 g should be administered. Essential amino acids and hypertonic dextrose solution, given as total parenteral nutrition (TPN), may lower the potassium level. As potassium enters the cells, it might be necessary to add potassium. If the BUN level approaches 100mg/dL, there are difficulties with clotting factors and platelets are dysfunctional, resulting in gastrointestinal bleeding, which further elevates the BUN level. Peritoneal, hemodialysis, or continuous ultrafiltration occasionally is required.

Hepatic Risk

Hepatic dysfunction is best estimated by the Child-Pugh criteria, which enumerate synthetic functions, including albumin, prothrombin time, excretory functions (bilirubin), nutritional status, the presence or absence of ascites, and encephalopathy (Table 11-3). In cirrhotic patients, the mortality accompanying noncardiac surgery is less than 5 percent for Child-Pugh Class A, between 5 and 10 percent for Class B, and between 20 and 50 percent or higher for Class C categories. When the blood ammonia concentration is higher than 150 ng/dL, an 80 percent mortality can be expected. When the albumin level is below 2.0 g/dL, a similar mortality is anticipated. Few patients with a bilirubin level higher than 4 mg/dL as a result of hepatic dysfunction survive an operation requiring a general anesthetic. Similarly, with a prothrombin time prolonged more than 2 seconds, a mortality of 40 to 60 percent can be expected. A point system for patients with liver disease undergoing nonshunt operations uses serum albumin concentration, prothrombin time, the presence of encephalopathy, and a history or presence of varices (Table 11-4, Fig. 11-2). Patients with hepatic dysfunction generally die of a high-output cardiovascular failure and low peripheral resistance.

In an elective situation, abstinence from alcohol, where alcohol is the damaging agent, is probably the most important feature for improving hepatic function. Patients with cirrhosis receive a large proportion of calories from carbohydrate, especially from alcohol, a cycle that it is necessary to break. These patients often are intolerant to the protein that they need because they are hypercatabolic, requiring 1.1 g of amino acids/kg/24 h, instead of the normal 0.55 g of amino acids/kg/24 h. Protein tolerance may be improved by the use of branched-chain amino acid–enriched nutritional mixtures or branched-chain amino acids alone. In patients with ascites, the conversion of uncontrollable ascites to ascites that can be controlled with medications substantially improves the operative risk. Spironolactone (Aldactone) and furosemide (Lasix) combined with fluid restriction to 1500 mL/day are therapeutic. Restriction of sodium to 500 mg/day is desirable but rarely achievable without the use of salt- free foods. A sodium limit of 2 g/day is more realistic.

Hemostatic Risk

See Chap. 3.

Nutritional-Immunologic Defects

The difficulty in assessing nutritional risk is that it is impossible to predict which malnourished patient is at risk for the immunologic abnormalities that accompany malnutrition. Statistically, malnourished patients as a group experience a higher complication rate. The defining characteristics of a severely malnourished patient include a weight loss of more than 15 percent over the previous 3 to 4 months, a serum albumin level of less than 3.0 g/dL, anergy to *injected* skin-test antigens, and a serum transferrin level of less than 200 mg/dL. If these characteristics are present, enteral or parenteral nutrition to improve the patient's well-being should be undertaken for 4 to 5 days preoperatively to normalize retinol-binding protein, thyroxin-binding prealbumin, and transferrin. Patients who do not have severe malnutrition should not be prepared for operation with total parenteral nutrition; in this group TPN increases the risk of nosocomial and non-catheter-related infections without improving other outcomes sufficiently to risk the increased infection rate. In a Veterans' Affairs Cooperative Trial, the severely malnourished patients' improvement in nutritional status was found to be correlated with a lower mortality and fewer septic complications.

Once the malnourished patient has been identified, nutritional preparation and reparation are undertaken. This may be achieved via nasoduodenal feeding tube at home for 7 to 10 days in most elective cases. If parenteral nutrition is required, patients can be discharged on home parenteral nutrition prior to operation.

Vascular Risk (Thromboembolism)

See Chap. [21](#).

DIABETES MELLITUS

Diabetes mellitus occurs in 2 to 3 percent of the general population, with a higher rate among older people. In two series the disease was discovered in the perioperative period in 16 percent and 23 percent of patients. The most commonly associated operative procedures were related to vascular disease, but in a high proportion of patients diabetes was discovered before an emergency procedure. Diabetic patients represent a special challenge during surgical care because impairment of the homeostatic mechanism for glucose may result in ketoacidosis if untreated or hypoglycemia if overtreated, and also because of the associated incidence of generalized small-vessel vascular disease.

Pathophysiology

The basic defect in diabetes is a lack of metabolically effective circulating insulin. The elevated blood glucose level is a result of deficient utilization on the part of peripheral tissues and increased output of glucose by the liver. In diabetes the breakdown of fatty acids is increased. Metabolism of the ketone bodies is limited, accumulating in the bloodstream and eliminated via the kidneys. Glycosuria produces an osmotic diuresis that is enhanced by the presence of ketone bodies with the associated loss of sodium and potassium. Evaluation of decompensated diabetes includes not only measuring the blood glucose level but also measuring serum acetone and electrolyte levels, carbon dioxide combining power, and blood pH.

Anesthesia may affect carbohydrate metabolism. Hyperglycemia may be increased by an accelerated breakdown of liver glycogen and a concomitant catabolism of muscle glycogen, with the formation of lactic acid. The anesthetic agents affecting glucose catabolism cause an exaggerated hyperglycemic epinephrine response and an increased resistance to exogenously administered insulin.

The stress of an operation aggravates hyperglycemia because of the increased secretion of epinephrine, growth hormone, and glucocorticoids. Increased epinephrine secretion results in an increased breakdown of liver glycogen to glucose, which is released into the general circulation. The glucocorticoids also increase hepatic glucose output via mobilized protein and exert an anti-insulin effect by stimulating a circulating insulin antagonist. The effects of epinephrine and glucocorticoids are offset by an increased secretion of endogenous insulin in the normal person but can require the administration of larger doses of insulin in diabetic patients. Treatment is directed at preventing ketoacidosis, hyperosmolar nonketotic coma, decreased cardiac output with associated poor peripheral perfusion, electrolyte imbalance, impaired polymorphonuclear leukocyte phagocytosis, and decreased wound healing, all of which are related to uncontrolled diabetes.

Management

For the diabetic patient essential laboratory studies include hemoglobin level determination, white cell count, urinalysis for glucose and acetone, fasting and timed postprandial blood glucose level determination, serum electrolyte levels, including potassium and carbon dioxide combining power, BUN level, and, for older patients, serum cholesterol level determination and electrocardiography. Measurement of arterial blood gases may be necessary if there is suspicion of ketoacidosis. Diabetic patients should have preference on the operative schedule to minimize the effects of fasting and ketosis. Preoperative medication should be kept to a minimum because diabetic patients, particularly elderly ones, are sensitive to narcotics and sedatives, and there is a danger of hypercapnia and hypoxia. The choice of anesthesia should not be influenced by the presence of diabetes. Spinal anesthesia has little tendency to evoke hyperglycemia apart from the stress of the operation. Nitrous oxide, trichloroethylene, and halogenated hydrocarbons have the least effect on carbohydrate metabolism. The degree of control during the perioperative period should be assessed by serial determination of the blood glucose level and urinalysis for glycosuria and acetonuria. It is safer to permit mild glycosuria and minimal elevation of the blood glucose level in the perioperative periods, particularly in the elderly and in cardiac patients. In the patient with postoperative hypotension, blood glucose level should be measured to rule out hypoglycemia as a cause.

Patients with mild diabetes mellitus frequently do not require insulin, and dietary control is sufficient. The cornerstone of all diabetic management is the dietary or parenteral intake. The preoperative diabetic intake should contain 140 to 200 g of carbohydrates, 60 to 100 g of protein, and with adequate vitamins and minerals should furnish 1200 to 2100 kcal daily. If parenteral fluids are required, there is some advantage to using fructose or sorbitol, which can be taken in amounts up to 50 g daily in the diabetic patient, but fructose can cause lactic acidosis. The goal of the dietary or parenteral fluid regimen is to keep the patient free of acetonuria without excessive hyperglycemia. Patients using oral agents with well-controlled diabetes should continue the use of these drugs until the day before operation, particularly if the medication is tolbutamide or phenformin. With longer-acting agents, such as chlorpropamide, the drug should be discontinued 72 h preoperatively if the administration of insulin is contemplated. Patients who take tolbutamide preoperatively usually require insulin during and immediately after a major operation. Patients receiving chlorpropamide usually do not require insulin during the immediate perioperative period.

Insulin Therapy

Several protocols for the administration of insulin have been proposed (Table [11-5](#)). One popular method of treatment uses a regimen in which the daily carbohydrate requirement is divided into four equal doses and given parenterally as 5–10% dextrose in water every 6 h. This initiation of the parenteral glucose infusion is accompanied by the subcutaneous injection of unmodified regular insulin in doses equal to approximately one-fourth the dose of insulin that the patient required prior to operation. Urine is checked regularly, and supplementary doses of crystalline insulin are given as indicated. Depending on the extent of glycosuria, 4 to 10 units of additional insulin is provided for each unit of positivity. Larger doses may be indicated when acetonuria, severe stress, infection, or marked hyperglycemia is present. The advantage of this method is that glucose and insulin are given at regular intervals, permitting adjustment in the dose during the day. It is preferable to monitor blood glucose levels. The major disadvantage of this regimen is that inadvertent interruption of glucose infusion may result in hypoglycemia. With this regimen, slight glycosuria is preferable provided there is no acetonuria.

The second basic regimen is for patients whose diabetes is under control with single-injection therapy using long-acting insulin and in whom a complicated postoperative course is not anticipated. On the day of operation, the patient receives 50 g of glucose in 1000 mL of solution. When the intravenous solution is started, insulin is administered at one-half the daily dose of that previously required. After operation and return to the recovery room or ward, the remainder of the usual daily dose of insulin is given subcutaneously. The amount of insulin given on the day of operation approximates that given the previous day. On the day after operation, the usual dose of insulin is given in the morning before breakfast or at the same time that an intravenous infusion is started. Modifications of this approach use small doses of regular insulin subcutaneously during the postoperative period based on the extent of glycosuria or, preferably, the serum glucose level. In patients treated with single daily injections but whose diabetes is not under control before operation, conversion to a regimen of soluble insulin is indicated.

Severe hyperglycemia in patients undergoing major operations is more effectively managed with intravenous regular insulin. The problem of insulin absorption by the fluid container has been overcome by the use of plastic containers, high concentrations of insulin, small amounts of albumin, and flushing the system. A specific infusion protocol is outlined in Table [11-6](#).

A simplified protocol has been proposed by Woodruff and associates. The patients receive their evening dose of insulin the preoperative day, but no subcutaneous insulin on the morning of surgery. The patient is scheduled as the first case of the day. Insulin and glucose are controlled with two separate infusion pumps; one pump infuses 5% dextrose in lactated Ringer's solution at 2 mL/kg/h, and the other dispenses insulin from a plastic bag containing 250 mL sodium chloride to which 50 units of U-100 regular insulin has been added. The rate of insulin infusion is based on the serum glucose level. Twenty units per hour is infused for glucose levels above 200 mg/dL, but no insulin for levels below 80 mg/dL. The surgical procedure is not begun until the level is below 200 mg/dL. Insulin therapy during emergent surgery or surgery complicated by infection requires greater amounts of insulin to maintain serum glucose levels below 200 mg/dL. In extreme cases, bolus injection of 0.1 to 0.4 unit/kg may be required as an additive.

Ketoacidosis

The preparation for surgical treatment of a patient with ketoacidosis is critical. Ketoacidosis may masquerade as a surgical emergency. The patient with frank diabetic coma is no candidate for surgical treatment regardless of the indication. Crystalline insulin should be used in all cases to establish control. Page and associates reported effective management of diabetic coma with continuous low-dose insulin infusion using an average of 7.2 units/h. Plasma glucose, ketone bodies, and free fatty acids decreased 58 percent in 4 h. The associated dehydration and electrolyte abnormality must be corrected; most patients with advanced coma require an average of 2 to 4 liters of fluid to overcome the dehydration. The serum potassium concentration should be determined at 6 to 8 h intervals. Potassium is added to the fluid in quantities of 40 mEq/L administered at a rate no higher than 25 mEq/h. The need for potassium usually does not exceed 80 mEq. There generally is no need to add glucose to intravenous fluid unless the blood glucose level falls below normal. Gastric atony is a frequent accompaniment of diabetic ketoacidosis, and suction frequently is required to minimize pulmonary aspiration. It is possible to correct ketoacidosis in sufficient time that the patient's surgical status is not compromised.

Nonketotic Hyperglycemic Hyperosmolar Coma

Hyperosmolar dehydration and coma is a relatively uncommon syndrome that usually occurs in elderly diabetic or nondiabetic obese patients and in patients receiving total parenteral nutrition. It may be an early indication of sepsis. The blood glucose level is frequently above 1000 mg/dL, and ketone bodies are absent from the plasma and urine. Treatment consists of large amounts of hypotonic solutions and intravenous insulin, often as much as 200 units of regular insulin per 24 h. If total parenteral nutrition is an inciting agent, it should be stopped. Marked lowering of the blood glucose level can result with small doses of insulin; a test dose of 10 units can be given to determine responsiveness. Sodium and potassium also must be given because large amounts of these ions are lost in the urine.

GENERAL CONSIDERATIONS

The response to injury and surgical procedures includes antidiuresis, an increase in extravascular volume, fever, and tachycardia. These are known to be the result of the release of cytokines and other agents. In patients undergoing general anesthesia, fever on the first postoperative night is usually attributed to atelectasis, but is probably the result of a resetting of the central thermostat to combat the hypothermia that occurs in the operating room. As heat preservation techniques become more widespread, it is possible that such fever will be eliminated.

Urine output falls, normally because of the release of *antidiuretic hormone (ADH)*. A tendency to hyponatremia is present in the immediate postoperative period. After intraabdominal procedures, diuresis that occurs on the second to the fourth postoperative day is a signal of normal convalescence, with reabsorption of fluid not only from within the bowel with return of function but also from within the third space and the peritoneal cavity. Diuresis coincides with decreased secretion of aldosterone and ADH because of refilling of intravascular volume. Because the degradation rate of albumin normally is related to the amount in the extravascular space, the outflow of albumin from the intravascular space to the interstices results in increased catabolism of albumin.

Ileus of the colon and stomach persists for 2 to 5 days after open intraabdominal procedures, but for a considerably shorter period after minimally invasive surgery. Return of intestinal motility is one area in which minimally invasive surgery appears to have a distinct physiologic advantage, explaining the shorter length of hospital stay. After open abdominal procedures, the small bowel continues to function throughout the postoperative period, allowing the use of enteral nutrition in the immediate postoperative period.

Wound pain can be severe for approximately 48 to 72 h. A considerably reduced requirement for pain medication can be achieved by the use of a subcuticular closure reinforced with Steri-Strips. The metabolic response to surgery and the postoperative wound pain can be improved by the use of adequate amounts of local anesthesia, even if the patient is undergoing general, epidural, or spinal anesthesia. Use of these agents at the beginning of the procedure decreases the metabolic response to surgery. When used at the end of the procedure, a mixture of short- and long-acting agents decreases the use of pain medication in the immediate postoperative period, and in small incisions may completely obviate the use of narcotics.

Postoperative fatigue may be the result of general anesthesia as much as the operation. It is essentially absent in patients in whom the afferent nerves have been severed, e.g., patients with paraplegia. In the elderly, it may take up to 2 or 3 months for the patient's fatigue to ameliorate. Absence of physical activity during convalescence results in extended fatigue. Infection or other untoward events in the postoperative period also result in a prolongation of fatigue.

FEVER

Pathophysiology

Fever is a disorder of normal body thermoregulation that is controlled by the anterior hypothalamus. Core temperature and its diurnal variation are centered on a set point, normally between 98 and 99°F (37°C). There is speculation that the evolution of fever arose as a protective mechanism to combat infection. It is recognized that certain organisms, such as bacteria and viruses, are heat-sensitive and can be destroyed in vivo by artificially inducing fever. Various pathophysiologic mechanisms, such as pyrogens, are responsible for the generation of fever. Pyrogens may arise from infectious agents such as viruses, bacteria, or fungi as a result of substances released by these organisms or because of the presence of these organisms. In addition, antigen-antibody complexes, steroids, and other inorganic substances have been demonstrated to produce experimental fevers.

All pyrogens appear to evoke a common mediator, endogenous pyrogen or interleukin-1, a monokine produced by leukocytes. Temperature-sensitive preoptic neurons reside within the hypothalamus. Interleukin-1 generates fever by altering the activity of temperature-sensitive neurons located in the anterior hypothalamus. When the set-point is raised, the body's mechanisms for increasing temperature, including heat conservation and increased heat production, are brought into play. The actual body temperature reaches that point a few hours after the set- point has raised. Interleukin-1 is extremely potent; only a few nanograms can affect the hypothalamus and increase core temperature.

When the hypothalamic thermostat is suddenly raised to a higher setting by pyrogen, the blood temperature remains relatively low. The individual feels chills and the skin is cold because of the vasoconstriction induced to conserve body heat. Shivering can occur, a powerful mechanism of heat production. This continues until the blood temperature reaches the set- point. If the set-point drops suddenly, the patient then goes through the flush phase, or *crisis*. During this phase the body attempts to rid itself of excess heat by vasodilatation and sweating. Before the advent of antibiotics, physicians anxiously awaited the flush phase, knowing that the fever would soon resolve.

Fever per se usually is not a significant physiologic problem unless core temperature is elevated above 105°F. The most important role for fever in a critically ill surgical patient is providing an early-warning sign for infection or inflammation. Potential sources include injury, catheters, urinary tract, lungs, surgical sites such as wounds or anastomoses, or thrombophlebitis in the pelvic veins. Drug reactions are another common cause of fever.

Perioperative Fever

(Table 11-7). Between 27 and 58 percent of patients develop a fever for at least 24 h after an operative procedure of modest magnitude. Persistence of a fever for 24 h is not cause for alarm unless it is high and associated with systemic symptoms such as rigor, hypotension, disturbances in mentation, decreases in urine output, or septic shock. If the fever persists for 48 h, it is considered significant. Fever in the immediate postoperative period usually is not serious, is not very high, and is self-limited. It is more likely that the postoperative fever is the result of overcompensation of the set-point in a patient who was cold.

The presence of fever on the first postoperative night is common, usually ascribed to atelectasis, but the usual sources of serious fever, which include wound cellulitis, urinary tract infection, preexistent central venous catheter fever, or drainage spread of an infected focus should not be disregarded. A delayed transfusion reaction, allergic in nature, can be the cause of fever.

Malignant Hyperthermia

This is a rare anesthetic complication that was first reported in 1960, when ten members of an Australian family died after being exposed to a general anesthetic. The incidence is estimated to be approximately 1 in 100,000 general anesthetic procedures. The syndrome consists of a rapid rise in body temperature, usually during the initiation of a general anesthetic after administration of succinylcholine or potent inhalation agents, particularly halothane. Metabolic acidosis and electrolyte imbalances quickly develop, with associated hypercalcemia. In most patients there is hypotonicity of skeletal muscle resulting in the acidosis. The final stages of the event are marked by temperatures approaching 42°C, oxygen desaturation, hypercapnia, and cardiac dysrhythmia.

Prevention is the safest method of limiting the risk to susceptible patients. A family history of complications associated with anesthetics is a warning of this possibly lethal complication. Once the syndrome unfolds, dantrolene is administered intravenously in a dose of 1 mg/kg and repeated as necessary to a total dose of 10 mg/kg. Support measures are initiated promptly, including positive pressure ventilation on 100 percent oxygen, correction of the acidosis and electrolyte imbalance, cooling blankets, monitoring of urine output, and treatment of possible myoglobinuria. After the acute episode, oral administration of dantrolene up to 1 to 2 mg/kg four times a day may be necessary for 1 to 3 days to prevent recurrences.

Fever Within 24 Hours

These fevers usually are attributed to atelectasis or the failure to clear pulmonary secretions. The sources, including wound, thrombophlebitis, urinary tract infection, and catheter sites, also should be evaluated. It is a waste of resources to obtain routine chest x-rays on patients who have a postoperative fever without systemic symptoms or supportive physical findings. The practice of culturing sputum, blood, urine, and blood withdrawn from a catheter in the absence of systemic symptoms is unnecessary. White blood counts are often elevated after operations and rarely determine whether the fever is a manifestation of a significant complication. High fevers with systemic symptoms, such as rigors, hypotension, and changes in mentation, usually are associated with severe wound complications, such as necrotizing fasciitis or an intestinal leak.

Fever at 24–48 Hours

Fever at 24 to 48 h usually is attributed to respiratory complications. Catheter-related problems, especially preexistent catheters or catheters placed in the operating room, also might be the source of fever. The pattern in catheter septic complications is that the fever generally does not cross the baseline. The wound should be carefully inspected for cellulitis and evidence of necrotizing fasciitis or clostridial myositis.

Fever After 48–72 Hours

In patients who are afebrile during the first 72 h, fever usually is a manifestation of a significant complication. Fever can be related to thrombophlebitis, especially in patients with previous episodes of thrombophlebitic complications.

The most common cause of fever after 72 h is wound infection, which may be latent or may be associated with increased wound pain. Urinary tract infections should also be suspected, particularly in patients with catheters. Less common infectious complications include pneumonitis, acute cholecystitis (especially acalculous cholecystitis in patients who are immobile or have received large volumes of blood), idiopathic postoperative pancreatitis (though this occasional complication usually is manifest in the immediate postoperative period), and drug allergy. Unusual nosocomial infections can occur in immunocompromised patients, especially transplant patients. An occasional patient sustains liver necrosis from a chlorinated hydrocarbon, but usually these are evident in the early postoperative period. Hepatitis B, hepatitis C, cytomegalovirus, and other viral illnesses may develop after transfusion.

Candidiasis may complicate intravenous total parenteral nutrition. *Candida* may not grow out on initial blood cultures while the patient is ill. If *Candida* is discovered growing from one or two other sites, such as skin, sputum, or urine, the patient should be treated with amphotericin B, and the TPN line should be removed.

Fever occurring after 1 week is almost always an indication of serious complications unless it is the result of a drug allergy. Late infections can be the result of a leaking anastomosis, an abscess adjacent to the anastomosis, or a deep wound infection that is suppressed by antibiotics.

WOUND COMPLICATIONS

Wound Infection

Predisposing Factors

Wound contamination occurs in the operating room, but not all wounds harboring bacteria become infected. While much of the contamination takes place within the operative field (e.g., a violated hollow viscus), the other major source of wound contamination is the environment. Efforts should be taken to reduce the bacterial count in the operating room. Deterrents include barriers, changes of clothing, appropriate covering of facial and other hair, and the use of scrub suits that minimize the shedding of personal bacteria. Shoe covers also can be worn, but their efficacy is questionable. Floors and walls should be mopped down between cases with antiseptic solution to decrease the bacteria ambient in the operating room. Ultraviolet light at entrances to sterile areas may be efficacious; frequent air changes to 20 times a minute are important.

Sterilization should be adequate. Breaks in technique are particularly important. The contamination of gowns, gloves, and instruments should be promptly remedied. Gloves should be immediately changed when holes are detected. It is important to wall off the wound and viscera and to keep them moist, especially during long procedures.

Staphylococcus aureus is the most frequently involved offending organism. Enteric organisms frequently contaminate wounds when bowel operations are performed. Hemolytic streptococci are responsible for 3 percent of wound infections. Occasionally “surgical scarlet fever,” may complicate these infections. Other less common pathogens include enterococci, *Pseudomonas*, *Proteus*, and *Klebsiella*.

The incidence of wound infection developing in clean, atraumatic, and uninfected wounds is between 3.3 and 4 percent. Clean wounds without emergent operation, drained wounds, or stab wounds have an incidence of 7.4 percent. The figure rises to 10.8 percent when the bronchus, gastrointestinal tract, or oropharyngeal cavity is entered in procedures with unusual contamination. Classification of wounds is given in Table 11-8.

With breaks in surgical technique, infection rates rise to 16.3 percent, and in operations involving perforated viscera, the rate of infection is reported as high as 28.3 percent. In the

latter situation, consideration should be given to delayed primary closure or to prolonged closed-suction drainage of a primary closure.

The rate of wound infection rises from 4.7 percent in the subpopulation of patients who are 15 to 24 years of age to 10 percent in patients over the age of 65 years. Diabetes is not an independent risk factor when adjusted to age, while steroids increase general wound infection rates from 7 to 16 percent. Obesity doubles the infection rate. Remote infections increase wound infection rates. Duration of operation is an important variable; 3.6 percent of procedures that take 30 min or less become infected, while 18 percent of procedures over 6 h in duration are followed by infection. Patients who are malnourished have an increased incidence of infection, but many factors, including cancer and prolonged operations, contribute as variables. Temperature is an important determinant of wound infection; patients who are kept warm during operation have a much lower incidence than patients who are allowed to get cold. Carriers of *S. aureus*, operating room personnel, or the patients themselves increase wound infection. Factors influencing wound infection are presented in Table [11-9](#).

Prevention

Wound infection rates can be minimized in a variety of ways, including (1) skin preparation, (2) bowel preparation, (3) prophylactic antibiotics, (4) meticulous technique, (5) temperature maintenance, and (6) appropriate drainage.

Any damage to the skin, particularly if skin preparation is performed the night before surgery, increases wound infection rates. Fine hair removal is not necessary unless it technically interferes with wound closure. If hair removal is required, clipping immediately before surgery or the use of a depilatory agent results in a wound infection rate of 2 percent versus 5 percent after shaving or clipping the night before or shaving preoperatively.

Bowel preparation decreases wound infection. Mechanical preparation with clear liquids and cathartics is most effective. Although it is difficult to demonstrate reduced wound infection rates with the use of adequate and appropriate systemic perioperative antibiotics, most authorities believe that there is a cumulative beneficial effect. The Condon-Nichols erythromycin base–neomycin is most commonly used, but metronidazole-based regimens are being used with increasing frequency.

Systemic antibiotics should be given immediately before the incision is made and serum levels maintained throughout the procedure above the minimally inhibitory concentration. Redosing is necessary in long procedures. Table [11-10](#) lists appropriate antibiotics. There is disagreement over continued use of antibiotics after the wound closure. Many believe that this is not necessary except in massive contamination, in which case prophylactic antibiotics become therapeutic. Some authorities believe that the infection rate after a clean elective operation can be halved from 4 to 2 percent with a single dose of a first-generation cephalosporin; others are concerned about the spread of resistant organisms and believe that widespread prophylaxis is unnecessary.

Meticulous technique, hemostasis, and gentle handling of tissues contribute to a lower infection rate. Tissues should be kept warm and moist, especially during long procedures. Evidence suggests that wound infections might be prevented if hypothermia does not occur. While warming techniques have improved, they have not completely obviated hypothermia. All exposed parts of the patient should be covered, and a warming pad should be placed before the procedure. Administered fluids and irrigation should be warmed.

Wounds requiring drainage are more likely to become infected (11 percent versus 5 percent in undrained wounds), but it cannot be concluded that the drains are responsible for the infection. The drains may have been used because of concern for adequate hemostasis.

Clinical Manifestations

Wound infections are classified as minor, e.g., purulent material around skin suture sites, or major, e.g., discrete collections of pus within the wound. Superficial infections are limited to the skin and subcutaneous tissue; deep infections involve areas of the wound below the fascia.

Clinical manifestations of a wound infection include *rubor*, *calor*, *tumor*, and *dolor*. The patient may have pain (*dolor*) that is unusually severe given the magnitude of the procedure or the length of time it remains after the procedure. The wound may be warm to the touch (*calor*), it may be swollen and edematous (*tumor*), and there may be surrounding redness and cellulitis (*rubor*). If drainage is not prompt, the resulting wound abscess may be accompanied by substantial cellulitis. Fever to 101–102°F (to 39°C) usually is present with some increase in pulse rate.

Wound infections usually are evident between the fifth and eighth postoperative days. When patients have received antibiotics, wound infections may become manifest weeks after an operative procedure. In cases of severe necrotizing fasciitis or clostridial myositis, manifestations may occur within 24 h.

Management

Management of wound infections depends on the extent of destruction and the type of wound infection. A simple collection of purulent material in the skin and subcutaneous tissue without major surrounding cellulitis is treated by opening the incision. The incision is opened enough to provide adequate drainage, and a wick is placed in the wound to prevent the skin and superficial subcutaneous tissue from closing before complete drainage.

Opening the wound alone is insufficient in severe clostridial myositis or necrotizing fasciitis with loss of viability of the fascia or muscle. Radical debridement is necessary to save the patient's life. Discharge of "dishwater pus" should alert the surgeon to the possibility of necrotizing fasciitis. A Gram stain will reveal a mixed flora of gram-negative rods and gram-positive cocci. Clostridial myositis is manifest by crepitus (gas in the tissues), which also may be present in necrotizing fasciitis, and vesicles on the skin. In both situations, the patient is more sick than expected with a simple wound infection. When wound infections are associated with surrounding cellulitis and edema, the use of antibiotics in addition to opening and debriding the wound is necessary. A Gram stain may identify some of the offending organisms. In the absence of specific information, the wound should be cultured and the patient placed on ampicillin (or clindamycin), gentamicin, and metronidazole or a combination of antibiotics that covers the organisms most likely to be present. If hemolytic streptococcus is the offending organism, penicillin should be administered for 1 week.

Diabetic patients are prone to Fournier's gangrene, which is a form of necrotizing fasciitis of the perineum or groin. These patients require wide debridement. Fournier's gangrene often is fatal, with mortality rates of 30 to 75 percent.

Wound Hematomas

Wound hematomas are caused by inadequate hemostasis. Anticoagulation, fibrinolysis, polycythemia vera, myeloproliferative disorders, and decreased or inadequate clotting factors all contribute to hematoma formation. Great care must be taken in closing incisions in a patient who is in shock or in induced hypotension, or wounds in which epinephrine has been used. Hematomas provide good culture media for bacteria that might contaminate the wound and a barrier to the apposition of the wound edges. When the hematoma resolves, an untoward cosmetic result may occur.

Wound hematomas are manifest by pain and swelling of the wound. Drainage, if present, usually is serosanguinous, indicating a collection of bloody material within the wound. In certain locations, such as the neck, a large hematoma can be life threatening, and the hematoma must be evacuated immediately.

When the hematoma is discovered early postoperatively, the patient should be returned to the operating room and, under sterile conditions, the wound opened, the hematoma evacuated, the responsible vessel found and ligated, and the wound closed primarily. If hematomas are discovered late, heat should be applied to the wound and the patient managed expectantly with the hope that the hematoma has not become contaminated. If the surgeon suspects the possibility of a postoperative hematoma because of predisposing factors or difficult hemostasis, a closed-suction drain should be used. The catheter should exit through a separate stab incision remote from the primary wound.

Wound Seromas

Seromas are lymph collections usually resulting from, and associated with, operations in which large areas of lymph-bearing tissues are transected. Examples include axillary dissection and groin dissection. Closed-suction drains are placed to prevent seromas and are kept in place until drainage stops, but drains are not foolproof. Seromas are fertile ground for bacteria to create wound infections. If seromas are discovered in the postoperative period, repeated aspiration is indicated, or, preferably, closed-suction drains are placed percutaneously until fluid no longer accumulates, and the skin flaps are allowed to adhere. Pressure dressing may be helpful in conjunction with aspiration and continuous closed-suction drainage.

Wound Dehiscence

Dehiscence is separation within the fascial layer, usually of the abdomen, whereas *evisceration* indicates extrusion of peritoneal contents through the fascial separation. The incidence of wound disruption is between 0.5 and 3.0 percent, averaging 2.6 percent when all abdominal procedures are considered. Dehiscence occurs in 1.3 percent of patients under 45 years of age, and 5.4 percent of patients over 45 years of age. Many patient characteristics may contribute to a fascial dehiscence, but generally dehiscence is caused by a technical factor. Patient characteristics that contribute to fascial dehiscence should sensitize the surgeon to take precautions using a mass closure or anterior fascial retention sutures, which may not prevent dehiscence but may prevent evisceration.

Malnutrition, hypoproteinemia, morbid obesity, malignancy with immunologic deficiency, uremia, diabetes (especially with poorly controlled blood glucose levels), coughing with increased abdominal pressure, and remote infection are contributory factors. Dehiscence occurs in over 5 percent of patients with cancer, compared with less than 2 percent in patients with benign conditions. Jaundice has been associated with an increased incidence of wound dehiscence, but it is unclear whether this is a direct effect of the bilirubin or is the effect of all of the deficiencies that are associated with end-stage liver disease. Preoperative biliary drainage does not result in a decrease in wound complications or dehiscence. Ascites increases the incidence of wound disruption.

Local factors increasing wound disruption include hemorrhage, infection, excessive suture material, and poor technique. Series are contradictory as to whether dehiscence is increased in vertical as compared with horizontal incisions, but all agree that when a stoma is brought out through an incision the incidence of disruption is increased.

A multicenter, randomized, prospective trial compared interrupted versus continuous polyglycolic acid suture closure of midline abdominal incisions. The overall dehiscence rate was 1.6 percent in the continuous suture group and 2.0 percent in the interrupted suture group. The dehiscence rate was significantly higher in the interrupted suture group when wounds were contaminated. Monofilament sutures have a lower incidence of disruption than braided sutures.

Vitamin C is essential for collagen synthesis and fibroblast formation, bacterial destruction, and superoxide production by neutrophils. Many elderly patients are subclinically scorbutic, a condition associated with an eightfold increase in the incidence of wound dehiscence. Zinc is a cofactor for various enzymatic and mitotic processes, particularly of the epithelial cells and the fibroblasts. Patients with diarrhea, hepatic insufficiency, or chronic stress may be subject to zinc deficiency that is associated with poor healing. Steroids,

administered topically or systemically, have a deleterious effect on wound healing, interfering with wound healing at every level, including inflammation, wound macrophage function, capillary proliferation, and fibroplasia. Vitamin A may counteract these effects.

Chemotherapeutic agents inhibit wound healing. In patients with neoplastic disease, chemotherapeutic agents should not be used during the first week and probably also during the second and third weeks after operation. In breast disease, the use of 5- fluorouracil in the first postoperative week results in an increased incidence of skin dehiscence and delayed healing. A trial of 5-fluorouracil in patients with colon cancer, administered during and after operation, however, did not have a deleterious effect on wound healing. Radiation is known to affect wound healing adversely because of a variety of factors, including interference with normal protein synthesis, mitosis, migration of inflammatory factors, and maturation of collagen. Radiation therapy also results in obliteration of the small vasculature and fibrosis, which contribute to the vicious cycle of local hypoperfusion and hypoxia.

Clinical Manifestations

Dehiscence without evisceration can be detected by the “classical” appearance of salmon- colored fluid draining from the wound, which occurs in about 85 percent of cases about the fourth or fifth postoperative day. With the appearance of such fluid, the patient should be returned to the operating room and the wound opened under sterile conditions. Dehiscences may become manifest when skin sutures are removed and evisceration of abdominal contents occurs. If evisceration occurs, moist sterile towels are applied to the extruded intestines or omentum and the patient returned to the operating room. Dehiscence presents late as an incisional hernia.

Treatment

Management depends on the patient's condition. In some circumstances, if there is no evisceration, it is preferable to treat the patient nonoperatively with a sterile occlusive wound dressing and binder, accepting a postoperative hernia. Repair of dehiscences varies but generally involves a mass closure. There is little reason to attempt to redo a closure in which sutures, appropriately placed 1 cm apart, have pulled through. Perioperative (preferably broad-spectrum) antibiotics should be given. The mortality associated with wound disruption has been reduced from over 30 percent to below 1 percent. The incidence of reported postoperative hernia is inaccurate but is probably over 30 percent.

COMPLICATIONS OF THE GENITOURINARY SYSTEM

Urinary Retention

Urinary retention frequently complicates operations in males, but rarely in females. While the incidence of urinary retention after major abdominal surgery ranges from 4 to 5 percent, the incidence after anorectal surgery may be greater than 50 percent. Voiding requires a coordinated interaction between release of the alpha-adrenergic receptors in the smooth muscle of the bladder neck and urethra and parasympathetic stimulation so the bladder can contract and the urine can exit. Stress, pain, spinal anesthesia, and various anorectal reflexes conspire to increase alpha- adrenergic stimulation, which prevents release of the musculature around the bladder neck. In one study using phenoxybenzamine, there were no episodes of urinary retention in 58 older male patients after inguinal herniorrhaphy, compared to a 26 percent incidence of retention in 44 control patients. A similar study using bethanechol to stimulate parasympathetic activity, however, showed no benefit. Use of prazosin hydrochloride in a randomized prospective trial of male patients undergoing surgery for hip and knee prostheses resulted in a significant decrease in the incidence of postoperative urinary retention: 21 percent in the treated group, compared to 59 percent in the control group.

Another influence on urinary retention, especially in older patients, is the volume of infused fluid. A minimal volume of fluid should be infused in situations in which blood loss is likely to be minimal and the patient is not dehydrated.

In urinary retention, patients experience urgency, discomfort, and fullness, and an enlarged bladder can be percussed above the pubic symphysis. Straight catheterization is undertaken initially and then usually a second time. The bladder should not be allowed to distend beyond 6 h, because this will impair the patient's ability to void when the catheter is removed. If catheterization is required more than twice, a Foley catheter is placed and left to drain for at least 2 to 7 days.

Acute Renal Failure

Etiology

Inadequate Resuscitation

In the presence of hypotension, catecholamine release and norepinephrine stimulation by the sympathetic nervous system decreases renal blood flow. In addition, the renin-angiotensin system is activated with shunting of blood away from the afferent arterioles, depriving the cortex and the tubules of blood flow. A basic dictum in the care of emergency patients is ensuring significant and adequate urine output before going to the operating room and, if there is time, resuscitating such patients. In the elderly, it should be remembered that it is not *rapid* rehydration or transfusion, but *overhydration* and *overtransfusion* that are responsible for congestive failure. If there is any question in the surgeon's mind about the patient's volume status before going to the operating room, a pulmonary artery catheter should be passed, the patient's fluid and volume status optimized, and the operation delayed so that urine output is adequate before general anesthesia is induced. In the presence of transfusion reaction, sepsis, myocardial dysfunction, or crush injury, diuresis should be established using mannitol and furosemide (the latter given only when volume status is restored). Bicarbonate should also be administered to alkalinize the urine. Other causes of acute renal failure are listed in Table [11-11](#).

Drug Toxicity

A second important cause of acute renal failure in the surgical setting is the use of nephrotoxic drugs, most commonly the aminoglycosides, vancomycin, amphotericin B, and occasionally high doses of penicillin G or sulfonamides. Pharmacokinetic monitoring is used to determine the minimal inhibitory concentration in the plasma required for the organism in question. Pharmacokinetic monitoring does not prevent and may not substantially decrease vestibular toxicity or the incidence of renal failure. Heroin, thiazides, lithium, and occasionally vitamin D also are responsible for renal dysfunction. Acetaminophen in high doses, in addition to causing hepatic failure, also may cause interstitial nephritis (Table [11-12](#)).

Pathophysiology of Renal Dysfunction

Prerenal Dysfunction

Prerenal dysfunction is characterized by a BUN-to-creatinine ratio of 20:1 or greater. This is commonly observed with dehydration, e.g., after a bowel preparation or fulminant diarrhea, or underresuscitation of patients before or after an operative procedure. If the patient with prerenal azotemia is taken to the operating room inadequately resuscitated, there is an increased likelihood of acute renal failure. The diagnosis is fairly clear when the patient is hypotensive, overtly hypovolemic with decreased skin turgor and has sunken eyes and flaccid skin.

Another type of prerenal azotemia is a complication of liver disease. The hepatorenal syndrome is divisible into two forms. The first is basically hypovolemic, and the second is a maldistribution of blood flow (Fig. [11-3](#), Table [11-13](#)). In Type I, there is a deficiency in intravascular volume secondary to the presence of a high-grade outflow block in the liver leading to production of ascites and an inability of the liver to return the lymph to the systemic circulation. Lymph leaks from the surface of the liver into the peritoneal cavity. The patients have a low cardiac output, a tendency to high peripheral resistance, a urinary sodium concentration below 10 mEq/dL, and they appear underperfused and vasoconstricted. The patients generally improve with ascites reinfusion, volume challenges, or a side-to-side type portal decompression. The Type II hepatorenal syndrome usually is associated with a failing liver, an elevated bilirubin level, and other stigmata of cirrhosis such as an elevated prothrombin time and a low serum albumin level. In contrast to the intravascular depletion seen in Type I, the intravascular volume may be adequate. Cardiac output is increased, and there is low peripheral resistance. Hepatorenal syndrome is a hemodynamic rather than a renal disease. Kidneys from patients with hepatorenal syndrome are normal. They can be transplanted with prompt function, which has been shown repeatedly. Recovery from hepatorenal syndrome is rare and depends on the recovery of the intrinsic liver disease.

Intrinsic Damage

Intrinsic damage in the surgical setting includes a series of merged forms of tubular damage, known as acute tubular necrosis. In addition to acute tubular necrosis, pigment nephropathy and drug nephrotoxicity are prominent. The most common cause of acute tubular necrosis in the surgical setting is diminished renal perfusion secondary to prolonged and sustained hypotension in the presence of sepsis, blood loss, hypovolemia, dehydration, or myocardial infarction. This results in prolonged ischemia of the renal parenchyma. The kidney attempts to maintain glomerular blood flow and filtration by afferent arteriolar dilatation and efferent arteriolar vasoconstriction. As hypotension continues, the renin-angiotensin system attempts to correct for the perceived hypoperfusion, increasing angiotensin. Release of norepinephrine by the sympathetic nervous system with angiotensin causes afferent arteriolar vasoconstriction. Tubular ischemia and hypoperfusion of the renal cortex result; if tubular hypoperfusion persists, acute tubular necrosis results. In the traumatic setting, the presence of myoglobin and, in the case of transfusion reaction, free hemoglobin also complicates the renal injury.

The use of large amounts of radiocontrast medium causes reversible renal failure that usually clears after 3 to 10 days if the patient is maintained in a state of adequate hydration. When added to renal hypoperfusion from other causes, however, the damage may be prolonged and irreversible. Another cause of acute tubular necrosis that is commonly observed in the surgical setting is the showering of atheromatous emboli during aortic vascular surgery. The application of a clamp close to the renal artery may break off plaque that showers the renal vasculature, resulting in multiple infarcts and obstruction of various arterioles.

Postrenal Failure

Postrenal failure is rare in the surgical setting but may result from ureteral clots or stones. Chronic postrenal obstruction may result from benign prostatic hypertrophy. Prolonged Foley catheter intubation can result in a posterior urethral stricture.

Prevention of Acute Renal Failure

For patients with chronic urinary tract infection, specific antibiotics selected on the basis of cultures and sensitivities should be administered to eradicate the infection. Patients with chronic obstruction from benign prostatic hypertrophy should be treated with preemptive balloon dilatation or transurethral resection. In patients with chronic renal impairment, saline loading and adequate diuresis are established before general anesthesia. A pulmonary artery catheter is placed after the patient is adequately hydrated and sodium-containing fluids administered until a urine output of 50 to 100 mL/h is obtained. A patient with inadequate urine output should not be subjected to general anesthesia, because acute tubular necrosis can result.

If an angiographic workup is necessary, it is best performed several days before the operation so that if renal failure complicates the angiogram the effect is minimized. Use of radiographic contrast agents has an incidence of renal failure of 1 to 10 percent. Injury is due to direct nephrotoxic effect and hypovolemia from osmotic diuresis. Prestudy hydration is partially protective. If possible, antibiotics that are not nephrotoxic should be used.

In low-flow states, mannitol, bicarbonate, and diuresis induced by furosemide should be used. Mannitol is known to increase renal cortical blood flow and produce an osmotic diuresis. It may protect the tubules by preventing precipitation of metabolites within them. Renal perfusion may be increased by a “renal dose dopamine” of 2 to 5 mg/kg/min. If there

is any question concerning the volume status, central venous pressure monitoring or, preferably, monitoring of left-sided filling pressure with a pulmonary artery catheter is appropriate.

Manifestations

Acute renal failure usually presents in the postoperative period as oliguria with a urine output of 0.4 to 0.5 mL/kg/h in an adult. Anuria, while uncommon, as a manifestation of acute tubular necrosis probably is the result of renal artery thrombosis or obstructive uropathy. The diagnosis of acute tubular necrosis is made by spot measurement of urinary sodium and potassium levels and osmolality. The fractional excretion of sodium (FE_{Na}) is calculated using levels of sodium and creatinine in the urine or plasma as follows:

$$FE_{Na} = U_{Na} \times V / P_{Na} \times (U_{Cr} \times V / P_{Cr}) \times 100\%$$

where:

U_{Na} = urinary sodium

P_{Na} = plasma sodium

U_{Cr} = urinary creatinine

P_{Cr} = plasma creatinine

V = urine volume in mL

An FE_{Na} of greater than 1 indicates intrinsic renal damage. Measurement of urinary specific gravity may be helpful in the absence of pigment or contrast; urinary specific gravity in acute tubular necrosis is approximately 1.010. A urinary sodium level of less than 10 mEq/L indicates a prerenal cause or intrinsic liver disease. BUN and creatinine levels also are elevated. In the case of renal failure, tubular cells and casts, red blood cells, and white blood cells on urinalysis are diagnostic. The urinary sodium level usually is greater than 40 mEq/L, the potassium level is about the same figure, osmolality is approximately between 300 and 350, and the fractional secretion of sodium is greater than 3. The BUN-to-creatinine ratio is less than 20. The following formula may be used to calculate the creatinine clearance: C_{cr} = V × U_{Cr}/P_{Cr}, where V = volume, U_{Cr} = urinary creatinine, and P_{Cr} = plasma creatinine.

Management

The management of renal failure can be divided into two periods, the first when the diagnosis is uncertain, and the second when the diagnosis has been made. If the patient is oliguric and thought to be hypovolemic, a volume challenge is in order. If the initial volume challenge results in elevation of jugular neck veins, skin turgor, and eyeball depth, a pulmonary artery catheter should be placed to measure the patient's true volume status. Once adequate volume status has been established, furosemide (20 to 40 mg) should be given to improve urine output. Mannitol can be given to increase renal cortical blood flow, provided the patient is not volume overresuscitated. Ethacrynic acid or a furosemide drip can be used to prevent renal failure and perhaps “wash out” those tubules in which urine flow can be reestablished. “Renal dose dopamine” also may be used in conjunction with diuretics; bicarbonate is given in an attempt to slightly alkalinize the patient. Any nephrotoxic drugs (see Table [11-12](#)) should be stopped immediately.

Management of Established Renal Failure

When the patient has been diagnosed with renal failure, there are a number of goals:

- (1) Avoid overhydration, which results in congestive heart failure and the need for hemofiltration or dialysis.
- (2) Avoid dialysis if possible.
- (3) Avoid toxic ionic damage, such as hyperkalemia.
- (4) Provide nutritional support.

The most immediate threat in many patients with acute renal failure is hyperkalemia. Serum electrolyte levels should be monitored frequently. When serum potassium level reaches 5.5 mEq/ dL, there is need for concern, because rapid rises to the point where ECG changes occur can happen very quickly. ECG changes include peaked T waves with progression to a sinus bradycardia and a sine wave rhythm, with hypotension and death. Deaths from hyperkalemia are avoidable. The emergency treatment of hyperkalemia includes the infusion of calcium, hypertonic dextrose solution, and insulin. Thereafter, sodium polystyrene sulfonate (Kayexalate) 5 g as enema or by mouth should be administered. Infusion of hypertonic dextrose solution and essential amino acids in the form of total parenteral nutrition will lower the serum potassium level. Administration of fluids should be limited to below measured and calculated losses.

Maintenance of nutrition is an essential part of the treatment of patients with acute tubular necrosis. Nutrition can be administered enterally or parenterally. Increased survival in patients with acute tubular necrosis has been demonstrated only with intravenous nutritional support. An intravenous “Giordano-Giovannetti diet” of essential amino acids and hypertonic dextrose solution, with a minimum of fat, decreases mortality in patients with acute tubular necrosis. The most significant improvement is seen in the more severely affected patients with complications and in those with oliguric renal failure requiring dialysis. There is some, but not statistically significant, improvement in patients with non-oliguric renal failure. Nutritional support may delay the need for dialysis or cellular filtration because it minimizes the free water produced by muscular breakdown. It is not clear whether semiessential amino acids, such as arginine and histidine, should be included in this essential amino acid diet. While it is true that these amino acids may become deficient with chronic dialysis, plasma amino acid patterns in patients with acute renal failure do not reveal any evidence of amino acid deficiency.

Dialysis is undertaken in patients with acute renal failure for critical ionic excesses, volume overload, or a BUN concentration higher than 80 to 100 mg/dL. Volume overload may be prevented by continuous hemofiltration. Under most circumstances, hemodialysis is the procedure of choice, but chronic peritoneal dialysis may be successful after the retroperitoneum is sealed.

Dialysis is not innocuous; it has an annual mortality of 5 to 10 percent. The hypotension that usually is seen at the termination of the dialysis run, when patients are dehydrated, can be injurious to the kidneys. There is an advantage to delaying dialysis until patients are hemodynamically stable. Once patients are on chronic dialysis, essential and nonessential amino acids are given by TPN or, preferably, enterally.

RESPIRATORY COMPLICATIONS

Pathophysiology

Respiratory complications are among the most common complications of surgery and the most lethal, responsible for 5 to 35 percent of postoperative deaths. Upper abdominal and thoracic incisions result in a significant decrease in vital capacity and functional residual capacity, most prominently in the first 24 h after operation. After upper abdominal surgery, vital capacity may be reduced by as much as 50 to 60 percent, while functional residual capacity is reduced by approximately 30 percent. The cause of these changes is multifactorial. Postoperative pain alters the mechanics of respiration; upper abdominal and thoracic incisions have the greatest impact. Narcotic analgesia carries it own inherent risks because it eliminates sighing and promotes atelectasis. Even if pain is eliminated by epidural anesthesia, there remains a demonstrable decrease in vital capacity and functional residual capacity, an observation that has led some to suggest diaphragmatic dysfunction. Diaphragmatic inhibition can result from inhibitory reflexes arising from sympathetic, vagal, or abdominal receptors.

Closing volume (lung volume at which airway closure is first detectable) decreases in the postoperative period. Loss of functional residual capacity, elimination of sighing (by narcotic analgesics), and the change to a postoperative breathing pattern of small frequent breaths combine to lower the end-tidal point to a level that falls below closing volume. This results in a rapid loss of alveolar volume and subsequent alveolar collapse. This predilection is accentuated in smokers, who characteristically have a higher closing volume as a result of their diseased airways.

In addition to altered mechanics and decreased closing volume, other physiologic causes that contribute to respiratory insufficiency include diffusion defects, abnormalities in the ventilation-perfusion ratio, reduction in cardiac output with concomitant persistent shunt, alterations in the hemoglobin level and persistent shunt, and shunting that is anatomic or related to atelectasis.

Measurements of ventilation and oxygenation have been applied to assess the pathophysiologic events. Ventilatory mechanics are evaluated by measuring the ventilatory rate, the vital capacity (VC), total volume (VT), and dead space (VD). VD/VT, which also is influenced by cardiac output, is used to assess CO₂ elimination. Compliance is a measurement of the distensibility of the lung.

The partial pressure of CO₂ in arterial blood (Pa_{CO2}) can be considered as a reciprocal function of ventilation and is normally 40 mmHg. The adequacy of intrapulmonary blood-gas exchange is determined by measuring the Pa_{CO2} and Pa_{O2} in relation to the fraction of inspired oxygen (FI_{O2}). One method of estimating the efficacy of oxygen exchange in the lung is to measure the alveolar-arterial oxygen tension difference [(A - a)D_{O2}]. Factors influencing the (A - a)D_{O2} include the degree of mismatching of ventilation to perfusion, shunts around the lung, the difference between the arterial and mixed venous oxygen content, the mixed venous oxygen content (which may reflect oxygen consumption), the cardiac output, the inspired oxygen concentration (FI_{O2}), the position of the oxygen hemoglobin dissociation curve, and the position of the Pa_{O2} on the curve. A nomogram can be used to define the calculated amount of blood shunted around the lung as a fraction of the total cardiac output (*Q* /*Q*_T) based on the measurement of Pa_{O2} and pulmonary alveolar oxygen tension (PA_{O2}) (Fig. 11-4). Small changes in the *Q* /*Q*_T are more readily detected when the patient is breathing 100 percent oxygen for 20 to 30 min. This may result in absorption atelectasis, which increases *Q* /*Q*_T. Therefore, “shunt fraction” usually is measured at the inspired O₂ concentration (FI_{O2}) required to maintain an adequate arterial P_{O2} (60 to 70 mmHg or greater). The ratio of arterial P_{O2} to alveolar P_{O2} is independent of inspired FI_{O2}. Determinations are affected by alterations in the cardiac output and pH.

Predisposing Factors

A number of recognized risk factors predispose the patient to the development of pulmonary complications. Some of these risk factors cannot be appreciably altered (age, cardiac disease, or surgical site), but many risk factors may be favorably influenced during the preoperative period.

Smoking

The surgeon should distinguish chronic parenchymal lung changes from the short-term, potentially reversible effects. Patients who smoke are encouraged to stop before their operation. A patient must abstain from smoking for at least 8 weeks to achieve any demonstrable benefit. Patients who have stopped smoking for more than 8 weeks before surgery have as low a risk of pulmonary complications as patients who have never smoked. Less than 20 percent of patients who are requested to stop smoking preoperatively actually are able to do so.

Age

It is difficult to determine the risk attributable to age independent of the associated changes that accompany the aging process. Advanced age alone should not govern the decision for or against surgery. Overall poor physical status, rather than advanced age, is the most important risk factor in determining perioperative respiratory complications.

Obesity

The predilection for development of postoperative respiratory complications in obese patients is related to the underlying pulmonary dysfunction characteristic of this patient population. Patients who exceed their ideal body weight by more than 30 percent have a demonstrable decline in their functional residual capacity, which is the result of a reduction in chest wall compliance. This intrinsic loss of functional residual capacity, when combined with supine positioning, renders the obese patient particularly prone to hypoventilation and hypoxemia in the postoperative period.

Chronic Obstructive Pulmonary Disease (COPD)

Patients with preexisting pulmonary disease are at increased risk of postoperative complications. Pulmonary function tests provide a means to quantify objectively existing pulmonary abnormalities. Specific findings on pulmonary function tests that should alert the surgeon include abnormalities in FEV_1 , forced vital capacity (FVC), peak expiratory flow rate, and FEV_1/FVC ratio (see Table [11-2](#)).

In patients with recognized COPD, the surgeon must plan to improve the likelihood of a successful outcome. This includes careful planning of the incision, use of local anesthetic agents, careful closure of the wound, placement of gastrostomy tube, and postoperative pulmonary toilet techniques. A transverse incision is preferable to a vertical incision. Use of local anesthetics such as lidocaine or bupivacaine decreases intraoperative and postoperative incisional pain. In selected patients it is possible to place a small catheter for continuous infusion of lidocaine in the intramuscular layers of the incision to render the incision anesthetic. This simple maneuver may significantly improve a patient's pulmonary function. Wound closure in the patient with COPD should emphasize a secure closure with minimal postoperative pain, such as a subcuticular closure. Pain control, such as the use of an epidural catheter, is a significant improvement. In the absence of a postoperative epidural catheter, patient-controlled analgesic devices result in fewer respiratory complications. A gastrostomy (rather than a nasogastric) tube is useful in patients with COPD, as multiple studies have shown that a nasogastric tube in the postoperative period is the one clinical factor associated with a statistically increased incidence of respiratory complications.

Cardiac Disease

The most obvious example of cardiac disease complicating postoperative respiratory function is congestive heart failure with accompanying pulmonary edema. Patients with jugular venous distention, a third heart sound, or previous history of pulmonary edema are most likely to develop this complication. Patients with significant cardiac dysfunction and surgical patients who receive large amounts of fluids during their resuscitation should be observed closely for the development of pulmonary edema in the postoperative period.

Atelectasis

Atelectasis, the collapse of alveoli with ongoing perfusion, is the result of changes in the normal dynamics of pulmonary function resulting from anesthesia, diaphragmatic dysfunction, postoperative incisional pain, and patient positioning. Emphasis should be on prevention rather than treatment. Atelectasis results in a perceptible increase in shunt fraction. Once collapsed, the alveolus is difficult to rerecruit. This may be due to the loss of surfactant but also is a physical property because of the surface tension considerations of the pulmonary parenchyma. Secretions may accumulate in the collapsed alveolus, with potential bacterial overgrowth.

Lung inflation in the postoperative period prevents and reverses atelectasis. The most frequently used techniques include coughing and deep breathing (CDB), chest percussion and postural drainage (CPPD), incentive spirometry, intermittent positive pressure breathing (IPPB), and continuous positive airway pressure (CPAP). Each technique employs a different method to achieve the same goal of lung inflation. Various studies support the use of any of these techniques as superior to no postoperative therapy. The techniques of coughing, deep breathing, and incentive spirometry are inexpensive and not labor-intensive. CPPD, IPPB, and CPAP require significant labor, machinery, or a combination of both. Patients who are at low risk for the development of respiratory complications, such as those undergoing a lower abdominal incision or extremity surgery, probably do not require specific therapeutic interventions. For patients who have moderate risk of developing perioperative respiratory complications, coughing and deep breathing or the use of incentive spirometry usually will suffice.

Three groups of medications have been applied to the prophylaxis and therapy of atelectasis: (1) expectorants, to provide more liquid and less viscous secretions, (2) detergents and mucolytic solutions, to alter the surface tension of secretions and render their elimination more likely, and (3) bronchodilators, used primarily by inhalation, to increase the size of the tracheobronchial tree and eliminate bronchospasm. The mucolytic agents, such as Mucomist or Alevaire, are indicated because inhaled air with a relative humidity lower than 70 percent inhibits ciliary activity and tends to desiccate secretions.

Pneumonitis

Pneumonitis is a nosocomial infection seen with increasing incidence on surgical services. While pneumonitis is the third most common nosocomial infection (after wound and urinary tract infections), it is associated with the highest morbidity and mortality. The organisms involved are powerful pathogens and include *Pseudomonas*, *Serratia*, *Klebsiella*, *Proteus*, *Enterobacter*, and *Streptococcus*. There is an emerging predominance of gram-negative organisms, particularly in patients in intensive care units. This may be a reflection of the widespread use of H_2 -receptor antagonist therapy. Use of these agents results in breakdown of the acid barrier, allowing overgrowth and colonization of the stomach by intestinal flora. This situation is further aggravated by the placement of a nasogastric tube and the use of supine positioning in the postoperative period. Fungal pneumonia is uncommon, but with the increasing use of ever broader spectrum antibiotic regimens, the emergence of this pathogen in the future is likely.

Clinical Manifestations

Patients with pneumonia manifest fever, productive cough, dyspnea, pleuritic chest pain, and a purulent sputum. Bloodstained sputum is rare but may be exacerbated by repeated suctioning attempts. Moderate hypoxemia is common, but severe hypoxemia is unusual unless the pneumonia is severe and widespread. The presence of hypotension usually indicates a gram-negative pneumonia. Auscultation reveals bronchial breathing, areas of dullness to percussion, and the presence of rales.

Management

Management depends on correctly identifying the responsible organism. The proper technique for obtaining an adequate respiratory specimen for the diagnosis of pneumonia is controversial. In the intubated patient, there is increasing emphasis on obtaining a bronchoscopic-guided alveolar lavage specimen or the use of a protected-brush catheter technique. Multiple studies in patients who are intubated have demonstrated that cultures obtained via routine endotracheal suctioning have little predictive benefit in correctly identifying the pathogen responsible for nosocomial pneumonia. Given the increasing incidence of gram-negative nosocomial infections in the intensive care setting, antibiotic therapy with an aminoglycoside and an antipseudomonal penicillin should be initiated when the diagnosis is made. Antibiotic usage must be evaluated with culture results and modified if necessary.

Aspiration

Aspiration can be a cataclysmic, instantaneously lethal event associated with large amounts of particulate-laden acid contents unless expeditiously managed by suction, lavage of the respiratory tree, protection of the airway with an endotracheal tube, and continued pulmonary toilet until all particulate matter has been removed from the tracheobronchial tree. The use of steroids to ameliorate the progression of the clinical syndrome of aspiration has been demonstrated to be effective only when they have been given before the aspiration event. The most likely setting of massive aspiration is during the emergency induction of anesthesia, particularly in patients with gastroesophageal reflux or a hiatal hernia. It is commonly assumed that a nasogastric tube prevents aspiration, but this is not always the case. If a nasogastric tube is improperly positioned or maintained, its presence may facilitate rather than prevent aspiration.

Clinical Manifestations

The clinical sequelae of aspiration usually are not subtle, with the presence of gastric contents in the mouth followed by wheezing, hypoxia, bronchorrhea, and cyanosis. In the conscious patient, cough productive of particulate matter may be present. In the unconscious patient, aspiration may present as a major airway obstruction. Suctioning reveals gastric aspirate in the oropharynx and trachea. If aspiration is untreated, or if it is significant in volume, the results resemble that of a pulmonary burn, with edema, wheezing, cyanosis, and tachycardia. Chest radiographs demonstrate progression of local damage and infiltration. Acute respiratory failure results. In over 50 percent of patients who suffer aspiration, the initial chemical pneumonitis results in bacterial colonization with subsequent development of pneumonia.

Management

The only effective treatment of aspiration is prevention by emptying the stomach and neutralization of gastric contents. Evidence has suggested that neutralization strategies that use histamine H_2 -receptor antagonists or proton pump inhibitors elevate gastric pH to a point at which bacterial overgrowth may occur. Treatment of the early phase of aspiration includes removal of debris and lavage of the upper airway. Endotracheal intubation usually is necessary to initiate treatment and to complete clearance of the tracheobronchial tree. Bronchoscopic clearance of the airway is useful to clear all particulate debris. Bronchodilating agents may help in relieving obstruction in those patients with audible wheezing from reactive airway response. Positive-pressure ventilation (PPV) and *positive end-expiratory pressure (PEEP)* are often necessary, and a pulmonary artery catheter is useful in determining volume status.

Pulmonary Edema

Pulmonary edema results when pulmonary-capillary hydrostatic pressure exceeds plasma oncotic pressure. The result of this imbalance is fluid transudation into the alveolus. The most common causes of pulmonary edema in the surgical patient are fluid overload or myocardial insufficiency secondary to myocardial infarction/ischemia. Additional causes of pulmonary edema include sepsis, valvular dysfunction, neurogenic stimulation, and hepatic failure.

Increased capillary permeability also can result in a transudation of fluid into the alveolus. Common causes of increased capillary permeability include sepsis, *acute respiratory distress syndrome (ARDS)*, and acute pancreatitis.

Clinical Manifestations

There are two peak phases of occurrence of pulmonary edema in the surgical patient. The first occurs during resuscitation. If resuscitation is too aggressive or overreplaces intravascular volume, pulmonary edema may result. The second peak incidence of pulmonary edema is in the postoperative period when fluid mobilization occurs. If the patient's cardiovascular and renal systems do not maintain an adequate pace to off- load the mobilized fluid, pulmonary edema may result, particularly in the elderly patient. When volume status is in doubt, placement of a pulmonary artery catheter is valuable because it allows for adequate estimation of vascular volume and of end points of resuscitation, and it aids in the phase of fluid mobilization.

The patient with pulmonary edema manifests dyspnea at rest, tachypnea, and air hunger. In the elderly or immunocompromised patient, changes in mental status, including lethargy and disorientation, may occur. Wheezing and signs of bronchospasm may be audible. In addition, rales that often are up to the clavicles on posterior auscultation, distended neck veins, cyanosis, and peripheral pitting edema may be present. Chest radiographs may reveal progression of pulmonary edema, vascular prominence, septal lines (Kerley's B lines), and peribronchial and perivascular cuffing.

Management

Management depends on the inciting cause. For patients with a volume overload, simple therapy including oxygen and digitalization can significantly improve the clinical condition. In most instances, the placement of a pulmonary artery catheter significantly aids in the diagnosis and management. The initial findings reveal an elevated pulmonary artery occlusion pressure in the range of 18 to 25 mmHg. Cardiac index is normally decreased, with a concomitant elevation of peripheral vascular resistance. The observation of a low or normal pulmonary artery occlusion pressure (8 to 15 mmHg) suggests an alternative cause, such as acute respiratory distress syndrome or sepsis. An ECG should be obtained to evaluate the presence of pump failure as a result of myocardial infarction. In the setting of an abnormal ECG, cardiac enzymes should be drawn for confirmation by evaluation of MB fraction.

Additional therapeutic measures include oxygen and positioning the patient in the upright position. Diuretics, such as furosemide, are used to promote an off-loading of volume. Intravenous nitroglycerin can promote an increase in venous capacitance and subsequently decrease preload. Low-dose dopamine (2 to 5 mg/kg/min) may also be used to promote diuresis. Inotropic agents, such as dobutamine or amrinone, may improve cardiac output. If these maneuvers fail to produce a sufficient cardiac index, the use of afterload reduction agents, such as sodium nitroprusside, should be considered. Pulmonary embolism is another potential source of pulmonary edema. In this setting, a ventilation-perfusion scan may not be accurate, and if there is sufficient suspicion for the presence of pulmonary embolism, an angiogram must be performed. (See Chap. 21.)

Fat Embolism Syndrome

Fat embolism is an extremely common pathologic finding after trauma. In an autopsy series of 300 accident victims, the incidence rate ranged from 80 to 100 percent, with the higher incidence occurring in patients who survived for 12 h or more after injury. In view of the high volume of fat in long bones, it is not surprising that fat embolism is common after bony trauma. The incidence of fat embolism ranged from 26 percent in patients with a single fracture to 44 percent in patients with multiple fractures.

The fat embolism syndrome of pulmonary dysfunction, coagulopathy, and neurologic disturbances associated with increased circulating fat globules is uncommon. Chan and associates, in a prospective series of 80 patients with tibial and femoral fractures, reported an incidence of 8.75 percent. Studying a larger series of 172 patients, ten Duis and associates identified the fat embolism syndrome in only 3.5 percent of patients. The fat embolism syndrome most commonly follows orthopaedic injuries but also has been reported to occur after prosthetic joint replacement, closed-chest cardiac massage, blast concussion, liver trauma, burns, extracorporeal circulation, rapid high- altitude decompression, bone marrow transplantation, and liposuction. The syndrome has also been reported after acute hemorrhagic pancreatitis and carbon tetrachloride poisoning.

Pathophysiology

Long-bone fractures in animal models suggest that the cause of fat embolism syndrome is the release of marrow fat into the circulation with subsequent lodging in the lungs. The classic presentation of this syndrome is in the patient with multiple long-bone fractures after a traumatic event; it becomes more prominent if the patient has subsequent reaming or rodding of these fractures in the early posttrauma phase. The release of marrow substance and fat from the damaged marrow allows the intravascular passage of these substances to the pulmonary-capillary bed. Fat emboli recovered from the lungs of patients who died with fat embolism presented a lipid profile similar to that of fat in bone marrow. Posttraumatic lipemia, the coalescence of chylomicra, and other metabolic derangements are also considered causes of the fat embolism syndrome.

Large emboli may cause mechanical obstruction of the major and secondary pulmonary vessels. Modalities that increase the activity of lipoprotein lipase, releasing free fatty acids, which are thought to be a toxin, cause a more explosive and damaging syndrome. The pathophysiologic agent is probably a substance such as oleic acid, which has served as the classic model for the production of acute respiratory distress syndrome in the laboratory. Other researchers have suggested that once fat emboli are lodged in the pulmonary vasculature, they are coated with platelets that subsequently lyse, releasing kinin-like substances and vasoactive substances, such as serotonin.

It is more difficult to explain the findings such as petechia and lesions in the brain in these patients given the filtering function of the pulmonary- capillary bed. Other explanations to reconcile these findings include coalescence of the chylomicra that give rise to the fat embolism syndrome, or bypassing the pulmonary filter in precapillary shunts that have opened because of increased pulmonary artery pressure. Since the histologic features of the cerebral lesions consist of petechial hemorrhages of the cortical white matter, brainstem, and spinal cord, these will be self-limited and are reversible provided the patient can be supported during this period.

Clinical Manifestations

Up to 75 percent of patients with fat embolism syndrome manifest some degree of respiratory insufficiency. This usually occurs soon after the injury but occasionally as long as 48 to 72 h afterward. Chest radiographic findings include characteristic bilateral alveolar infiltrates. The syndrome may evolve into acute respiratory distress syndrome, and a minority of patients require intubation and respiratory support. Central nervous system (CNS) involvement occurs in the majority (as many as 86 percent) of these patients but does not develop in the absence of pulmonary abnormalities. Neurologic impairment may precede the pulmonary findings by 6 to 12 h. The most common neurologic presenting symptoms include confusion and disorientation with eventual progression to coma. Hypoxemia may exacerbate the confusion and disorientation, but oxygen does not reverse these CNS findings. Focal neurologic findings are uncommon. The characteristic petechial rash occurs in the axillae, neck, and skin folds. This same rash also may occur in the oral mucosa or conjunctiva. The rash tends to be transient, is usually present between 12 to 24 h, and rapidly disappears. Fever and tachycardia are common, with the fever out of proportion to injury.

The examination of urine for the presence of fat globules is not specific. A great number of patients have fat globules present in their urine after a traumatic event, but only a few of these patients develop the fat embolism syndrome. Associated clinical findings in fat embolism syndrome include an unexplained drop in hematocrit level, thrombocytopenia, hypocalcemia, and hypoalbuminemia. Hypoxemia is noted on arterial blood gases analysis. Serum lipase levels may be elevated in patients with the fat embolism syndrome. This finding might be a result of the escape of lipoprotein lipase and does not represent pancreatic lipase.

Management

Patients with long-bone fractures tend to be most prone to the development of fat embolism syndrome. Any patient with hypoxemia, fever, unexplained confusion, or tachypnea should have the fracture immobilized. Multiple studies have demonstrated that early surgical immobilization and fixation significantly decrease the incidence of pulmonary complications of the fat embolism syndrome in the posttrauma population. Adequate fluid resuscitation, transfusion, and, according to some, administration of TPN are associated with a decreased incidence of both encephalopathy and fat embolism syndrome. Encephalopathy is treated by oxygenation and supportive measures. There is not much support for the routine use of steroids in this disease. Heparin has been reported by some to increase lipase activity and thereby increase the clearance of fat. Other investigators have questioned the use of heparin because lipase actually may increase the toxicity of this syndrome by releasing oleic acid.

Other suggested therapeutic modalities include the use of low-molecular- weight dextran (40,000 MW), which is thought to act by decreasing blood viscosity, reducing platelet adhesion, providing partial reversal of thrombocytopenia, and reducing red cell aggregation. The use of ethyl alcohol has been advocated because of its ability to reduce lipase activity and decrease the production of free fatty acids, but this is unproved. Supportive modalities include ventilatory support as necessary and adequate fluid resuscitation. Death as a result of the fat embolism syndrome in the properly supported patient occurs infrequently.

Acute Respiratory Distress Syndrome (ARDS)

By definition, acute respiratory failure is the clinical situation in which the patient is incapable of maintaining adequate oxygenation, adequate ventilation, adequate tissue delivery, or some combination of these defects. The cause of this failure may be single and identifiable, such as pneumonia, decompensation of preexisting chronic obstructive pulmonary disease, or traumatic flail chest. Alternatively, it may represent the end point of a poorly understood pathway with a common final denominator of lung damage and subsequent decompensation of oxygenation and ventilation. An example of this is acute respiratory distress syndrome.

Ashbaugh and Petty are credited with the first description of ARDS. They described a syndrome characterized by atelectasis, reduced pulmonary compliance, and refractory hypoxemia. Estimates have suggested that the incidence of ARDS is about 100,000 cases per year. In 1992 a joint American-European conference on ARDS was convened with the goal of developing a working definition of ARDS. As part of this process, “*adult respiratory distress syndrome*” was changed to “*acute respiratory distress syndrome*” in recognition of the fact that pediatric patients also develop this disease process. The most widely accepted definition of ARDS is the syndrome that includes (1) lung injury, acute in nature, (2) bilateral infiltrates on frontal chest radiograph, (3) Pa O₂/FIO₂ less than 200, and (4) pulmonary-capillary wedge pressure less than 19 mmHg with no evidence of congestive heart failure. These same findings associated with a Pa O₂/FIO₂ less than 300 are defined as “acute lung injury” (ALI).

Etiology and Pathophysiology

Attention has focused on the concept of an abnormal cytokine response to injury. What is not established is whether the extent or components of this response are abnormal. Important components of this abnormal response include some activation of the complement cascade, activation of the thromboxane-leukotriene pathway, disorders in nitric oxide production, degranulation of neutrophils, and production of increased permeability factors by macrophages. All of these various factors have been implicated in the resultant transudation of fluid and reactive materials that are the hallmark of the alveolar flooding that characterizes ARDS.

The signature of ARDS is a ventilation/perfusion mismatch. Alveolar hypoventilation occurs because of collapse, expansion of the alveolar membrane, or presence of exudate and fluid within the alveolar spaces that are still available for perfusion. This “alveolar block,” when combined with changes in perfusion relationships, results in the increased hypoxemia manifested as a shunt as well as increased dead space. Shunt fraction, a reflection of the magnitude of ventilation/perfusion abnormality, is calculated by the formula Cc - Ca/Cc - Cv, where Cc is pulmonary capillary O₂ content, Ca is arterial O₂ content, and Cv is mixed venous O₂ content. Under normal circumstances, less than 5 percent of the blood flow across the pulmonary bed is a shunt fraction. In severe cases of ARDS, this shunt fraction can exceed 40 percent.

The classic radiographic appearance of “whiteout” on chest radiograph implies a uniform lung injury in ARDS. Computed tomography (CT) scans have demonstrated that this concept is misleading. CT scans of a patient with significant ARDS characteristically demonstrate regional changes in lung function and normal areas of lung interspersed with markedly diseased portions of lung. Dependent lung regions suffer the most significant consolidation, while the nondependent regions remain air-filled. The suggestion is that in some areas regional lung compliance is not reduced during ARDS but that lung volume is drastically diminished.

A newly described concept of ventilator lung injury has been termed “volutrauma.” Volutrauma is different from barotrauma and refers to the maldistribution of inspired tidal volume secondary to positive-pressure ventilation and the heterogeneous nature of lung injury in ARDS. Barotrauma is simply extra alveolar air. Volutrauma does not result in dissection of air from the alveolus but rather is characterized by direct tissue damage at the alveolar/capillary interface.

The mechanism of injury in volutrauma is thought to be overdistention or “stretching” of the alveolus beyond its normal maximum. This overdistention results in capillary fracture and parenchymal inflammation. This initial overstretching may be the primary insult, which results in a cascade of injury leading to increased capillary permeability and culminating in hypoxemia. The traditional therapy for ARDS may lead to a clinical syndrome that actually resembles ARDS. The heterogeneous nature of ARDS promotes the maldistribution of delivered tidal volume such that the majority of ventilation and pressure is transmitted to “normal” alveolar units rather than to the diseased areas of the lung. Based on the relationship of normal lung volume and transpulmonary pressure, the maximum alveolar pressure should remain less than 35 cmH₂O.

Management

Current mechanical ventilation strategies have emphasized the need to reduce volutrauma. These include: (1) early use of PEEP adjusted to the inflection point, (2) pressure-limited ventilation with plateau pressures less than 35 cmH₂O, (3) permissive hypercapnia (P_{CO₂} up to 90 mmHg, pH > 7.20, tidal volume 4–8 mL/kg), and (4) use of inhalational nitric oxide.

PEEP has remained the mainstay of treatment of ARDS for the past 30 years. End points based on Pa_{O₂}/Sa_{O₂}, intrapulmonary shunt, pulmonary compliance, oxygen delivery, and functional residual capacity have been suggested as defining “optimum PEEP.” PEEP is important not only as a method of recruiting collapsed alveolar units but also as a means of attenuating lung injury associated with positive-pressure ventilation. A pressure-volume curve is constructed to determine the point at which the pressure-volume relationship of the lung results in a rapid rise in the slope of this curve. PEEP is then set at a level equal to or just above the inflection point. Early in ARDS this value typically is 10 to 15 cmH₂O, with values of up to 20 cmH₂O necessary in patients with stiff chest-wall compliance. This use of PEEP may prevent loss of functional residual capacity and prevent alveolar collapse at end-expiration.

Positive-pressure ventilation is delivered predominantly as a volume- controlled breath. In this instance, the desired tidal volume is set on the ventilator, and that volume is delivered regardless of other variables. Peak airway pressures vary inversely with pulmonary impedance during volume- controlled ventilation. Monitoring of the peak airway pressure is used to evaluate the progression of ARDS. Pressure-controlled ventilation has been reported to improve oxygenation while maintaining lower peak airway pressures. Pressure-controlled ventilation is characterized by a square waveform of pressure and a decelerating waveform of flow. The rapid rise in flow reduces patient/ventilator dyssynchrony and reduces the work of breathing. The rapid rise in pressure allows peak airway pressure to be applied throughout inspiration, increasing mean airway pressure, enhancing ventilation-perfusion relationships, and allowing for improved distribution of inspired gases. The cumulative effect of these changes results in a more favorable ventilatory pattern while reducing volutrauma. Current practice is to use pressure- controlled ventilation in patients requiring PEEP greater than 10 cmH₂O, patients with elevated airway pressures (peak inspiratory pressure greater than 45 cmH₂O), and patients with hypoxemia despite adequate PEEP.

Permissive hypercapnia is not a therapeutic maneuver but rather a strategy of ventilatory support. Use of permissive hypercapnia includes the appropriate application of PEEP and the implementation of pressure- controlled ventilation, but the typical end points for acid-base balance are drastically altered in this strategy. Permissive hypercapnia relies on intentional hypoventilation and hypercapnia in an attempt to limit potentially detrimental effects of increased peak airway pressures. Permissive hypercapnia also limits the number of breaths necessary per minute, which reduces the risk of barotrauma and volutrauma. Appropriate initiation of permissive hypercapnia allows a gradual increase in Pa_{CO₂} over a period of 10 days as renal compensation maintains pH above 7.20. Permissive hypercapnia is a strategy that must be adopted early in the treatment of ARDS. It has little value when applied late in the course of ARDS, when volutrauma already has occurred. The lower level of acceptable pH is open to debate. Reports of pH as low as 7.0 with no adverse cardiovascular or neurologic effects have been published. Well- controlled prospective randomized trials comparing the morbidity, mortality, ventilator days, and length of stay in the intensive care unit for patients managed with permissive hypercapnia versus traditional strategies in ARDS have provided suggestive, but not unequivocal, evidence of its benefits.

In the weaning process, the first priority is to reduce the FI_{O₂} to less than 0.5 to avoid maintaining the collapse of the alveoli. Next, the number of intermittent mandatory ventilation (IMV) breaths should be decreased to a level that permits a normal pH and a Pa_{CO₂} of 35 to 45 mmHg at a respiratory rate of less than 30 per min. This is continued until only two mechanical breaths per minute are required. PEEP is lowered in increments of 2 to 3 cmH₂O/min while monitoring the Pa_{O₂}. When adequate oxygenation is maintained with a PEEP of 5 cmH₂O, the IMV is 0, the continuous positive airway pressure (CPAP) is 5cmH₂O, and the criteria of the second column of Table 11-14 are met, the patient generally no longer requires ventilatory support.

Newer Modalities

Inhalation of nitric oxide at small doses (0.1 to 20 parts per million) has reduced pulmonary hypertension and improved oxygenation in a variety of patients. Initial experience with the use of nitric oxide in ARDS has suggested a response rate of 60 to 70 percent. Partial liquid ventilation (PLV) or perfluorocarbon-assisted gas exchange (PAGE) has been applied. Perflubron, a perfluorocarbon with twice the density of saline that has an oxygen-carrying capacity of 50 mL/100 mL of perflubron, has been used. PAGE refers to the combination of PLV and traditional gaseous ventilation. The functional residual capacity is filled with perflubron (30 mL/kg), and mechanical ventilation is initiated. This strategy is sometimes referred to as liquid PEEP. Early limited studies with PLV have shown: (1) limited distribution of gas to the dependent regions of the lung, (2) more uniform ventilation with PLV, (3) improved gas exchange, and (4) reduced histologic evidence of lung injury with PLV. Because perflubron is immiscible with most substances, it facilitates removal of cellular debris and mucus.

CARDIAC COMPLICATIONS

Myocardial Infarction

Perioperative myocardial infarction (MI) probably is the leading cause of death in the elderly after noncardiac surgery. Mortality from perioperative MI ranges from 54 percent to 89 percent; 80 percent of the deaths occur within 48 h of operation. This is surprising, considering that the mortality rate for acute MI without shock, unassociated with operation, is approximately 12 percent.

The presence of coronary artery disease increases the incidence of perioperative MI from the control level of 0.1 to 0.7 percent to 1.1 percent after operation. In patients over 40 years of age, with or without coronary artery disease, the infarction rate is 1.8 percent. In patients with previous MI, the reinfarction rate ranges from 5 to 8 percent. The most important variable in a patient with a previous MI is the time that has elapsed since that MI. For patients who are operated on within 3 months the reinfarction rate is 27 percent, between 3 to 6 months the rate is 11 percent, and after 6 months the reinfarction rate stabilizes at 5 percent.

Identification of the Patient at Risk

The most widely used criteria and computation for the multifactorial index score to estimate cardiac risk in noncardiac surgery is that originally published by Goldman (see Table 11-1). Indications on physical examination include jugular-venous distention and an S3 gallop, more than 5/min premature ventricular beats, rhythm other than sinus, age over 70 years, and transmural or subendocardial infarction in the previous 6 months. Emergency operation, intrathoracic, intraperitoneal, or aortic sites of surgery, evidence for important valvular aortic stenosis, and “poor general medical condition” also are indicators. On the basis of the computational score, patients with more than 25 points (Class IV, the highest-risk class) have a mortality rate from cardiac causes of 56 percent and a morbidity rate of 22 percent with life-threatening, nonfatal cardiac complications.

The history is important in evaluating the risk of myocardial infarction. A history of dyspnea on exertion, orthopnea, paroxysmal nocturnal dyspnea, peripheral edema, and particularly angina pectoris (especially at rest) should lead the surgeon to obtain a more detailed history and a cardiac evaluation. A preoperative ejection fraction of less than 0.35 determined by radionuclide imaging and ventriculography is associated with a 75 to 85 percent incidence of perioperative MI, compared to a 20 percent incidence in patients with an ejection fraction greater than 0.35.

Clinical Manifestations

Most cases of perioperative MI occur on the operative day or during the first 3 postoperative days. Although infarction has been associated with all anesthetics, the incidence is higher after general anesthesia for abdominal or pelvic surgical treatment. The most important precipitating factor is shock, either during the operation or in the early postoperative phase. The more prolonged the shock, the greater the risk of coronary thrombosis and myocardial ischemia. The ECG may show ST depression and T-wave flattening with the loss of as little as 500 mL of blood in patients with previous coronary occlusion.

Diagnosis can be difficult because chest pain often is absent or obscured by narcotics. Chest pain occurs as a primary manifestation in only 27 percent of patients, which is less than the 97 percent generally reported in patients in whom a coronary occlusion is not related to surgery. It is appropriate to consider routinely monitoring patients with previous infarction in an intensive care unit. The sudden appearance of shock, dyspnea, cyanosis, tachycardia, arrhythmia, or congestive failure should be an indication of the diagnosis. Dyspnea, cyanosis, and arterial hypotension require a differential diagnosis between cardiac and respiratory problems. The ECG may provide the diagnosis with a characteristic infarction pattern, but this is not an unequivocal finding. In older patients, ST-segment and T- wave changes may be associated with myocardial ischemia, and the same changes may be observed with postoperative shock. A study of arterial gases may provide a differential diagnosis of respiratory problems. Left ventricular failure with pulmonary edema generally is not accompanied by carbon dioxide retention, and, in contrast to airway obstruction and alveolar hypoventilation, there usually is a reduction in arterial CO₂ tension and respiratory alkalosis when cardiac failure accompanies MI. The CPK- MB isoenzyme is the most precise method for detection of myocardial necrosis after operation. If MI is suspected, serial studies, including ECG and measurements of aspartate transaminase (SGOT) and CPK-MB, should be done daily. Isotope scanning of the myocardium using technetium pyrophosphate may detect a recent acute infarction.

Management

Preoperative preparation of patients with signs of cardiac insufficiency should include digitalization for patients with enlarged hearts or histories of previous cardiac failure. Routine digitalization is not indicated. Anemia, if present, requires treatment, and attention should be directed toward the regulation of fluid and electrolyte balance and hypovolemia. Patients on propranolol should continue to receive the drug until the morning of the operation. Operation is contraindicated for a period of at least 3 months, and preferably 6 months, after myocardial ischemia or infarction, except in an emergency. During the operation, a broad array of factors precipitating MI should be avoided. These include hypoxia,

hypotension, hemorrhage, dehydration, electrolyte disturbance, and arrhythmias. The regulation of blood pressure during anesthesia probably is the most important measure in the prevention of myocardial ischemia and infarction. When blood pressure falls significantly in the absence of blood loss, the prompt correction of hypoxia by adequate ventilation with oxygen and the administration of vasopressors is indicated. Digitalization may be required when shock is combined with heart failure. The administration of blood or fluid is indicated to maintain blood volume.

Treatment of MI itself consists of relief of pain and anxiety using morphine and sedation. Relief of hypoxia is accomplished with 33 to 50 percent oxygen delivered via a mask or nasal catheter. Suctioning of the tracheobronchial tree may be required to clear obstructing secretions. Critically ill patients are best managed in an intensive care setting with invasive monitoring using arterial lines and a pulmonary artery catheter. Shock is treated by vasopressor agents. Promptness in instituting vasopressor therapy increases the chance of its effectiveness. Rapid digitalization is applicable in the treatment of shock when the myocardial insufficiency may be responsible for the severe hypotension. Digitalization also is indicated for the treatment of heart failure, which is a frequent manifestation of postoperative MI. In addition to digitalization, parenteral diuretic therapy may be used in the treatment of cardiac failure. Some have advocated the use of anticoagulant therapy after the danger of excessive bleeding from an operative site has passed. When acute MI is detected, early emergency cardiac catheterization, angioplasty, or stenting may reverse an evolving MI. Rarely, emergency coronary artery bypass surgery may be indicated.

Arrhythmias

If arrhythmia is defined as a sequence of abnormal beats sustained for more than 30 s, an overall incidence of 73 percent of intraoperative cardiac arrhythmias has been reported in one series and 62 percent in another. The incidence was higher in intubated patients and in patients who had undergone neurosurgical and thoracic procedures. Twenty-one percent were ventricular in origin. In another study, 84 percent of patients experienced significant arrhythmias. The majority were associated with the intubation and extubation phase of anesthesia. Of these, 43 percent were ventricular arrhythmias. Once anesthesia is over, the incidence of arrhythmias in noncardiac surgery is approximately 2.4 percent, and the majority of patients are asymptomatic.

Sinus tachycardia, not an arrhythmia, is by far the most common disturbance in rhythm, followed by premature ventricular contractions and sinoatrial arrhythmia, which can be a normal variant. Bradycardia and trigeminy also have been reported in a significant number of patients.

With cardiac procedures, the incidence of arrhythmias is approximately 50 percent, while thoracic procedures have an incidence of 20 to 30 percent. Pneumonectomy has been associated with an arrhythmia rate of 30 to 40 percent, leading to the practice of digitalizing the patient before pneumonectomy. In one report, the incidence of supraventricular tachycardias, including supraventricular tachycardia, atrial flutter, and fibrillation, after pneumonectomy was 22 percent. Of these, more than half occurred within the first 72 h of operation, and one-third occurred on the first postoperative day.

Etiology

The causes of arrhythmia include intrinsic cardiac disease, perioperative release of catecholamines because of stress or pain, and organ manipulation that stimulates a reflex response.

Electrolyte Abnormalities and Metabolic Disturbances

Hypokalemia is associated with paroxysmal atrial contractions and ventricular contractions or, in rare cases, ventricular fibrillation. Hyperkalemia, especially when severe, results in conduction abnormalities and, when potassium levels reach 6.5 to 7.5 mEq/L, conduction defects resulting in asystole. Hypocalcemia increases the QT interval and can cause ventricular arrhythmias. Hypercalcemia causes bradycardia and heart block. Hypomagnesemia can cause ischemic heart disease and increase ventricular irritability, and it may make the heart prone to the dysrhythmic effects of hypoxia and other disturbances.

Cardiac Medications

Digitalis may predispose surgical patients to serious dysrhythmias, and digitalis toxicity can result in supraventricular-atrial flutter with varying block, premature ventricular contractions, and ventricular tachycardia or fibrillation. Some of the common antihypertensive medications or beta blockers such as propranolol, reserpine, methyldopa, and quinidine may cause sinus bradycardia or induction block.

Anesthetic Agents

While no longer used, cyclopropane and, less commonly, halothane may result in ventricular dysrhythmias. Parasympathetic stimulation, such as the use of neostigmine, physostigmine, and the relaxant succinylcholine, may cause bradycardia. Atropine is administered simultaneously to prevent bradycardia.

Other Factors

Hypercapnia suppresses the intrinsic pacemaker activity and, at times, sinoatrial node function. It also may lead to ectopic pacemaker or aberrant reentry mechanisms, resulting in atrial or ventricular arrhythmias. Thyrotoxicosis commonly results in atrial dysrhythmias, the most common of which is atrial fibrillation. Pheochromocytoma, with its elevated plasma catecholamines, also can cause a variety of arrhythmias.

Management of Preexisting Arrhythmias

Approximately 33 percent of patients with preoperative arrhythmias undergoing peripheral vascular surgery develop cardiac complications, as compared with 9 percent of the patients with normal sinus rhythm. When the preoperative evaluation is accomplished with ECG, rhythm strip, and, if necessary, 24-h continuous Holter monitoring, the preexisting arrhythmia should be controlled. Digoxin (Lanoxin) and other cardiac glycosides can be used to control the ventricular rate in patients with supraventricular tachycardia, taking care not to push the patient into digitalis toxicity.

Multifocal premature ventricular contractions or runs of ventricular tachycardia should be treated aggressively, especially in elderly patients with coronary artery disease or congestive heart failure. Reversible causes, such as electrolyte disturbances, drug toxicity, hypoxia, etc., should be controlled. Intravenous lidocaine should be given. For significant conduction defects, cardiac pacing should be considered. In third-degree AV block, secondary to Mobitz II block and sick sinus syndrome, a permanent pacer should be implanted before elective surgery. Percutaneous pacing might be indicated in patients with new-onset ischemia if operation cannot be delayed. In patients with ischemic development of left fascicular block, hemodynamic monitoring with a pulmonary artery catheter is essential.

Management of New-onset Arrhythmias

A 12-lead ECG and a rhythm strip should be obtained. Carotid sinus pressure may slow the rhythm sufficiently so that the critical issue, which is whether P waves are present and, if so, what their morphology is, can be detected. The presence of P waves implies a supraventricular origin. The variable morphology of the P waves may suggest an ectopic focus, a multifocal atrial tachycardia or supraventricular tachycardia. Absence of P waves indicates atrial fibrillation. A small sine wave usually is evident. The QRS response indicates whether there is an abnormality in ventricular function. A narrow and normal-appearing QRS complex suggests that the arrhythmia is supraventricular. Wide, aberrant QRS responses indicate a ventricular origin or a supraventricular rhythm with aberrant conduction. AV conduction block rhythms also result in a wide QRS complex.

The management of new-onset arrhythmias essentially consists of determining (1) the hemodynamic stability of the patient, (2) the ventricular rate (rapid or slow), (3) the site of origin (atrial or ventricular) of the arrhythmia, (4) the need for cardioversion, and (5) identifying and correcting the underlying cause of the arrhythmia. In hypotensive patients with an acute tachyarrhythmia, cardioversion should be performed. The initial impulse should be at 100 joules (J), and if the arrhythmia is not eliminated, cardioversion should be repeated with the voltage rapidly increased to 360 J.

The rapidity with which therapy is attempted, and what is attempted, is directly dependent on the ventricular rate, which is critical. The ventricular rate may be too fast or too slow. In bradycardia, defined as a heart rate of less than 60 beats/min, chest pain, dyspnea, altered mental status, evidence of myocardial ischemia, or ventricular asystole demands rapid treatment with 0.5 mg of atropine intravenously. If bradycardia persists, atropine is given every 5 min to the maximum dose of 0.04 mg/kg. If a single dose of atropine does not result in the rapid reversal of bradycardia, a transcutaneous or, preferably, a temporary transvenous pacemaker should be inserted promptly. If the ventricular response rate is too great (a tachyarrhythmia is defined as a heart rate of more than 100 beats/min), hemodynamic instability may result, especially when ventricular response rates of 130 to 140 beats/min are achieved.

Sinus Tachycardia

Sinus tachycardia is not a dysrhythmia but may be mistaken for one and can cause alarm when the heart rate is markedly elevated. Sinus tachycardia usually suggests sympathetic stimulation by a variety of conditions, including pain, hypovolemia, hypoxia, acidosis, sepsis, congestive heart failure, hypoperfusion, or hypercapnia. While thyrotoxicosis may cause sinus tachycardia, particularly in thyroid storm, an atrial dysrhythmia is more common. Treatment should include 2 to 4 liters of oxygen/min and is directed at identification and correction of the underlying cause.

Paroxysmal Supraventricular Tachycardia

The paroxysmal supraventricular tachycardia dysrhythmias, which are reentry in nature, have a ventricular rate between 150 and 250 beats/min. They may be caused by hypoxia, myocardial ischemia or infarction, congestive heart failure, or thyrotoxicosis. With a rapid ventricular response, synchronized direct-current countershock is indicated. Carotid sinus massage or Valsalva maneuver usually is unsuccessful. Primary treatment is adenosine 6 mg intravenously, which may be repeated after 1 to 2 min with 12 mg; this should be infused slowly to prevent the hypotension that may be associated with the medication. If the arrhythmia persists, verapamil 2 to 5 mg intravenously is given as an initial dose, with a second dose given after 15 to 30 min. Verapamil should be used with caution in the elderly because hypotension may result in patients with poor ventricular function.

Atrial Fibrillation

Atrial fibrillation can be troublesome because the lack of the “atrial kick” can result in a 10 to 15 percent diminution in cardiac output, which may be poorly tolerated in certain elderly patients. Causes include thyrotoxicosis, valvular heart disease, hypertension, coronary artery disease, pulmonary embolism, and myocardial ischemia. Atrial fibrillation is very common after pulmonary resection, particularly pneumonectomy. Direct-current cardioversion is indicated when patients are hemodynamically unstable. However, if the atrial fibrillation is long-standing, there may be a clot in the atrial appendage, and the patient should undergo anticoagulation therapy. Verapamil is the drug of choice once the initial treatment controls cardiac rate, and digitalis glycoside is used to maintain control once the heart rate is decreased. When atrial fibrillation supervenes in a patient on digitalis glycosides, quinidine or procainamide usually is successful in converting fibrillation to a sinus rhythm. Part of the success of the conversion depends on the size of the atrium; if the atrium is more than 4 cm in diameter, atrial fibrillation may be refractory to conversion.

Sustained Supraventricular Tachycardias

Atrial tachycardias and atrial flutter, especially multifocal or ectopic atrial tachycardias with 2:1 or 3:1 block, may be a result of digitalis toxicity. Serum levels of potassium and digoxin should be obtained promptly, and digoxin discontinued temporarily. Potassium should be supplemented if serum level is low, and procainamide and quinidine might be necessary.

Atrial Flutter

The ventricular rate in atrial flutter is sustained and high, but it is regular. Ventricular response varies with the degree of AV conduction and usually determines the patient's hemodynamic stability. Atrial flutter can be associated with mitral or tricuspid valve disease, sustained pulmonary hypertension, or cor pulmonale. If the patient is hemodynamically unstable, synchronized direct-current countershock should be applied. Once the heart rate is controlled, digitalis is used to maintain it. In patients who have had previous digitalis therapy, quinidine or procainamide may be effective.

Ventricular Tachycardia

Ventricular tachycardias are the most dangerous arrhythmias. A sustained ventricular tachycardia cannot be tolerated. The underlying cause usually is intrinsic disease of the heart, including cardiomyopathy, coronary artery disease, or mitral valve prolapse. Hypoxia and cardiac glycoside toxicity are rare causes. The rhythm is variable, and it may be regular or irregular. The critical diagnostic feature is a wide, bizarre QRS complex with a rate varying between 100 and 225 beats/min. The patient may be pulseless, in which case a precordial thump and immediate defibrillation at 200 J should be performed while cardiopulmonary resuscitation is attempted. Treatment is the same as that of ventricular fibrillation. If a pulse is present, lidocaine 1 mg/kg is given, followed by a second dose of 0.5 mg/kg and a lidocaine drip 2 to 4 mg/ min. Bretylium may be used in refractory cases at a dose of 5 to 10 mg/kg and may be repeated every 15 min to a dose of 30 mg/kg.

Ventricular Fibrillation

Ventricular fibrillation usually is fatal. Its underlying causes include ischemia, hypoxia, digitalis toxicity, or hypokalemia. No P waves or normal QRS complexes are seen on the ECG but rather a fine or coarse fibrillation. There is no pulse or cardiac output. Cardiopulmonary resuscitation is initiated and direct-current countershock repeated up to three times. Epinephrine is given intravenously, and, if it is effective, cardiopulmonary resuscitation may maintain some aspect of the circulation. Lidocaine 1 mg/kg is given, followed by a second dose of 0.5 mg/kg and a continuous infusion of 2 to 4 mg/min. Bretylium may be given as well in a dose of 5 to 10 mg/kg and can be repeated at 15-min intervals to a maximum dose of 30 mg/kg. Administration of epinephrine may be repeated every 3 to 5 min, followed by defibrillation. The patient usually will become acidotic, and in addition to intubation and hypocapnia to control the acidosis, bicarbonate should be used. Arterial blood gases analysis should be used to guide the resuscitation.

Hypertension

Approximately 40 percent of the population, at one time or another, has a systolic blood pressure of more than 140/90 mmHg. There is controversy about whether this is dangerous preoperatively. Long-term hypertension causes arterial medial hypertrophy, accelerated atherosclerosis, increased left ventricular work, and cardiac hypertrophy. Untreated hypertension leads to diminished renal blood flow and premature renal failure. The pathophysiology of very wide swings in blood pressure, and particularly in peripheral vascular resistance in the postanesthetic period because of sympathetic stimulation as a result of pain, anesthesia, or hypovolemia, is the result of vascular smooth muscle hypertrophy and wall thickening. Diminution of intravascular diameter associated with hypertension exaggerates changes in vascular resistance brought about by the inevitable sympathetic discharge that occurs after an operation. The central nervous systems of the elderly are vulnerable under these circumstances because the autoregulation of blood flow is diminished with the thickening of the arterial media, resulting in an increased risk of ischemic strokes.

Preoperative Hypertension

It is not clear that patients with preoperative hypertension are at risk for increased cardiac morbidity and mortality. For those patients whose diastolic blood pressure is lower than 110 mmHg, it appears that there is no increased perioperative risk of MI or sudden cardiac death, although there may be a slight increase of intraoperative hypotension and myocardial ischemia. Preoperative hypertension that is untreated or poorly controlled does increase the risk of perioperative blood pressure lability, which may result in an increased incidence of stroke, transient neurologic event, arrhythmias, postoperative myocardial ischemia, and possibly postoperative renal failure. Whether *systolic* hypertension is a critical issue is unclear, but evidence suggests that it is not. In the event of previous hypertension with a *diastolic* pressure higher than 110 mmHg, new-onset hypertension, sudden increases in hypertension, or recent deterioration in critical end- organ status (renal failure, eye, or heart), elective operation should be postponed until the hypertension is controlled.

In patients with uncontrolled new-onset hypertension or poorly controlled preexistent hypertension, operation should be delayed until adequate blood pressure control is achieved. In patients with mild to moderate hypertension, the following factors call for delay of an operative procedure:

- (1) Electrocardiographic changes of myocardial ischemia or infarction
- (2) New-onset dysrhythmias
- (3) Emergence of left ventricular hypertrophy on ECG
- (4) New-onset or unstable angina pectoris
- (5) Congestive heart failure, whether established or new
- (6) A recent neurologic deficit
- (7) New onset of high-grade hypertensive retinopathy

Antihypertensive medications should be continued until the operation. Although beta blocking agents, which previously were used for the treatment of angina or hypertension, may complicate intra- and perioperative hemodynamic monitoring, abrupt stoppage of these medications has resulted in ventricular tachycardia, myocardial infarction, unstable angina, and sudden death. Preoperative discontinuation of quinidine can result in anxiety, tremors, diaphoresis, malaise, and tachycardia.

Postoperative Hypertension

Approximately 25 percent of patients with vascular disease and previously established hypertension who are operated on experience postoperative hypertension. Postoperative hypertension develops in over 50 percent of patients after abdominal aortic aneurysm and 20 percent of patients after carotid endarterectomy. This is important because systolic blood pressures higher than 200 mmHg result in bleeding from the suture line, hemorrhagic cerebral infarction, myocardial ischemia or infarction, and acute renal failure. If such hypertension complicates a carotid endarterectomy operation, the elevated blood pressure results in a statistically significant increase in neurologic deficits. Similarly, cardiac operations in which cardiopulmonary bypass is used are commonly followed by hypertensive episodes, which may jeopardize the integrity of vascular anastomoses. Patients with uncontrolled hypertension undergoing plastic surgical procedures experience an increased incidence of hematomas.

If hypertension develops during anesthetic induction or emergence, the adequacy of ventilation, hydration, and fluid status should be established. The anesthesiologist should be prepared to use sodium nitroprusside or nitroglycerin if the increase in pressure cannot be otherwise controlled. Nearly 80 percent of postoperative hypertensive episodes occur within the first 3 h of emergence from anesthesia. Obvious factors include the presence of an endotracheal tube in a patient who is awake and agitated, inadequate analgesia, acute bladder distention, and fluid overload. More subtle causes of postoperative hypertension include tracheal stimulation, hypothermia, hypercapnia, and hypoxemia. Diuretic therapy may be required to diminish the intravascular volume and hypoxemia. The efficacy and adequacy of ventilation and oxygenation should be evaluated. If the blood pressure is uncontrollable, sodium nitroprusside or labetalol is given. Effective management of postoperative hypertension includes intravenous or sublingual nitroglycerin or nifedipine. Other antihypertensive agents, such as methyl dopa or hydralazine, can be used in less emergent situations.

Hypertension occurring later in the postoperative course usually is related to hypervolemia secondary to fluid mobilization into the intravascular space, inadequate analgesia, or failure to resume previous antihypertensive medications. If fluid overload is mild, fluid restriction may be sufficient. Occasionally, diuretic therapy may be necessary. Sufficient doses of analgesics should be provided and preoperative medications resumed as soon as feasible.

HYPERCOAGULABLE STATES

Acquired Hypercoagulable States

Lupus Anticoagulant Factor (Anticardiolipin Syndrome)

Lupus anticoagulant factors are antibodies that interfere with the in vitro partial thromboplastin time (PTT) by prolonging phospholipid-dependent clotting factors and also interfere with heparin monitoring. The presence of these antibodies is associated with an increased risk of arterial and venous thrombosis. A positive history of thrombotic events is obtained in many patients. Several specific assays are available to detect the presence of lupus anticoagulants.

A problem with the treatment of acute venous thromboembolism is the distortion of the activated partial thromboplastin time (APTT). The patients normally do not require anticoagulation therapy. Patients undergoing major surgical procedures should receive prophylactic anticoagulation therapy and mechanical prophylaxis, such as sequential compression boots, against venous thromboembolism.

Heparin-Induced Thrombocytopenia

Heparin-induced thrombocytopenia is a form of consumptive platelet activation. It is idiosyncratic in that the effect of heparin is not dose- dependent, and very small quantities of heparin can elicit the syndrome. Porcine heparin may be less likely to induce thrombocytopenia. The mechanism is thought to be autoantibody formation directed toward heparin and platelet surface antigens. Platelets are then activated and the clot consumed.

Mild transient thrombocytopenia occurs 2 to 4 days after heparin exposure; it may occur earlier if the patient has been exposed to heparin in the past. A more severe syndrome includes marked severe hyperthrombocytopenia 6 to 12 days after heparin exposure and is associated with thrombosis. Arterial thrombosis is common and includes aortic and lower extremity vascular bypass grafts, femoral arteries, coronary arteries, and cerebral arteries. Thrombosis of the vena cava and the feeding branches of the iliofemoral and femoral veins also have been reported.

Mortality is significant, and phlegmasia cerulea dolens results in an amputation rate of up to 30 percent. While arterial thrombosis may be effectively managed by surgical thrombectomy, ultimately antithrombotic agents, such as dextran and warfarin, must be used. Venous thrombosis is treated by mechanical interruption of the vena cava, such as by Greenfield filter, and by warfarin therapy when platelet levels are adequate.

The diagnosis is based on clinical suspicion and elimination of other causes of thrombocytopenia. It is possible to perform platelet aggregation studies using donor platelets and the patient's serum, plasma, and in vitro heparin. Discontinuation of the heparin results in lower morbidity and mortality, especially if the syndrome is detected early.

Inherited Thrombotic Disorders

The surgeon occasionally is required to treat patients who have known inherited thrombotic disorders. More often, the patient gives a history of unexplained venous thromboses without a clear family history. It is useful to investigate these patients in order to prevent the occurrence of thrombosis after operation. Indications are spontaneous venous thrombosis, especially in unusual sites such as the mesenteric or the cerebral veins, a family history of thrombotic problems, and recurrent thrombosis with no apparent precipitating factor. Arterial thrombosis is notably absent, except as complicating mesenteric venous thrombosis where venous thrombosis could lead to bowel necrosis. Heparinization is indicated until the dose of warfarin sodium is adjusted. Prophylactic anticoagulation should occur before any type of surgical procedure.

Antithrombin-III Deficiency

Perhaps the most important inhibitor of coagulation is antithrombin III. In addition to acting on thrombin, antithrombin III inactivates factors Xa, IXa, XIa, plasmin, kallikrein, and factor XIIa. The physiologic range of antithrombin III is narrow. Its deficiency, an autosomal-dominant inherited trait, with resultant moderate decreases in concentration may have significant clinical impact in the form of thrombosis. Recurrent thrombosis occurs in approximately 60 percent of patients, and pulmonary embolism occurs in up to 40 percent. Antithrombin-III deficiency can be confused with several acquired conditions that include estrogen therapy, oral contraceptives, heparin, l-asparaginase, cirrhosis, nephrotic syndrome, and disseminated intravascular clotting. Treatment is with heparin. Patients undergoing operation should be given fresh-frozen plasma to raise their level of antithrombin III. Warfarin is effective.

Protein C Deficiency

Protein C is a vitamin K–dependent inhibitor of the procoagulant system. It forms in the liver and causes its inhibitory action by inactivating factors V and VII:C. Its inhibitory action is greatly facilitated by protein S (see below). Approximately 4 to 5 percent of patients under the age of 45 years with unexplained venous thrombosis have this diagnosis, but evidence suggests that it might be as important as antithrombin III deficiency. It is an autosomal-dominant gene disorder of which two types of deficiency are known: patients in whom there is an absence of protein (CRM-) and those who have a dysfunctional protein (CRM+). Thrombosis occurs when serum activity falls to less than 70 percent. Since the levels are affected by warfarin, anticoagulation therapy with warfarin is sufficient.

Protein S Deficiency

Protein S is a vitamin K–dependent protein that is produced by hepatocytes and megakaryocytes. Its primary role is to function as a cofactor in protein C's inhibitory actions on factors V and VIII. Protein C deficiency occurs as an autosomal-dominant disorder, but protein S deficiency can occur in the heterozygous and homozygous forms; the homozygous form generates symptoms earlier in life. As in protein C deficiency, there is a quantitative and qualitative subgrouping for protein S–deficient patients. Thrombosis occurs when serum activity falls below 60 percent of normal. Patients may acquire protein S deficiency in acute thrombotic disease and in disseminated intravascular clotting.

POSTOPERATIVE PAROTITIS

Postoperative parotitis is a serious complication that is associated with a high mortality related mainly to the primary disease. Studies indicate a recrudescence in occurrence, probably related to the increasing age of the surgical population. The right and left glands are involved equally, and the disease presents bilaterally in 10 to 15 percent of cases. Seventy-five percent of patients are 70 years of age or older, and the majority have associated diseases. Patients who undergo major abdominal operations or who have a fractured hip, debilitating disease, or severe injury are among the most commonly afflicted.

Causes include poor oral hygiene, dehydration, and the use of anticholinergic drugs. In one large series, one-third of the patients with acute suppurative parotitis had carcinoma, and one- half had preexisting major infection elsewhere in the body. In only one-third of the cases in this series did the acute suppurative process develop in the postoperative period.

The majority of infections are from staphylococci, and the pathogenesis is thought to be a transductal inoculation of the parotid gland. Poor oral hygiene and lack of oral intake to stimulate parotid secretions predisposes to bacterial invasion of Stensen's duct. The inflammatory lesions of early parotitis are confined to an accumulation of cells within the larger ducts. The parenchyma of the smaller ducts are initially spared, but once penetration of the parenchyma occurs, multiple abscesses form and later coalesce. If the process continues, the purulent material penetrates the capsule and invades the surrounding tissue along one of three routes: downward into the deep fascial planes of the neck, backward into the external auditory canal, or outward into the skin of the face.

Clinical Manifestations

The interval between operation and the onset of parotitis varies from a few hours to many weeks. The patient initially presents with pain in the parotid region that is usually unilateral but may become bilateral in a short time. Initially the gland is slightly swollen and exquisitely tender. Because of the septate anatomy of the gland, fluctuance rarely is demonstrable. The course of postoperative parotitis is rapid and fulminating with severe cellulitis on the affected side of the face and neck. The temperature and leukocyte count may be extremely high. Obstruction of the airway might necessitate tracheostomy, and the abscess can rupture into adjacent structures of the ear, mastoid process, pharynx, or anterior and posterior triangles of the neck. Parotitis is differentiated from benign postoperative swelling of the parotid glands that occurs more frequently in black patients and may be related to straining or to the administration of atropine or neuromuscular depolarizing drugs.

Management

Prophylaxis includes adequate hydration and good oral hygiene, which is aided by allowing the patient to take ice chips and hard candy to stimulate salivary flow. Prophylactic antibiotics have no apparent value.

When considering the diagnosis, pus should be expressed from Stensen's duct and culture and sensitivity tests performed. While awaiting results, a broad-spectrum antibiotic is begun that acts against staphylococci. In one series of 66 glands cultured, 64 contained staphylococci. Some cases were combined with streptococci, gram-negative bacilli, and pneumococci. If there is considerable pain and the disease is less than 48 h old, irradiation of the gland in small doses is indicated. Irradiation may provide symptomatic relief by reducing the secretions of the obstructed gland, but this type of therapy does not affect the course of the disease as much as antibiotics or surgical drainage.

Frequent observation of the patient is essential. If the disease persists or progresses, drainage should be considered as early as the third day. If there is moderate improvement, drainage may be delayed for a day or two, but it should not be delayed beyond the fifth day. An incision is made anterior to the ear, extending down to the angle of the mandible, and flaps are reflected to expose the gland. A hemostat is inserted through the capsule and opened in the direction of the course of the branches of the facial nerve, establishing multiple drainage sites; the wound is packed lightly open. Deferring drainage until fluctuation is apparent is unwise. Stimulation of the salivary flow by massage of the gland or other means is contraindicated once the inflammatory process is established.

Prognosis

In one series, the mortality rate approximated 20 percent, but this was frequently related to the patient's basic disease. Thirty-six percent of the patients who died demonstrated active parotitis. In 80 percent of patients treated with incision and drainage, the parotitis was palliated or cured.

COMPLICATIONS OF SURGERY OF THE GASTROINTESTINAL TRACT

Ileus and Partial Small-Bowel Obstruction

Ileus is defined as nonmechanical obstruction that prevents normal postoperative progression of the return of bowel function, food intake, and discharge. Some of the complications that are called ileus probably include partial mechanical obstructions that are difficult to differentiate from nonmechanical obstruction. Ileus is thought to arise from a neural inhibition that interferes with coordinated intrinsic bowel wall motor activity and effective propulsive peristalsis. It is believed that the small bowel normally does not manifest ileus postoperatively, because the small bowel continues to function throughout and after operation; tube feedings into the small bowel can begin almost immediately after operation. When inflammation is adjacent or there are several anastomoses in the small bowel, a 24-h ileus might be experienced. If adjacent inflammation continues, the ileus may be prolonged. Gastric ileus can persist from 24 to 48 h, and sometimes longer. Colonic ileus lasts from 3 to 5 days. The end of colonic ileus is signaled by the passage of flatus.

Ileus increases with increased manipulation, inflammation, peritonitis, and large amounts of blood left in the peritoneal cavity because of inadequate hemostasis or deficient coagulation factors. Blood in the retroperitoneum often produces ileus. Hypokalemia, hypocalcemia, hyponatremia, and hypomagnesemia prolong postoperative ileus. Opiates and phenothiazines, especially when given in large doses and in patients who are habituated, contribute to a delay in the resolution of ileus.

While in the United States water-soluble dyes such as diatrizoate (Hypaque) are used for radiologic examinations, enteroclysis with thin barium is used in other countries where there is less fear of barium accumulating in an intestine that may be obstructed. Failure to pass contrast past a fixed point is pathognomonic of intestinal obstruction.

If ileus persists postoperatively, a long tube (Kaslow or Miller-Abbott) may be effective in reversing intestinal obstruction without operation. The serum albumin should be measured, because patients who are very hypoalbuminemic can experience a prolonged ileus. Infusion of 12.5 g of albumin every 8 or 12 h to raise the serum albumin level above 3.0 mg/dL often results in the passage of flatus and a return of bowel function. The treatment of ileus is purely supportive. Rarely, if ileus is prolonged and it is not mechanical obstruction, TPN may be required to support nutrition until the ileus resolves.

Ileus and mechanical obstruction may be difficult to distinguish. Postoperative mechanical obstruction, particularly in carcinomatosis, can be difficult to distinguish from a prolonged ileus. Radiographic examinations, particularly enteroclysis, might be helpful. Long-tube decompression may be helpful for postoperative obstruction.

Anastomotic Leaks and Fistulas

General Considerations

There are a number of features that are essential for normal, complication- free healing of gastrointestinal tract anastomoses:

- (1) The two ends of bowel or the two hollow viscera being anastomosed should have adequate blood supply without ischemia and should bleed freely.
- (2) The orientation of the bowel and its mobilization should result in an anastomosis that lies well and is free of tension.
- (3) Meticulous technique should be used in anastomosing bowel. The sutures should be appropriate for the anastomosis —small, fine, nonabsorbable sutures for the serosa and a variety of sutures in the mucosa. A single-layer anastomosis with excellent apposition of the serosa can be expected to undergo normal healing.
- (4) Preparation of the large bowel, including antibiotics and especially a mechanical (cathartic) preparation, is critical.
- (5) Hemostasis should be meticulous, especially in and around the anastomosis. Hematomas at the suture line should be avoided.
- (6) No contaminated material, fibrin, etc. should be left in the immediate vicinity of the anastomosis.
- (7) If possible, the anastomosis should be reinforced with a serosal patch or omentum onlay.
- (8) Nutritional preparation of patients, when the operation is not an emergency, should be optimal. Patients who are at risk nutritionally for disrupted anastomoses include those with an albumin level in the hydrated state of less than 3.0 mg/dL and weight loss of more than 10 to 15 percent in the previous 3 to 4 months. Confirmatory findings include the inability to perform tasks previously done, anergy to injected skin test antigens, and a true transferrin level of less than 200 mg/dL. Hand dynamometry and other tests of physiologic capacity also are confirmatory.

Factors that increase the likelihood of anastomotic leakage include emergency procedures, poorly prepared patients, inadequately resuscitated patients, prolonged intraoperative hypotension, and hypothermia. Three etiologic factors are: (1) poor surgical technique; (2) distal obstruction; and (3) inadequate proximal decompression.

Duodenal Stump Blowout

Duodenal stump blowout remains a disastrous complication with a high mortality. A review of gastrectomies performed at the Mayo Clinic in 1956 revealed that 4.5 percent of patients subjected to the procedure for gastric ulcer had some evidence of leakage, and 5.6 percent of patients in whom the same procedure was performed for duodenal ulcer revealed similar evidence. In that study, drains were inserted into the stump region, and in many patients increased drainage represented the evidence of a leak. Edmunds and associates reported a 1.1 percent incidence of dehiscence of the stump with a mortality of 0.6 percent.

Duodenal stump leakage occurs most commonly after operation for a duodenal ulcer and frequently when gastrectomy is performed as an emergency procedure to stop hemorrhage. In a majority of cases, the leak is a result of technical error and failure of the suture line. A scarred and edematous duodenum, obstruction of the afferent loop, and local pancreatitis predispose to leakage. Complications of duodenal leakage include peritonitis, subhepatic abscess, pancreatitis, sepsis, establishment of an external fistula with fluid, and electrolyte abnormalities.

Specific measures can be taken to avoid this complication. With marked inflammatory disease in the duodenal region, vagotomy and gastroenterostomy represent safer procedures. When resection has been undertaken and duodenal closure is difficult, catheter duodenostomy may be used as an adjunct. Rodkey and Welch reported that in 51 cases with difficult duodenal stump closures in planned duodenostomy, there was one death, and in five patients drainage from the fistula continued for more than 48 h after the catheter was removed. As a compromise between primary closure and planned duodenostomy, some surgeons advise drainage of the right upper quadrant, with the drain placed in the region of the duodenal stump so that if perforation occurs, the contents discharge along the tract. This does not provide the safety factor of planned duodenostomy because the drain tract may wall off from the stump before the perforation becomes established.

Duodenal blowout is most likely to occur between the second and the seventh postoperative days, manifested by sudden pain, elevation in temperature and pulse rate, and general deterioration of the patient's condition. Adequate drainage must be instituted immediately, which is best accomplished by an incision below the right costal margin and insertion of a large sump catheter that is passed down to the duodenal stump area with constant suction applied. Attention must be directed toward fluid and electrolyte therapy, and TPN should be instituted. Fistula closure can be anticipated within 2 to 3 weeks.

Intestinal Leaks and Fistulas

Leakage of intestinal anastomoses usually is manifest by fever, leukocytosis, unexplained ileus in the absence of intestinal obstruction, and a complicated postoperative course. There may be localized swelling of the abdominal wall and point tenderness. CT scanning usually is diagnostic. Percutaneous drainage often is effective in reversing the sepsis.

If the leak is small or if previous drains were placed to the area of the anastomosis, the patient is not deteriorating, the sepsis is contained, and the patient is not jeopardized, nasogastric suction, broad-spectrum antibiotics, total parenteral nutrition, and adequate drainage may result in containing the leakage without further need for operative intervention. If the patient is in jeopardy, the sepsis is uncontrolled, and there is no effective drainage, the patient's abdomen should be reexplored. Under no circumstances should the anastomosis be resutured; it will almost always leak again for the same reason it leaked initially. The anastomosis must be completely resected and redone. If contamination is not massive, the blood supply is adequate, and the patient is not hypotensive, a redone anastomosis will most likely heal. If there is gross contamination and purulence, the patient is hypotensive, septic, or hemodynamically unstable, then separation of the two ends of the bowel and diversion should be performed. When the patient has stabilized and the area of sepsis is resolved, reanastomosis may be attempted.

A fistula represents an anastomotic leak that has developed a pathway to the skin. The typical presentation of a patient with an enterocutaneous fistula is characterized by fever, ileus, leukocytosis, malaise, and some manifestation of sepsis. On the fourth or fifth postoperative day, increased wound pain and redness lead to drainage of purulent material from the wound. This usually is followed within 24 h by leakage of intestinal contents through the wound. Postoperative fistulas usually result from operations involving inflammatory bowel disease, cancer, or lysis of adhesions.

The treatment plan for patients with enterocutaneous fistulas is to allow the fistula to close spontaneously; it is more likely to do so without operative intervention in the absence of distal obstruction. Operative intervention in patients with fistulas is dangerous; the mortality in patients with enterocutaneous fistulas from all causes is 10 to 20 percent.

Therapy of an Established Fistula

The treatment of patients with gastrointestinal-cutaneous fistulas can be divided into five phases: (1) stabilization, (2) identification and diagnosis, (3) decision, (4) operation, and (5) healing.

Stabilization

The typical presentation is a wound that developed an infection and was drained, subsequently followed by the appearance of enteral contents. Another situation occasionally occurs in which there is localized swelling and redness, an abscess from a disrupted anastomosis that points to the skin. Before drainage, a fistulogram should be obtained with water-soluble dye. Once the fistula is defined, rapid and complete resuscitation should take place. The patient has probably not eaten for about a week, there are 3–4 liters of electrolyte-rich solution in the gut, the serum albumin level is probably decreased, and the patient is most likely anemic because of sepsis. Resuscitation should take place using crystalloid, red blood cells when appropriate, and, if there is no capillary leakage syndrome, albumin to raise the albumin level to 3.0 mg/dL. The patient is largely given nothing by mouth, so that intestinal secretion is avoided. Hard candy can be used to stimulate salivary flow to avoid parotitis. Nasogastric tubes are unnecessary unless the patient has an obstruction, in which case the passage of a long Miller-Abbott or Kaslow tube in the early postoperative period should be considered. Antibiotics are not used unless the patient is septic. A sump-type drain is placed around the skin, and the skin and skin edges are protected with Stomadhesive and ion exchange paste to keep the pH of the skin acidic and prevent activation of the pancreatic enzymes that require a basic pH.

Nutritional support is delayed only if there is anticipated drainage of abscesses because hematogenous spread, while rare, may occur and result in seeding the TPN catheter. TPN is initiated with a 5 or 6 percent amino acid solution, 15 to 25 percent dextrose, and 20 percent of the caloric content as fat. It is started at 40 mL/h and advanced rapidly until the patient's metabolic needs are met. The adequacy of nutritional support can be monitored with short-turnover proteins, such as retinol-binding protein, thyroxin-binding prealbumin, and transferrin, or by indirect calorimetry.

Enteral nutrition also is effective, but in all series the rate of closure using enteral nutrition is decreased slightly when compared to parenteral nutrition. Enteral nutrition, while it temporarily increases the drainage from the fistula, results in increased hepatic protein synthesis, decreased contamination, decreased translocation, and other beneficial metabolic effects. It might not be possible to meet the patient's entire metabolic needs with enteral feedings. It is estimated that as little as 20 percent of the caloric needs supplied enterally provide all of the benefits usually associated with enteral feedings. A combination of enteral and parenteral feedings may be more effective than total parenteral nutrition alone.

If enteral nutrition is used, 4 feet (1.2 m) of small bowel must be present to absorb the feeding. If the material is given into the duodenum, or beyond the ligament of Treitz, which is best from the standpoint of safety in order to avoid aspiration, it should not be given at full strength but started at isosmolar or slightly below isosmolar levels and the rate increased to 125 mL/h. The osmolality is increased to the point where it may be 400 to 500 mOsm, but not greater. Many patients will not tolerate osmolalities greater than 300 to 400 mOsm when feedings are given into the small bowel. If the material must be given into the stomach, hypertonic material can be given initially. The osmolality may be increased, and when 600 mOsm is reached, the rate can then be increased. When hyperosmolar material is given into the stomach, peristalsis stops, and the stomach ceases emptying and secretes fluid until the material is diluted to isosmolar. Peristalsis then returns and the material is transferred in 2- to 4- mL aliquots every 30 seconds across the pylorus.

Identification and Diagnosis

After the patient is stabilized, the fistula controlled, and the skin protected, the characteristics of the fistula are defined in order to determine whether or not the fistula is likely to close spontaneously, or whether operation is required. The most important exercise in making that decision is a fistulogram. The prognostic indicators that should be obtained from the fistulogram include the degree of bowel continuity, the size and depth of the defect, whether there is distal obstruction, the nature of the bowel adjacent to the fistula, and whether there is a large abscess adjacent to the fistula (Fig. 11-5). The location of the fistula also is important. Ileal fistulas, gastric fistulas, and fistulas at the ligament of Treitz close less frequently than fistulas in the area of the jejunum, esophagus, pancreas, biliary tree, or colon. The fistula should be inspected to make certain that it is not a mature fistula, i.e., intestinal epithelium that has grown up to the skin. Fistulas that have matured as an enterostomy never close and always require operation to resect the area and

perform an anastomosis.

Spontaneous closure usually occurs within 5 weeks of adequate nutritional support, enteral or parenteral, in a patient who is sepsis-free. Duodenal and esophageal fistulas may close within 3 weeks. While remote sepsis (e.g., urosepsis, pneumonitis) may not interfere with spontaneous closure, local sepsis almost always prevents spontaneous closure. If the patient is systemically septic, spontaneous closure will not occur.

Decision

If spontaneous closure is to occur, the drainage of the fistula usually decreases. Somatostatin has been used to promote closure. Somatostatin will not close a fistula that is anatomically unfavorable for closure. Randomized prospective trials suggest that somatostatin is not effective in speeding up closure but might be useful in pancreatic and biliary fistulas as well as duodenal fistulas.

The critical decision in the management of these patients is *whether* operation is indicated, and if so, *when*. If spontaneous closure has not occurred during 5 weeks of sepsis-free parenteral nutrition and drainage has not decreased dramatically, it is unlikely that the fistula will close. The timing of operation is important. It is best performed when the skin around the fistula is in a good state and the patient has shown no recent improvement. This is a judgment that can be made readily with the use of laboratory data. If the short-turnover protein levels are increasing, the serum albumin concentration is approaching 3.0 g/dL, and the patient is maintaining the albumin level without infusions of exogenous albumin, operation can take place. A rise in the level of serum transferrin at the onset of the treatment or after 3 weeks is predictive of spontaneous closure. A rise in levels of transferrin, retinol-binding protein, and thyroxin-binding prealbumin predict survival, but failure to maintain or increase these is indicative of mortality.

If the patient is to be operated on, careful preparation should be undertaken. The skin should be washed with chlorhexidine gluconate (Hibiclens) for 2 or 3 days before surgery, and a sump should be positioned with Stomadhesive to protect the skin. Fistula or abscess drainage should be cultured and appropriate antibiotics covering aerobes and anaerobes given just before the start of the incision. If yeast is present, prophylaxis by fluconazole (Diflucan) is appropriate. Formal bowel preparation, neomycin enemas, and irrigation to the colon may be useful adjuncts.

Operation

A significant mortality exists in patients with enterocutaneous fistulas. Mortality is approximately 10 to 11 percent if patients are reoperated on during the first 10 days or after 4 months, but a mortality of 20 percent is expected if operation is performed between 10 days and 4 months after the fistula appears. All adhesions must be taken down and all abscesses drained. Resection and end-to-end anastomosis yields the lowest incidence of failure and the lowest incidence of complications. Bypasses or intentionally staged procedures lead to higher rates of complications and may result in total failure of closure of the fistula. The authors do not recommend using stapled anastomoses under these circumstances. An exception to the dictum that fistulas should be excised with end-to-end anastomoses is the duodenal fistula. If a duodenal fistula does not close, it is suggested that a vagotomy and gastrojejunostomy be performed, a feeding jejunostomy and a gastrostomy be placed, and the area of the fistula be drained, which usually closes.

Ancillary procedures at the time of resection and operation include the placement of gastrostomy and a feeding jejunostomy, which is essential. A No. 12 latex Robinson nephrostomy catheter is preferred, as it allows greater flexibility with respect to the tube feedings. The anastomoses should be protected where appropriate by omentum onlay. If the colonic anastomosis is unsatisfactory or suspect, it should be protected by colostomy or ileostomy.

A chronic pancreatic fistula that communicates with the duct may be dealt with by excising the fistula down to the pancreas, identifying the leak, and performing a distal pancreatectomy and splenectomy if the spleen cannot be spared. A Roux-en-Y anastomosis can be used to provide internal drainage for the pancreatic fistula. Alternatively, the fistula tract can be dissected free starting at the skin and anastomosed to a Roux-en-Y.

Healing

After the procedure has been performed successfully and there are no or few perioperative complications, feeding is delayed 7 to 10 days postoperatively, particularly if the anastomoses are tenuous. Many of these patients will have difficulty in eating because they have not eaten for some time. Some patients lack taste sensation; zinc sulfate or lactate given once a day often restores taste within 10 to 14 days. It may be necessary to allow the patient alcohol to induce eating, and families can bring home-cooked meals that may be more palatable than hospital food. After a prolonged period of time without food, most patients require cycling of their enteral or parenteral nutrition at night to enable them to eat during the day.

Colocutaneous Fistulas

Colocutaneous fistulas are generally the result of colonic anastomotic leaks or unrecognized trauma to the colon during operation. Creation of an anastomosis in the presence of infection or fecal contamination, such as for acute diverticulitis or traumatic colon injuries, increases the likelihood of fistula formation. In contrast to enterocutaneous fistulas, fluid and electrolyte abnormalities and skin digestion are rare, but infectious complications are significant, with abscess formation and wound infections.

Nutritional status is maintained by using low-residue elemental enteral diets. Localized infection is controlled with percutaneous drainage of intraabdominal abscesses and local care of wound infections. Antibiotics are used as indicated. Spontaneous closure of colocutaneous fistulas is very likely, with rates approaching 75 percent. The presence of sepsis involving the fistula, distal obstruction, anastomotic dehiscence, Crohn's disease, or carcinoma are associated with persistence of the fistula. Earlier operative intervention is indicated if peritonitis or septicemia is present. The lack of spontaneous closure by 5 weeks is an indication for surgical repair. Definitive operation involves resection of the fistula and affected colonic segment with primary anastomosis and temporary diversion of the fecal stream by colostomy if necessary. The success rate of surgical repair of colocutaneous fistulas is 70 to 80 percent.

Postgastrectomy Syndromes

Dumping

Dumping is the result of the loss of the pyloric valve that normally is the mechanism by which hyperosmolar material is prevented from entering the duodenum and small bowel. This function is lost as a consequence of pyloroplasty, pyloromyotomy, gastrojejunostomy, gastric resection, and Billroth I or Billroth II anastomosis after which hyperosmolar material enters the small bowel. This results in physiologic changes including the release of vasoactive substances, such as serotonin, bradykinin, substance P, and peptides including vasoactive intestinal peptide, possibly pancreatic polypeptide, insulin, glucagon, neurotensin, and enteroglucagon. There also is a loss of plasma volume as the small intestine secretes actively to dilute the hyperosmolar contents, resulting in a decrease in plasma volume and hence hypotension. Rapid absorption of glucose, secretion of insulin, and the rapid entry of glucose and potassium in the cell results in hypokalemia, which may also be responsible for some of the symptoms.

Symptoms of dumping consist of early postprandial bloating, borborygmus, cramps, sensation of light-headedness, palpitations, sweating, and hypotension. Most patients admit to dumping immediately after gastric procedures, but most usually improve so that at the end of 4 to 5 months there are few patients who remain seriously symptomatic. Separating solids and liquids (eating the solids at the meal and drinking the liquids afterward) diminishes the symptoms. Carbohydrate-rich foods are more likely to provoke dumping; these include milk shakes, cereal with significant sugar content, etc., all of which should be avoided. Acquired lactase deficiency also may be present. In time, most dumping subsides or is minimized as the patient adapts. In more severe cases, long-acting octreotide may oppose some of the actions of the released peptides and ameliorate the symptoms. If the patient remains severely symptomatic, the anastomosis may be converted from a Billroth II to a Billroth I. The duodenum seems more resistant to an osmotic load and, because it contains any number of polypeptides, may provide an antagonist to dumping. If this strategy fails, a 6-cm reverse loop of jejunum may slow the transit of hypertonic solution in the small bowel sufficiently to ameliorate the symptoms of dumping.

Postvagotomy Diarrhea

Most patients have increased bowel movements or softer stools after truncal vagotomy. Between 5 and 20 percent of patients have diarrhea that is troublesome, and 1 to 2 percent of patients with truncal vagotomy have severe, disabling diarrhea. The latter may take two forms: diarrhea that is present every day, or a poorly explained episodic diarrhea occurring about every 7 to 10 days and lasting about 3 days. This does not respond to opiates, and it subsides spontaneously.

Many factors have been blamed for postvagotomy diarrhea, including stasis and overgrowth of bacteria, malabsorption of fat, and increased and incoordinate bile flow into the small bowel. The cause is not clear but is ascribed to the denervation of the celiac plexus, resulting in a dysmotility or dysfunction of small bowel motility.

Treatment is difficult. Episodic postvagotomy diarrhea is almost impossible to correct. Neomycin or tetracycline to sterilize a blind loop syndrome and cholestyramine to bind bile salts have been tried with little success. In patients who are incapacitated, a 10-cm reversed jejunal loop 100 cm distal to the ligament of Treitz has been advocated. A difficulty with the reverse loop is that although it initially may be effective, it ultimately enlarges and elongates and becomes obstructive, requiring shortening. A shorter reverse loop might be more appropriate (Fig. 11-6).

Afferent Loop Syndrome

This complication is a mechanical problem peculiar to the gastroenterostomy reconstruction after gastrectomy. Optimal function of this arrangement requires free flow of material from the duodenum (afferent loop) to the intestine draining the gastric pouch (efferent limb). The afferent loop syndrome can be acute or chronic. In its classic form, the syndrome almost always occurs when the afferent loop is anastomosed to the greater curve after a Billroth II gastrectomy. Obstruction of the afferent loop can result from adhesions, kinking, intussusception, volvulus of the afferent loop, stomal ulcer, or obstruction of the efferent limb. In the chronic form, as the afferent loop becomes obstructed, bile, pancreatic juice, and duodenal secretions accumulating in the afferent loop are suddenly regurgitated into the stomach. In the acute form, as the secretion increases, hemorrhagic pancreatitis (a Pfeffer loop) or perforation can occur. The ligament of Treitz may be located too close to the midline, resulting in kinking when the afferent limb is anastomosed to the greater curve. In these circumstances, this type of anastomosis should be avoided.

The clinical history is distinctive and often diagnostic. Patients experience a classical syndrome in which eating is regularly followed by right upper quadrant epigastric distention and pain, borborygmus, and cramps, all of which are suddenly relieved by an almost projectile vomitus of clear bile that is never mixed with food.

The diagnosis can be made on the basis of history, endoscopy, and, in some cases, by an upper gastrointestinal series showing a massively distended afferent loop. Operation is required for the relief of these symptoms; in most instances the afferent loop is anastomosed into a Roux-en-Y efferent loop approximately 60 cm downstream to prevent the reflux of bile into the stomach. A vagotomy is necessary because bile and pancreatic juice are diverted from the gastroenterostomy, increasing the propensity for a marginal ulcer.

Alkaline Reflux Gastritis

In alkaline reflux gastritis, also called bilious vomiting, the stomach is sensitive to bile, patients complain of severe continual burning epigastric pain, and eating is associated with

pain rather than relief of pain. Vomiting may be present. The vomitus may be bilious and may contain food. Endoscopy reveals large amounts of bile emanating from the afferent loop, and there is a beefy-red gastritis with an abrupt cutoff of the gastritis to normal jejunal mucosa at the anastomosis. Ritchie and associates demonstrated increased bile salts in bile gastritis. Biopsy analysis reveals acute and chronic inflammation, evidence of a decrease in parietal cells, an increase in mucus-secreting cells, and intestinalization of the gastric glands.

The most effective treatment is the combination of cholestyramine and sucralfate. Histamine H₂-receptor antagonists or omeprazole often make the syndrome worse. Aluminum-containing antacids and metoclopramide have variable effects. Cranberry juice may afford some relief. If the diagnosis is made and the pain is intractable, and relief by medication is unattainable, a Tanner-19 procedure with a vagotomy and a long bile- containing loop anastomosed 60 cm downstream is our preference, though some prefer a Roux-en-Y. The advantage of the Tanner- 19 is the absence of the Roux-en-Y syndrome, and if the diagnosis is actually efferent loop obstruction, a Tanner-19 type of operation will correct the obstruction.

Nutritional Complications

Late complications of the postgastrectomy syndrome include cases in which patients become “nutritional cripples,” with weight loss of 10 to 30 pounds resulting from excision of the majority of the stomach. While fat absorption may be near normal (fecal fat excretion of less than 5 g/24 h) after a Billroth I procedure, fat excretion averages approximately 9 g/24 h in most patients with a Billroth II anastomosis. This fat malabsorption may be accompanied by chronic nutritional deficiency, failure of absorption of fat- soluble vitamins, and chronic bile salt diarrhea. Iron and calcium also are absorbed primarily in the duodenum. After a Billroth II procedure, many patients are persistently hypocalcemic, and iron-deficiency anemia may be present. The loss of intrinsic factor after partial gastrectomy necessitates monthly injections of vitamin B₁₂; otherwise a megaloblastic anemia might result. If iron, calcium, or fat malabsorption is crippling, conversion from a Billroth II to a Billroth I anastomosis can be considered. After conversion to a Billroth I anastomosis, the mixing of food with bile and pancreatic juice may be restored to near normal, and fat absorption may improve. Weight gain and the reversal of malabsorption of fat-soluble vitamins results.

Stomal Complications

Planning, siting, and dialogue with the patient and an enterostomal therapist help to avoid patient dissatisfaction with a stoma. An enterostomal therapist is an essential member of the team. Some patients will have difficulty obtaining a satisfactory stoma. These include patients who are obese, in whom the thickness of the abdominal wall prevents sufficient mobilization of the colon or, especially, the ileum to obtain a well-functioning stoma without compromising blood supply. Patients with numerous abdominal scars, especially concave scars, and patients with many fat folds on the abdomen also present difficult stomal problems.

The stoma should be placed through the rectus muscle, splitting the muscle in a longitudinal fashion so that it can be closed easily, e.g., ileostomy accompanying a ileal pouch/anal anastomosis procedure, or a temporary colostomy with a coloanal type of anastomosis. Stomas work best when they are on a convexity of the abdomen when the patient is upright. Five centimeters of flat, unscarred, nonconcave abdominal wall for the stoma appliance to seal is optimal. The stoma should be a sufficient distance from the umbilicus and the anterior superior iliac spine, whether in the right or left lower quadrant. There should be no tension on the bowel as it is brought up through the fascia, and the bowel should be fixed to the fascia with fine nonabsorbable sutures to prevent it from falling back into the abdomen. The fascial opening should be adequate; two fingersbreadth is reasonable in a normal-sized patient. The stoma should not be placed in a skin crease or at the belt line, where care of the stoma would be difficult.

Recurrence of Disease

The incidence of complications for ileostomies is about 4 percent for patients with ulcerative colitis and up to 30 percent in patients with Crohn's disease. Crohn's disease at the stoma may take the form of recurrence in the bowel, in which case granulomatous, ulcerated, swelling, and ultimately fibrotic bowel develops at the stoma. Crohn's disease may also result in peristomal fistulas at a distance or immediately adjacent to the stoma. A combination of ciprofloxacin and metronidazole should be initiated. There is little advantage in resiting the stoma, because whatever led to the recurrence of Crohn's disease usually leads to recurrence in the new site as well. Recurrence of Crohn's disease in the peristomal skin results in a sharply punched out, but undermined, violaceous ulceration that is difficult to treat.

Stomal Necrosis and Retraction

Inadequate vascularization of the stoma can lead to ischemia or necrosis in the immediate postoperative period. The development of duskiness or frank necrosis of the stoma should prompt an evaluation to determine the extent of involvement. A good light and a lubricated glass or plastic test tube inserted into the stoma can help in ascertaining the level of viability. If the necrosis is superficial to the fascia, no immediate action is required; the necrotic tissue will slough or require debridement. Stricture formation usually results. If the necrosis extends below the fascia, immediate laparotomy and reconstruction of the stoma is indicated to prevent the development of peritonitis.

Inadequate mobilization of the mesentery or poor fixation of the stoma to the skin or fascia can result in retraction of the stoma, usually during the early postoperative period. Retraction below the level of the fascia requires immediate laparotomy to prevent further fecal contamination of the peritoneal cavity. Retraction above the fascia does not require intervention, but it usually results in poor fitting of the appliance and leakage because of loss of the spigot configuration.

Skin Complications

Skin complications are usually a result of siting and the inability to obtain an appropriate seal around the stoma. To avoid this complication, the stoma should be one-quarter inch above the skin level so that a satisfactory fit of the appliance is obtained. In difficult situations, flexible appliances especially fitted for a particular patient, may be appropriate. Once the skin breaks down, it is difficult to heal, and the intervention of an ostomy nurse is invaluable.

The critical issue for healing of peristomal skin problems is the placement of a Stomadhesive that can be left in place for 5 to 7 days. Medication can be applied under the Stomadhesive on the skin to effect healing. Caraya powder, ion-exchange paste (to keep the pH of the intestinal contents at acidic rather than basic pH, preventing activation of enzymes), and, when yeast is present in the effluent, nystatin powder and systemic fluconazole are helpful in solving peristomal problems. Cellulitis of the surrounding skin requires antibiotic therapy. Appropriate cultures should be taken to rule out unusual pathogens. Any tendency to high ileostomy output, such as in short-bowel syndrome, should be treated with opioids, thickening agents such as Kaopectate, and, if necessary, codeine. Bulk formers, such as Metamucil or fiber tablets, also thicken ileostomy contents so that they will be more manageable. When breakdown is severe and the situation uncontrollable, the patient may need to be admitted to the hospital for placement of a sump tube within the ileostomy. Parenteral nutrition may be necessary to decrease ileostomy output.

Stomal Stricture

Although stomal stricture is a late complication, it is caused by the development of serositis in the immediate postoperative period. Primary maturation by approximating the intestinal mucosa to the skin at the time of stoma construction has drastically reduced the development of serositis. The most common cause of stomal stricture is necrosis or retraction, resulting in mucocutaneous separation, exposure of the serosa, and subsequent serositis.

A superficial stricture of the stoma can sometimes be dilated, but a minor procedure is often indicated. Under local anesthesia, the stoma is separated from the skin, the skin opening is enlarged, and a new maturation is performed. If the stricture is at the fascial level, the fascial opening is enlarged and the bowel fixed to the newly enlarged opening. It rarely is necessary to reexplore the abdomen and to free the fascial stricture from within the abdomen.

Peristomal Hernias and Prolapse

Peristomal hernias and prolapse may be coincidental. Stomal prolapse occurs when there is vigorous peristalsis and insufficient fixation of the bowel to the underside of the anterior abdominal wall. There usually is some component of a peristomal hernia; a local procedure (resecting the stomal prolapse and fixing it again in place) results in recurrence. In a good percentage of cases the entire stoma must be repositioned. If the stoma is temporary, accelerated closing of the stoma is appropriate; it is often much easier to eliminate an ostomy rather than fix a stomal prolapse. In the event of a peristomal hernia, it is necessary to site the stoma in a new area. The fascial defect of the peristomal hernia should be closed in the usual fashion, making certain that a good hernia repair can be obtained.

METABOLIC COMPLICATIONS

Syndrome of Inappropriate Secretion of Antidiuretic Hormone (SIADH)

The syndrome of inappropriate secretion of antidiuretic hormone (SIADH) is perhaps the most common metabolic complication after surgery and is especially frequent in the elderly. It usually is easily managed but if not promptly recognized and treated may result in central nervous system damage, seizures, and even death. The basis for the inappropriate secretion of *ADH* is that when the various receptors perceive that the volume within the vascular system is not adequate, ADH and aldosterone are secreted. In the case of inappropriate ADH secretion, the secretion of ADH is more prolonged or more intense than after normal operative procedures. The development of this syndrome is augmented by the tendency to overhydrate patients postoperatively, especially with hypotonic solutions.

If the extent of the inappropriate secretion of ADH causes slight edema and a serum sodium concentration that is below the range of 125 to 130 mEq/L, fluid restriction is all that is necessary because the syndrome ultimately will correct itself. If, however, the serum sodium is more significantly reduced and the patient, especially an elderly patient, manifests central nervous system symptoms such as confusion and somnolence, medical therapy is necessary. If the symptoms are not severe, mannitol may be given slowly. This provokes a diuresis of excess water secreted with a minimum of sodium, adjusting the tonicity. Small doses of furosemide can be added if sodium is given intravenously or orally. If the situation is acute and there has been a seizure brought on by hyponatremia, administration of 3% saline solution is the therapy of choice. Three percent saline should be given in small increments of 50 to 100 mL over 3 to 4 h, provoking a diuresis of excess water and restoring the serum tonicity toward normal. This must be done slowly because permanent central nervous system damage may result from too rapid correction of the hyponatremia.

This syndrome is far better prevented than treated. Patients postoperatively should be appropriately resuscitated, but not overresuscitated. Free water in the form of 5% dextrose in water should be minimized. In the normal course of postoperative recovery, a certain amount of proteolysis and lysis of fat normally takes place, which generates between 500 and 800 mL of electrolyte-free water per day.

Disorders of Thyroid Metabolism

Thyroid Storm

While not as common as in the past, thyroid disorders, especially in patients with undiagnosed latent hyperthyroidism, can be very dangerous. Thyroid storm is the most common and the most dangerous. It occurs in patients with existing thyrotoxicosis that is unrecognized or uncontrolled. The abrupt administration of iodides without the protection of antithyroid drugs may also precipitate a thyroid storm. Any traumatic event, such as surgery, infection, or embolism, may complicate thyrotoxicosis and provoke thyroid storm.

Tachycardia, high fever, and mental status changes are the most prominent early symptoms. Patients are intensely irritable and manifest confusion, delirium, or perhaps coma. Tachycardia is initially sinus, but if thyroid storm persists, high-output cardiac congestive failure may result in tachyarrhythmias. Once hypotension supervenes, it is a preterminal event. Irreversible cardiac failure usually is the mode of death. In other patients, gastrointestinal side effects, such as diarrhea, nausea, vomiting, and abdominal pain also may be present, but these symptoms are secondary.

Treatment of thyroid storm, which has a mortality of 10 to 20 percent, involves control of the catecholamine-induced cardiac symptoms. Propranolol or other beta blockers are given 1 mg/min intravenously to a maximum of 10 mg to control the tachycardia and stabilize the cardiovascular system. Inotropic agents that do not promote tachycardia, such as dobutamine, may be necessary. Diuretics can be helpful, but if furosemide is given, care should be taken to replace the potassium. Propylthiouracil 200 mg and potassium iodide 5 to 10 drops are given to decrease triiodothyronine (T₃) and thyroxine (T₄) release. Hydrocortisone 200 mg intravenously as an initial dose and 100 mg every 8 h thereafter diminishes thyroid hormone release.

Myxedema Coma

An even rarer form of thyroid dysfunction that may result in a fatality is myxedema coma. The principal patients at risk are those with chronic hypothyroidism that is unrecognized or inadequately controlled and whose hypothyroidism is provoked by the stress of operation. Any number of inciting factors may provoke myxedema coma, including trauma, infection, gastrointestinal bleeding, surgery, and the administration of narcotics and phenothiazines. In these patients, there are signs of a metabolic slowdown, including hypothermia, hypoventilation, bradycardia, obtundation, psychotic behavior, seizures, or coma. Once the situation is suspected, serum levels of T₄, thyroid-stimulating hormone (TSH), and serum cortisol should be drawn and therapy initiated before the return of the laboratory values. Supportive care includes warming, hydration, and assisted ventilation. L-thyroxine 300 to 500 mg is given intravenously, followed by a maintenance dose of 50 to 100 mg/ day. Serum cortisol level should be measured and care should be taken to make certain that the patient is not hypoadrenal; 100 mg cortisol is sufficient to ensure that hypoadrenalism is not a problem.

Adrenal Insufficiency

Adrenal insufficiency occurs because of suppression of the pituitary-adrenal axis by previous administration of steroids or the destruction or exhaustion of the adrenal glands. In the former, postoperative adrenal insufficiency, particularly in a very stressful situation, may occur within 1 year of discontinuing steroids despite the absence of any symptoms indicating adrenal insufficiency.

Adrenal insufficiency complicating other forms of severe illness, such as sepsis, hypovolemia, shock, or trauma, can result in unexplained hypotension that will end fatally unless diagnosed and treated. In patients with carcinoma, unsuspected bilateral adrenal metastases may occur, but this has become less common with the frequent use of computed tomography and magnetic resonance imaging.

Acute adrenal insufficiency is manifested by unexplained hypotension, fever, abdominal pain, light-headedness, weakness, palpitations, mental status changes, nausea, and vomiting. Confirmatory laboratory findings include hypoglycemia, hyponatremia, and occasionally hyperkalemia. If the diagnosis is suspected, serum cortisol level is measured and treatment initiated. The mortality and morbidity associated with hypoadrenalism are significant because of prolonged hypotension and hypoglycemia resulting in central nervous system damage.

Treatment consists of 200 mg hydrocortisone given intravenously. Hypotension should resolve within 1 or 2 h if the diagnosis is correct. A solution of 5% dextrose in saline should also help to correct the hypotension. If not, 400 mg hydrocortisone in divided doses over 24 h should be given and can rapidly be tapered to a maintenance dose.

Liver Failure

The most common cause of liver failure is preexisting liver disease. Patients with cirrhosis, alcoholic hepatitis, or fatty infiltration secondary to alcoholism requiring surgical intervention are prone to the development of liver failure postoperatively. In patients with established liver disease such as cirrhosis, general anesthesia should be avoided if possible; in cirrhosis, the portal vein's contribution to hepatic perfusion of the liver is diminished, and the hepatic artery supplies at least 50 percent of hepatic blood flow. Since the hepatic artery is a splanchnic artery, general anesthesia, hypotension, shock, blood loss, and hypovolemia result in splanchnic vasoconstriction and hepatic ischemia as portal flow and hepatic artery flow are markedly decreased. A regional or epidural anesthetic is preferred in patients with liver disease.

Any stress, even in the absence of infection, may result in liver failure beginning on the third to fifth postoperative day. Somnolence, jaundice, diminished urine output, and ascites are manifestations of incipient hepatic failure. Liver function tests reveal an elevated bilirubin level, usually of the indirect fraction, some minor elevation of transaminase level, a decreased albumin level, and a lengthening of prothrombin time. Treatable reversible causes include hypovolemia, hypokalemia, hypomagnesemia, gastrointestinal bleeding, constipation, and remote infection. Spontaneous bacterial peritonitis must be ruled out by paracentesis, culture, and Gram stain. Other infections must be promptly treated.

Treatment of postoperative liver failure includes the correction of electrolyte abnormalities (especially hypokalemic alkalosis), the administration of neomycin, cathartics, or lactulose orally or by enema, and provision of nutritional support. Nutritional support is best carried out enterally if the patient tolerates tube feedings or by total parenteral nutrition. A modified, low aromatic, high branched-chain amino acid formulation is appropriate, preferably with 15 to 25% dextrose and less than 20% of the caloric requirements as fat.

Hemodynamic parameters and urinary output should be carefully monitored. A high-cardiac-output, low-peripheral-resistance hemodynamic state usually complicates hepatic failure. The cardiac index is elevated above 4 L/min/m². Systemic vascular resistance usually is low, with systemic hypotension and warm, well-perfused skin and extremities. Urinary output is decreased, and urinary sodium concentration is less than 10 mEq/ L. The BUN level rises out of proportion to that of creatinine, suggesting the Type II hepatorenal syndrome. Hemodynamic support can be initiated with dopamine and alpha-adrenergic agents to restore peripheral resistance. If the liver does not recover postoperatively, renal and high-output cardiac failure result in death.

PSYCHIATRIC COMPLICATIONS

There is an increased incidence of postoperative psychiatric complications in older patients. In most instances there is nothing more than disorientation in the evening, known as "sundowning." In such cases, setting the patient up in front of a television set or providing other external stimuli, such as a newspaper or music in the background, can help to reorient the patient. Delirium is diagnosed in 20 percent, depression in 9 percent, dementia in 3 percent, and a functional psychosis in 2 percent of elderly postoperative patients.

The first account of postoperative psychiatric disturbance presented by a surgeon was that of Dupuytren who wrote in 1834 that "the brain itself may be overcome by pain, terror, or even joy and reason leaves the patient at the instant when it is most necessary to his welfare that he should remain calm and undisturbed." In 1910, Da Costa indicated that the anticipated frequency for such complications is as high as 1 in 250 laparotomies, while Lewis suggested an incidence of 1 in 1500. Scott described 11 cases in 2000 surgical procedures.

Titchener and associates evaluated 200 patients admitted to the surgical service of the Cincinnati General Hospital using interviews and the Minnesota Multiphasic Personality Inventory to substantiate a psychiatric diagnosis. Although the patients in the study represented a lower socioeconomic status group, 86 percent of the sample had distressing psychologic symptoms, disabling patterns of behavior, or both; 21 percent were found to have neuroses, 11 percent psychophysiologic reactions, 14 percent psychoses, 34 percent character behavior disorders, and 3 percent chronic brain syndrome.

General Considerations

"Postoperative psychosis" cannot be considered a distinct clinical entity. No single factor has been shown to be responsible, and the physical illness and operative procedure may merely reveal a latent psychotic tendency. Illness, particularly when prolonged, and surgical procedures represent threats to the integrity of the organism on somatic and psychologic grounds. In nearly every person informed of the need for a surgical procedure, some degree of anxiety arises. There may be fear of loss of life or loss of body part or function, including, e.g., fear of castration in pelvic and hernia operations. The psychodynamic processes at work during the preoperative, postoperative, and convalescent periods may be classified as involving (1) psychophysiologic factors, (2) somatopsychic factors, or (3) psychosocial factors. Psychophysiologic factors represent processes originating from psychologic stress that act along neurogenic or humoral pathways to modify the healing process. A poorly functioning gastroenterostomy or marginal ulcer in a patient with emotional stress is an example. The somatopsychic factors have to do with the psychologic adaptation involved when the surgical procedure imposes a somatic defect, such as an ileostomy or colostomy. The psychosocial factors refer to patients' concern with the effects of their physical illness or surgical procedure on their position in society. All these may interplay and contribute to anxiety, neurotic symptoms, severe depression, and frank psychosis.

Clinical Manifestations

The occurrence of psychiatric derangement during illness is variable, and the duration of latent interval between surgical treatment and the psychologic disturbance may be days to weeks. Winkelstein and associates reported that, in the recovery room, patients who were subjected to surgical procedures under general anesthesia exhibited a lack of concern about the operation and an absence of affective response, despite the fact that they were sufficiently oriented to be interviewed. After 24 h, however, the patients responded with the concerns and emotions that had been so conspicuously absent in the immediate postoperative period. Both psychologic and pharmacologic factors are implicated in this response, since patients under spinal anesthesia exhibit immediate and overt emotional reaction.

The manifestations are extremely variable. Fear can be accompanied by depression or elation and overactivity. Clinical presentation may disclose acute delirium with confusion and disorientation or merely a vague alteration in perception and mood. The manic reaction may incorporate psychomotor excitement, delirium, delusions, visual or auditory hallucinations, agitated depression, and feelings of persecution. The psychotic reactions observed in 44 of 200 patients in the Cincinnati series are indistinguishable from the range of psychoses observed under other circumstances. The acute brain syndrome, or delirium, was manifest in 20 patients.

Delirium may begin with an inappropriate remark or a dramatic agitated outburst and is frequently the first sign of continued mental deterioration leading to a chronic brain syndrome, particularly in an elderly patient. Delirium must be regarded as a potentially dangerous situation. It occurs most commonly in elderly patients who have lost the closeness and support of family or friends and in patients who are immobilized for long periods.

Depressive reactions represented the second most important psychosis in the Cincinnati series, occurring in 4.5 percent of the sample. The patient is characteristically uncooperative in an active way or recovery may be impeded by listlessness, anorexia, and disinterest. The depressive reaction may be accompanied by physiologic changes. For example, Moore and associates demonstrated effects of emotion on the pituitary-adrenal axis during the immediate and postoperative period. Suicide is a major risk in patients with depressive reaction.

Another category includes the paranoid psychotic disorder. Although it is not rare for schizophrenic reaction to have its onset in the surgical patient, no acute breaks of the

schizophrenic type were noted among the 200 patients studied by Titchener and associates. There is no contraindication to operations for patients with schizophrenia. Manic excitement is a particularly difficult problem in the management of surgical patients and requires close cooperation of psychiatrist, surgeon, and anesthesiologist.

Knox indicated that the incidence of postoperative psychosis is not related to the duration of preoperative hospital stay. The duration of illness, particularly when prolonged, does determine the patient's psychologic reaction to surgical experience. By contrast, emergency operation often results in reactions marked by acute anxiety, nightmares, insomnia, irritability, and protective withdrawal from all stimuli. Age is an important factor, the highest incidence occurring in children under the age of two and in elderly patients. Knox presented evidence of constitutional predisposition, and although 17 percent of his patients had had previous surgical treatment uncomplicated by psychiatric disturbances, 11 percent had a previous psychiatric illness. Twenty- two percent of patients had a family history of mental illness of serious proportion. There is an increasing incidence of delirium in response to anesthesia and surgical treatment in patients who are alcoholic, while patients suffering from extensive trauma may have organic psychosis. Acidosis, acetonuria, hyperglycemia, hypercalcemia, hypomagnesemia, and hepatic insufficiency can all cause postoperative mental aberrations, and cerebral hypoxia frequently results in behavioral changes. Medications such as barbiturates, anticholinergics, and cortisone also have been implicated.

Management

There is a need to integrate psychologic treatment with the management of the surgical problem. As Titchener and Levine emphasize, it is not necessary, possible, or advisable for the management of psychologic problems to be turned over to psychiatrists, and it is frequently preferable that the measures be initiated by the surgeon in charge. Verbal communication between the surgeon and the patient is the best means of overcoming emotional or mental difficulty. The anesthesiologist is regarded as an impersonal figure who performs a task without emotional impact on the patient. The surgeon should be aware of the patient's feelings, attitudes, and needs for specific information on diagnostic procedures, operating approach, postoperative possibilities, and the like. Discussion of expected feelings and sensations may have a direct effect on the adequacy of adjustment. Also, changes in treatment to increase the patient's positive adaptation to the illness should be constantly considered. The incidence of significant postoperative disturbance suggests the need for a "mental check" to be incorporated into the usual postoperative surgical rounds. Efforts should be directed at removing toxic causes of the acute brain syndrome, removing unnecessary stimuli without isolating the patient, and providing psychologic or pharmacologic tranquilization.

The physician's psychologic approach should include repeated explanation and inquiries about the patient's concerns. In some instances, specific counseling and directive treatment, which may require direct intervention in the patient's personal or family affairs and the assistance of the social services department, is indicated.

The provocative patient who emits anger or attempts to irritate others as a mechanism for covering fear or relieving guilt needs understanding and an attitude of firmness rather than anger from the physician. A patient's attempting to sign out of the hospital against advice is a mechanism of expressing anger or fear and should be handled by the surgeon in a way that allows the patient to change his or her mind without embarrassment.

Consultation with a psychiatrist is indicated in the case of any acute and severe emotional disturbance. The referral should be candidly discussed between the surgeon and the patient. Patients must come to the conclusion that they require help to resolve their problems. Referral also is indicated for long-standing disturbances discovered during hospitalization and is frequently appropriate in patients with psychosomatic illness. Browning and Houseworth, in a study of patients with peptic ulcer, demonstrated that the removal of symptoms without altering the psychosomatic disorders led to the formation of a new set of symptoms. The surgeon should be prepared to differentiate organic from functional disorders, psychosis, and depressive states, because psychiatric consultation may not be available or may be refused by the patient. Specific drugs may be prescribed. Most postoperative traumatic neuroses, manifested by anxiety and reliving the operative experience, can be managed with minor tranquilizers such as diazepam or lorazepam or hypnotics such as flurazepam. Psychoses, such as schizophrenia, mania, and depression, may respond to a phenothiazine derivative. Benzodiazepine derivatives are preferable for nonpsychotic anxiety. Tricyclic antidepressants have an effect that is often delayed 1 to 3 weeks, and they are associated with acute cholinergic side effects and changes in cardiac conduction. These drugs are rarely used in the postoperative period. The newer class of S1 antagonists are rapid acting but may have significant side effects, including dyspepsia and other gastrointestinal complaints.

Delirium Tremens and Other Forms of Delirium

As drug and alcohol abuse become more prevalent, the surgeon must be aware that a relatively normal-appearing patient may undergo withdrawal from alcohol, benzodiazepines, or narcotics in the postoperative period. It may be difficult to distinguish between withdrawal from one type of medication and another.

Delirium is defined as a transient disorder with an impairment of cognitive functions and widespread disturbances in cerebral and entire body metabolism. It is a potentially fatal complication if not corrected. Delirium usually follows operation within 48 h but may be delayed. There usually is a lucid interval postoperatively while withdrawal from the agent is progressing. The patient typically is hyperactive with irritability, delusions, hallucinations, restlessness, and agitation and is noncooperative with treatment. There is a rare hypoactive variant in which the patient is apathetic or somnolent. Speech ranges from the incoherent to the rambling and irrelevant. There may be urinary and stool incontinence, and a loss of motor coordination may be noted.

The cause is multifactorial (Table [11-15](#)). Certain features, such as rapid- onset nystagmus, may indicate withdrawal of psychoactive medications such as sedatives or tranquilizers. Hypoxia, disorders in calcium metabolism, acidosis, and other endocrinopathies should also be ruled out by obtaining "confusion chemistries." A toxicology screening should be ordered to assure that drug withdrawal is not an issue.

Haloperidol 2 to 15 mg orally twice a day or 1 to 5 mg intravenously followed by 5 to 10 mg/h may help agitation, but an occasional patient, particularly among the elderly, may have a paradoxical response. Extrapyrimal symptoms that sometimes complicate the use of haloperidol are relieved with parenteral bntzropine mesylate. In patients with atropine-type psychosis, hepatic failure, or alcohol or drug withdrawal, a neuroleptic agent is not the drug of choice. If benzodiazepine or promethazine hydrochloride (Phenergan) drug abuse is suspected, a small dose of Phenergan with gradual tapering may prevent the emergence of full- blown withdrawal. Prophylactic medication with lorazepam should be administered in the perioperative period to patients with severe alcoholic histories who are candidates for delirium tremens.

Depression

Postoperative depression is not uncommon. It characteristically occurs rather late in the postoperative period. Major manifestations are inability to sleep, anorexia, loss of energy, and a reversal of the day-night rhythm. There may be agitation, but apathy and withdrawal are more common. There is decreased communication and movement. Suicidal tendencies may be present.

A peculiar form of depression occurs in patients in whom life-threatening situations have occurred, such as fistulas, chronic sepsis, major surgical catastrophes, etc. In these patients, depression usually does not occur until the acute syndrome has passed. It is as if these patients hold themselves together until it is clear to them that they will survive. At this point they allow themselves the "luxury" of emotional disintegration.

The use of the newer class of S1 antagonists and second- generation S1 antagonists are useful, although they may have gastrointestinal side effects. A minimal daily dose of 20 mg of paroxetine hydrochloride (Paxil) or 50 mg of sertraline hydrochloride (Zoloft) may not be sufficient, and Paxil may be required at a dose of 40 mg/day and Zoloft at 100 mg/day. Psychiatric consultation can be useful. As the patient understands the cause and leaves the hospital, the depression usually reverses and the medications are discontinued.

Special Surgical Situations

Very young and old patients are particularly vulnerable to the development of psychiatric complications after surgical treatment. Psychotic disturbances have been found in 2 to 3 percent of patients after cataract extraction. The combination of surgical procedure and awareness of the implications of the illness is critical in the patient with cancer. Special consideration is indicated for patients undergoing mastectomy and gynecologic procedures, cardiac surgical treatment, dialysis and transplantation, and prolonged periods in an intensive care unit. The management of drug addicts is assuming greater importance.

Pediatric Surgery

In children, severe anxiety states may be precipitated by the shock of operation. Levy reported that of a group of 124 children who had operations, 20 percent showed residual emotional disturbances. This occurred most frequently in the 1- to 2-year-old group; after the age of three there was a sharp decrease. The age distribution was attributed to a greater dependence on home and mother, and Levy suggested postponement of elective surgical treatment until the child could comprehend the situation. Postoperative reactions consisted of negativism, disobedience, tantrums, defiance, destructive behavior, and dependency manifested by clinging to the mother or attendant. The responses have been related to a feeling of betrayal and the consequent desire for revenge and rebellion. When a child is suffering from fears engendered by an operation, a second operation usually intensifies the earlier fears.

Prophylactic therapy is important. The maturity of the child's emotional adaptation is more a factor in the response than the operation per se. Parental absence frequently is associated with emotional difficulty. Prugh and associates compared two groups, one treated without organized consideration for emotional needs and another in which these needs were considered and ample opportunity for play was provided. Moderate or severe anxiety reactions immediately after leaving the hospital were observed in 92 percent of the control group and in 68 percent of the "treated" group, with a peak incidence in children under the age of three. Three months after discharge, the incidence of persisting anxiety had fallen to 58 percent for the control group and 44 percent for the "treated" group. The younger children reacted more severely, with apprehension, feeding disturbances, and depression. The pattern for the 4- to 6-year-old group was a tendency toward obsessive worries, phobias, and accentuated aches and pains. The 6- to 10-year-olds manifested conversion symptoms, compulsive behavior, and restlessness.

Surgery in the Aged

Elderly patients are more prone to becoming emotionally disturbed when confronted with new situations, especially if they have inadequate comprehension and a generalized feeling of insecurity. The operative procedure also presents an obvious physical threat to the integrity of the nervous system. Titchener and associates reported a 25 percent incidence of significant, and in some cases irreversible, change in cerebral function in the Cincinnati study patients over the age of 65 years. Some degree of depression was observed in 90 percent of the older patients and was disabling in about 50 percent. The indifference of family, friends, and society contributed to the evolution of a paranoid cycle. Efforts should be made to familiarize the patients with the hospital and personnel, and visitors should be encouraged to maintain human contact and prevent withdrawal. Collaboration with a social worker frequently is indicated for long-term rehabilitation.

Gynecologic and Breast Surgery

Removal of the breast and a variety of gynecologic procedures are highly represented in most series on postoperative psychosis. Maquire and associates found that 1 year after mastectomy the women reported a 20 percent incidence of depression, a 10 percent incidence of anxiety, and a 38 percent incidence of sexual difficulties. Contact with other

mastectomy patients expedites psychologic rehabilitation. Routine counseling lowers the postoperative psychiatric morbidity from 38 to 12 percent. Hysterectomy is associated with emotional disturbance more frequently than other gynecologic operations. The more the procedure antedates menopause, the greater the likelihood of associated psychologic disturbance. The loss of menstrual function is perceived by the woman as a blow to normal feminine esteem. Hollender reported that of 203 women admitted to psychiatric hospital, nine had pelvic surgical treatment as a precipitating event, compared to a total of five women admitted after operations of other kinds. Lindemann noted that the relative frequency of restlessness, insomnia, agitation, and preoccupation with depressive thoughts was greater after pelvic operations than after cholecystectomy.

Cancer Surgery

Cancer patients are exposed to two major threats—disease and extensive surgical treatment. They are concerned with death or injury during operation and with disruption of their pattern of living as a result of the effects of cancer or the surgical procedure. Anorexia, insomnia, tachycardia, fear, and panic may occur. Acute depression with suicidal tendencies has been reported in patients anticipating surgical procedures. Postoperatively, depression is related to an anticipated interference with valued activities. Sutherland and associates demonstrated that colostomy imposed on almost all patients a new order of living, and the patients were motivated to avoid social rejection. A rigid life arose from the fearful expectation of rejection because of the colostomy, combined with the fear of death from cancer. There is a tendency toward seclusion, withdrawal, and nonparticipation. Periods of depression are frequent. The study reported that loss of an important bodily part or function is more depressing than the fear or expectation of death. The management of patients with carcinoma must be based on an appreciation of their frequently suffering a sense of isolation, guilt, and abandonment.

Cardiac Surgery

Serious psychiatric disturbances have been observed to occur with considerable frequency after mitral valvulotomy and open-heart surgery. Studies have reported incidences of 16 to 19 percent of serious emotional disturbance after mitral valve surgery. Bolton and Bailey, however, in an evaluation of 1500 consecutive patients, noted an incidence of psychosis in 3 percent with no relation to age, sex, severity of heart disease, duration of failure, or complications of surgical treatment. Egerton and Kay noted delirium in 25 of 60 adults after open-heart surgery.

Manifestations generally occur after an initial lucid interval of 3 to 5 days and resolve shortly after the patient is transferred from an intensive care unit to a standard hospital ward. Postoperative incapacitation and increased time on the heart-lung machine apparently are factors increasing the likelihood of delirium, and incidence is unrelated to age and sex. Zaks has suggested that cardiac operation may produce organic brain damage, sensitizing patients and increasing the incidence of postoperative psychologic symptoms. A prediction equation was successful in differentiating reactors from nonreactors. Using the ego strength variable of the Minnesota Multiphasic Personality Inventory, a significant inverse correlation was found between the reaction and the incidence of acute psychotic episodes after cardiac operation. The incidence of psychosis is greater in males, older patients, and those expressing minimal preoperative anxiety. A preoperative psychiatric interview reduces the incidence of postoperative psychosis by 50 percent.

After operations on the heart, patients with emotional disturbances have been found to manifest perceptual distortion, visual and auditory hallucinations, disorientation, and paranoia. Egerton and Kay reported that 28 percent of adult patients who underwent open-heart surgery had delirious states ranging in duration from several nights to several weeks, averaging 5 days. The delirious patients had no psychologic sequelae, and no relation was established between the incidence of delirium and the duration of cardiac bypass, but open-heart procedures were more likely to produce delirium than other intrathoracic operations. Precipitating factors for delirium included dehydration, hyponatremia, and the performance of a tracheostomy, and predisposing factors included a familial history of psychosis, previous brain damage, overwhelming personal problems, and the presence of a rheumatic valvular lesion. Other psychiatric disturbances noted in patients after open-heart surgery were disabling anxiety state, conversion hysteria, tension headaches, and exacerbation of peptic ulcer. The almost total absence of delirium and other emotional disorders in children is particularly interesting and may be related to the fact that the concept of death as a permanent biologic process usually does not develop until the age of nine.

Dialysis and Transplantation

The suicide rate is 300 times greater in dialysis and transplantation patients than in a comparable healthy population. Uremia, debilitating disease, and the undergoing of repeated procedures are contributing factors. Wright and associates followed 11 patients on chronic dialysis and noted several stresses affecting them, such as unpredictability of well-being, tensions arising in marital relations from guilt and anger, effects of separation on the families, and financial anxiety. After each episode of dialysis, the main response was relief. Cramond and associates noted that their patients first denied their illness but later realized the loss of health and independence and the uncertainty of their future. This has been referred to as a “mourning reaction.” At times, the patients wished to be dead. They felt that life dependent on chronic dialysis was not worth living. Some patients passed from the “mourning reaction” to a state of active depression.

All patients undergoing dialysis become extremely dependent and emotionally attached to the staff and often react emotionally with a sense of loss when any replacement of staff occurs. During the course of the dialysis program, regression occurs relatively frequently, and the patient becomes withdrawn and pretends to sleep. Insomnia and frightening dreams also occur, and the frequency of emotional disturbances suggests that psychiatric assistance has an important role in a dialysis program.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 12 - Physiologic Monitoring of the Surgical Patient

Albert J. Varon
Orlando C. Kirton
Joseph M. Civetta

The primary reason for the surgeon's involvement in bedside critical care is the opportunity to understand and enhance the patient's physiologic response and to recognize and correct the pathophysiologic challenges. To do this effectively, the surgeon must understand physiologic monitoring. Without a thorough knowledge of the physics and methods of monitoring, ensuring the quality of numbers obtained, perceiving their importance, and using measurements as a guide for therapy, selection of proper therapy would be difficult, without foundation, rote, or naive. Thus there are many stimuli to obtain a fundamental knowledge of physiologic monitoring. This chapter is designed to initiate a lifelong process, one that extends the capabilities of the surgeon, improves patient outcome, and advances surgical science.

HEMODYNAMIC MONITORING

The traditional clinical evaluation, usually the initial assessment tool, is often unreliable in critically ill patients, since there may be major changes in cardiovascular function that are not accompanied by obvious clinical findings. Invasive hemodynamic monitoring at the bedside provides information about cardiorespiratory performance and guides therapy on a rational physiologic basis.

Arterial Catheterization

Indications

Arterial catheterization is indicated whenever there is a need for continuous monitoring of blood pressure and/or frequent sampling of arterial blood. States in which precise and continuous blood pressure data are necessary include shock of any etiology, acute hypertensive crisis, use of potent vasoactive or inotropic drugs, high levels of respiratory support (high intrathoracic pressure), high-risk patients undergoing extensive operations, controlled hypotensive anesthesia, and any situation in which any of the factors affecting cardiac function is rapidly changing. This is particularly true in patients with shock, because indirect measurement of blood pressure by a cuff has been proved inaccurate. Sequential analyses of blood gas tensions and pH are necessary in any acute illness involving cardiovascular or respiratory dysfunction or when hyperventilation is instituted in patients with central nervous system injuries. An indwelling arterial catheter also can provide ready access for other blood samples necessary to chart the progression of multisystemic illness.

Inserting arterial lines is a relatively safe and inexpensive procedure. There are no absolute contraindications to arterial catheterization per se, although bleeding diathesis and anticoagulant therapy may increase the risk of hemorrhagic complications. Severe occlusive arterial disease with distal ischemia, the presence of a vascular prosthesis, and local infection are contraindications to specific sites of catheterization.

Clinical Utility

With an indwelling arterial catheter and monitoring system, the systolic blood pressure (SBP), diastolic blood pressure (DBP), and *mean arterial pressure (MAP)* can be displayed continuously. The pulse rate can be calculated from the arterial tracing when the electrocardiogram (ECG) is not available (e.g., during electrocautery use in surgery).

Direct measurements of arterial pressure correlate rather poorly with indirect measurements. The disparities are due in part to physiologic considerations but are largely conditioned by the frequency response of the monitoring systems. Because blood pressure trends are probably more important than absolute values, the most important aspect of direct arterial pressure monitoring is that it constantly reminds the clinician to pay attention to the patient, to think about what is happening, and to reason why changes are occurring.

To obtain accurate data when measuring any pressure within the vascular system, the clinician must understand the monitoring system and methods of calibration. Minor details such as the use of long tubing and the presence of air bubbles or blood clots in the system can make the measurements unreliable.

Observation of the arterial pressure waveform obtained with an arterial catheter and monitoring system may permit a qualitative assessment of the patient's cardiovascular status. The shape of the arterial pressure tracing represents a particular stroke volume ejected at a particular state of myocardial contractility. Qualitative interpretation can be made in a hypovolemic patient with a small stroke volume that will create a smaller pressure wave. As intravascular volume is replenished, the stroke volume increases, and the arterial pressure tracing will increase in size until it attains normal shape. If myocardial contractility is diminished, the rate of increase in aortic pressure will diminish, and the upslope of the arterial pressure tracing will become less vertical and assume a more tangential trajectory with the apex moved to the right.

Although quantitation of stroke volume has been attempted using computers to solve the equations necessary to relate the shape of the peripheral arterial pressure tracing to actual stroke volume ejected, critical illness introduces too many variables for this measurement to be reliable. The location of the dicrotic notch on the arterial waveform also has been advocated as an indicator of the systemic vascular resistance; however, Gerber and associates were unable to demonstrate any statistically significant correlation.

Analysis of the SBP variation during mechanical ventilation may offer important information about the nature of low-flow states. The normal decrease in SBP after a mechanical breath is more pronounced during hypovolemia but practically nonexistent during congestive heart failure.

Sites of Catheterization

Many anatomic sites have been used to access the arterial circulation for continuous monitoring. The superficial temporal, axillary, brachial, radial, ulnar, femoral, and dorsalis pedis arteries have all been used. Although the selection of anatomic site for arterial catheterization usually has an institutional bias, specific advantages and disadvantages should be considered.

The dual blood supply to the hand and the superficial location of the vessel make the radial artery the most commonly used site for arterial catheterization. Cannulation is technically easy, as is securing the catheter in place, and there is a low incidence of complications. The mean and end-diastolic radial pressures are usually accurate estimates of the corresponding aortic pressures; however, the systolic pressure at the radial artery is often much higher than that of the aorta due to overshoot caused by the resonant behavior of the radial artery. This exaggeration is accentuated in stiff, arteriosclerotic radial arteries.

Most authors recommend assessing the adequacy of collateral circulation before cannulation of the radial artery. The most commonly used test is the modified Allen test. The patient is instructed to elevate one hand, make a fist, and clench it firmly, thus squeezing the blood from the vessels of the hand. After the examiner compresses at the same time both the radial and ulnar arteries, the patient lowers and opens the hand in a relaxed fashion (carefully so as not to overextend it). The examiner then releases the pressure over the ulnar artery, and the time for return of color is noted. It is considered normal if the capillary blush of the hand is complete within 6 s. Other methods such as ultrasonic Doppler technique, plethysmography, and pulse oximetry also have been used to assess the adequacy of the collateral arterial supply.

The axillary artery has been recommended as suitable for long-term direct arterial pressure monitoring. Its use has been associated with relatively few complications and no reported permanent sequelae. The major advantages include its larger size, freedom for the patient's hand, and close proximity to the aorta so that there is better representation of the aortic pressure waveform and minimal systolic pressure overshoot. Pulsation and pressure are maintained even in the presence of shock with marked peripheral vasoconstriction. Also, because of the extensive collateral circulation that exists between the thyrocervical trunk of the subclavian artery and the subscapular artery (which is a branch of the distal axillary artery), thrombosis of the axillary artery will not lead to compromised flow in the distal arm. Major disadvantages are its rather deep location and mobility, which increase the technical difficulty for insertion, and its location within the neurovascular sheath, which may increase the possibility of neurologic compromise if hematoma occurs.

The femoral artery also has been used for continuous blood pressure monitoring. Major advantages are its superficial location and large size, allowing easier localization and cannulation when the pulses over more distal vessels are absent. The major disadvantages are the presence of atherosclerotic occlusive disease in older patients and the problems associated with maintaining a clean dressing in the presence of draining abdominal wounds and ostomies in surgical patients. Furthermore, bleeding at this site may be difficult to control or may occur in an occult manner into the abdomen or thigh. Despite these potential disadvantages, studies have failed to demonstrate a higher complication rate in patients with femoral artery catheters.

The dorsalis pedis artery has no significant cannulation hazards if collateral flow can be demonstrated to the remainder of the foot through the posterior tibial artery. This can be done by occluding the dorsalis pedis artery, blanching the great toe by compressing the toenail for several seconds, and then releasing while observing return of color. A Doppler technique also can be used. Major disadvantages are its relatively small size (which makes it more difficult to cannulate) and overestimation of systolic pressure at this level.

The superficial temporal artery has been used extensively in infants and in some adults for continuous pressure monitoring. Because of its small size and tortuosity, however, surgical exposure is required for cannulation. Furthermore, a very small but worrisome incidence of neurologic complications due to cerebral embolization has been reported in infants.

The brachial artery is not used often because of the high complication rate associated with its use for cardiac catheterization. Although this artery has been used successfully for short-term monitoring, there are little data to support the use of prolonged brachial artery monitoring. If collateral circulation is inadequate, obstruction of the brachial artery may be catastrophic, leading to loss of the forearm and hand. Other problems include the difficulty in maintaining the site in awake, active patients and the possibility of hematoma formation in anticoagulated patients. The latter may lead to median nerve compression neuropathy and Volkmann's contracture.

Complications

Common problems associated with arterial catheterization are failure to cannulate, hematoma formation, and disconnection from the monitoring system with bleeding. The majority of reports that describe the complications following radial artery cannulation have stressed the high incidence of early radial artery occlusion and the rarity of late ischemic damage. Recannulation of the occluded artery generally occurs but may take several weeks. The incidence of radial artery thrombosis has declined progressively as a result of the understanding of the effects of different catheter sizes (smaller is better) and materials (Teflon is better) and of the use of continuous heparin flow instead of intermittent flushing. Factors associated with an increased risk of radial artery occlusion include female gender, low cardiac output states, use of vasoconstrictor drugs, severe peripheral vascular disease, small wrist circumference, insertion by surgical cut-down, multiple puncture attempts, hematoma formation, and increased duration of cannulation.

Infections related to arterial catheterization also can occur. Factors associated with an increased risk of infection include placement of the catheter for more than 4 days, insertion by surgical cut-down rather than percutaneously, and local inflammation. The rate of catheter-related infection varies from 0 to over 9 percent, but the risk of catheter-related septicemia is very low.

Other possible complications include retrograde cerebral embolization (when flushing catheters), arteriovenous fistulas, and pseudoaneurysm formation. Finally, inadvertent injection of vasoactive drugs or other agents into an artery can cause severe pain, distal ischemia, and tissue necrosis.

Central Venous Catheterization

Indications

The most common indications for central venous catheterization are to secure access for fluid therapy, drug infusions, or parenteral nutrition and for *central venous pressure (CVP)* monitoring. Central venous catheters also have been used to aspirate air in case of embolism during neurosurgical procedures in the sitting position, for placement of cardiac pacemakers or inferior vena cava filters, and for hemodialysis access.

There are no absolute contraindications for CVP catheter placement, although bleeding diatheses may increase the risk of hemorrhagic complications. Vessel thrombosis, local infection or inflammation, and distortion by trauma or previous surgery are considered contraindications to specific sites of catheterization.

Clinical Utility

While central venous lines are placed primarily for venous access, useful information occasionally can be obtained by measuring the CVP. The CVP may be useful in a hypotensive trauma patient to differentiate a pericardial tamponade from hypovolemia. Analysis of the CVP tracing also may be helpful in the differential diagnosis of certain cardiac arrhythmias (a waves are absent in atrial fibrillation) and in the diagnosis of tricuspid insufficiency (prominent v waves).

A properly placed catheter can be used to measure right atrial pressure, which, in the absence of tricuspid valve disease, will reflect the right ventricular end-diastolic pressure. CVP, therefore, can give information about the relationship between intravascular volume and *right* ventricular function but cannot be used to assess either of these factors independently. CVP cannot be used to assess left ventricular function in critically ill patients because ventricular disparity and independence of right and left atrial pressures have been confirmed repeatedly in these patients. Furthermore, CVP is only a single parameter, in contradistinction to the more complete information concerning pressures, flow, and venous gas measurements available with pulmonary artery catheters.

When monitoring CVP, the catheter should be attached to a pressure transducer for electronic measurement rather than to a water manometer. Water manometry does not permit visualization of the pressure tracing and cannot provide reliable measurements because of the frequency-response limitations of a fluid-filled column that cannot respond to the full range of pressure variations.

Sites of Catheterization

There are many anatomic routes to obtain access to the central venous circulation. The most commonly chosen sites include the subclavian, internal jugular, external jugular, femoral, and brachiocephalic veins. The patient's anatomy and the operator's experience are the major factors influencing site selection.

The subclavian vein can be cannulated with a high rate of success and may be the easiest to cannulate in situations of profound volume depletion. Another advantage of this approach is the ease with which the catheter and the dressings can be secured. Disadvantages include the higher risk of pneumothorax and the inability to compress the vessel if bleeding occurs.

The internal jugular vein has been cannulated with success rates similar to those of the subclavian approach. The major advantages of internal jugular vein catheterization are the lower risk of pneumothorax and the ability to compress the insertion site if bleeding occurs. In addition, the right internal jugular vein provides a straight path to the superior vena cava, facilitating placement of catheters and pacemakers. The internal jugular vein, however, may be more difficult to cannulate in patients with volume depletion or shock. Fixation and dressing of catheters are also more difficult.

Cannulation of the external jugular vein has a lower incidence of complications but a higher incidence of failure. Since catheters inserted through the neck are more difficult to fix and dress than those in other sites, this approach is not suitable for prolonged central venous access.

Although some authors have reported no higher incidence of complications from femoral cannulation than from subclavian or internal jugular sites, concerns over the risk of infection and thrombosis continue to limit general acceptance of long-term femoral cannulation in critically ill patients. Other peripheral veins, such as those in the antecubital fossa, have been used for central venous access, but the high incidence of thrombophlebitis and the fact that many catheters cannot be passed into the central venous circulation make these routes undesirable in critically ill patients.

Complications

Complications can be divided into technical or mechanical complications, usually occurring during catheter placement, and long-term complications related to the length of time that the catheter remains in place. The list of technical and mechanical complications is truly impressive: catheter malposition, dysrhythmias, embolization (air or catheter fragments), vascular injury (hematoma, vessel laceration, false aneurysm, or arteriovenous fistula), cardiac injury (atrial or ventricular perforation or cardiac tamponade), pleural injury (pneumothorax, hemothorax, or hydrothorax), mediastinal injury (hydromediastinum or hemomediastinum), neurologic injury (phrenic nerve, brachial plexus, or recurrent laryngeal nerve), and injury to other structures (trachea, thyroid, or thoracic duct). Pneumothorax is the most frequently reported immediate complication of subclavian vein catheterization, and arterial puncture is the most common immediate complication of internal jugular vein cannulation. The literature suggests that serious mechanical complications of central venous catheterization, although extremely rare, are associated with a high mortality rate.

Long-term complications related to the length of time the catheter is in place are due to infection or thrombosis. Norwood and associates studied triple-lumen catheter infections in septic and nonseptic critically ill surgical patients. They found no catheter-related infections or instances of septicemia in the nonseptic patients, but the incidence of catheter-related infection in the septic group was 26.3 percent, with a 9.6 percent incidence of septicemia. The catheter infection rate per 100 days, however, was only 0.9 for both septic and nonseptic patients combined, which is very similar to rates previously published for single-lumen catheters. Surface-modified central venous catheters have been developed to reduce catheter-related infection. Catheters impregnated with silver sulfadiazine and chlorhexidine resist bacterial adherence and biofilm formation. These catheters have been reported to have a significantly lower proportion of catheter-related infection compared with standard catheters.

At least three types of thrombi can develop in patients with central venous catheters: mural thrombus, catheter thrombus, and “fibrin sleeve” or sleeve thrombus. Any of these thrombi may break loose spontaneously or may be set loose when the catheter is removed. Generally, however, symptoms or clinical consequences do not occur. Superior vena cava syndrome does occur, especially in long-term patients who have had many catheters placed.

Pulmonary Artery Catheterization

Indications

Several studies in critically ill patients have shown that the clinical assessment is inaccurate in predicting cardiac output, pulmonary artery occlusion pressure, and systemic vascular resistance and that the information obtained from pulmonary artery catheterization prompts a change in therapy in 40 to 60 percent of patients. Although the pulmonary artery catheter permits a more accurate hemodynamic assessment and therapy may be modified as a result, this does not prove that knowledge of these data and alteration of the therapy improve overall patient outcome. Some studies indicate that preoperative invasive hemodynamic monitoring and cardiac function optimization in high-risk patients are associated with reduced intraoperative and postoperative cardiac complications and decreased mortality. While there are not enough carefully designed studies to definitely establish the benefit of hemodynamic monitoring to the individual patient, it is reasonable to assume that more precise bedside knowledge of fundamental cardiovascular parameters would facilitate earlier diagnosis and guide therapy. Whether morbidity can be decreased and overall survival can be improved also depend on the patient's overall response, not just on improved cardiovascular function, and thus should not be considered a necessary requirement for initiating invasive monitoring.

In general, a pulmonary artery catheter is indicated whenever the data obtained will improve therapeutic decision making without unnecessary risk. Table [12-1](#) represents the indications most often noted in the medical literature. Variables that are particularly important in assessing benefit versus risk of perioperative use of a pulmonary artery catheter include disease severity, magnitude of anticipated surgery and fluid shifts, and practice setting. There are no specific contraindications to pulmonary artery catheterization, but the same cautions as those attached to central venous access apply.

Clinical Utility

The pulmonary artery catheter has provided a “quantum leap” in the physiologic information available for the management of critically ill patients. The information that can be obtained includes CVP, *pulmonary artery diastolic pressure (PADP)*, pulmonary arterial systolic pressure (PASP), *mean pulmonary artery pressure (MPAP)*, *pulmonary artery occlusion (“wedge ”) pressure (PAOP)*, *cardiac output (CO)* by thermodilution, mixed venous blood gases by intermittent sampling, and continuous mixed venous oximetry. On the basis of this information, a multitude of derived parameters also can be obtained (see below).

When the pulmonary artery catheter balloon is inflated (1.5 mL), the blood flow in a distal segment of the pulmonary artery is occluded, creating a conduit through which left atrial pressure (LAP) can be measured (Fig. [12-1 A](#)). In a tubular system, flow can only be created if there is a pressure differential at both extremes. If there is no pressure differential, flow cannot be present. Using this principle in reverse, a stagnant system in which no forward flow is present would permit an accurate measurement of a distal pressure from a proximal location (see Fig. [12-1 B](#)). In fact, simultaneous PAOP and LAP measurements in patients have validated this principle. The PAOP is a reliable index of the LAP even in the presence of elevated pulmonary vascular resistance. Although the PADP also has been used as an index of LAP, it is not as reliable as the PAOP, particularly if there is tachycardia or increased pulmonary vascular resistance.

The PAOP represents the LAP as long as the column of blood distal to the pulmonary artery catheter tip is patent to the left atrium. This may not be so if the catheter is positioned in an area of the lung where the alveolar pressure exceeds pulmonary venous pressure (zone 2, as described by West) (Fig. [12-2](#)) or both pulmonary artery and venous pressures (West's zone 1), causing intermittent or continuous collapse of the pulmonary capillaries. The PAOP may then reflect alveolar pressure and not LAP. This is particularly important if patients have low pulmonary vascular pressures (i.e., hypovolemia) and/or are treated with high levels of *positive end- expiratory pressure (PEEP)*. Fortunately, since the pulmonary artery catheter is flow-directed, it is most likely to pass into dependent areas of the lung where blood flow is high and both pulmonary artery and venous pressures exceed alveolar pressure (West's zone 3). In this location, the continuous column of blood between the distal lumen of the catheter and the left atrium will remain patent, and the PAOP will reflect LAP. Another factor favoring appropriate catheter position is that when the patient is supine, the volume of lung located above the heart and the hydrostatic gradient favoring the formation of zones 1 and 2 are decreased. If there is any doubt, a lateral chest x-ray can be used to determine the location of the catheter tip in relation to the left atrium. If the tip of the catheter is below this chamber, zone 3 conditions will exist even if high levels of PEEP are used.

In the absence of mitral valve disease or premature mitral valve closure due to aortic regurgitation, the LAP reflects the *left ventricular end-diastolic pressure (LVEDP)*. If there are no alterations in left ventricular compliance (the relationship between pressure and volume), LVEDP will reflect *left ventricular end-diastolic volume (LVEDV)*. In the intact ventricle, LVEDV reflects the end-diastolic stretch of the muscle fiber, which represents the true preload (discussed later).

Raising intrathoracic pressure introduces an artifact that affects all intrathoracic vascular pressures to an extent that depends on the state of pulmonary compliance. In patients with acute respiratory insufficiency, compliance is often diminished, and the “stiff” lungs do not transmit alveolar pressure as readily to the pulmonary circulation. In these patients, the PEEP artifact on the PAOP measurement usually should not exceed 1 mmHg for every 5 cmH₂O of PEEP applied. A greater discrepancy can be seen if the patient is hypovolemic or if the catheter is malpositioned as described above. Another method of evaluating the effects of PEEP on the PAOP measurement is to observe the decrement in PAOP when PEEP is briefly removed. Presumably, this decrement remains relatively constant and can be subtracted from subsequent pressure measurements. Although removal of PEEP may decrease arterial oxygen tension and increase physiologic shunt, these changes are rapidly reversible. If a physician believes that the PAOP should be measured off PEEP, this probably should be done when PEEP is discontinued for other reasons (suctioning or changing breathing circuits), and increased concentrations of oxygen should be given before and after PEEP is stopped. Patients who are receiving very high levels of PEEP or whose condition deteriorates when PEEP is discontinued (such as immediate bradycardia) should not have PEEP removed for the exclusive purpose of measuring PAOP.

Since intravascular pressure measurements are affected by the intrathoracic pressure changes during respiration, they should be performed at end-expiration and obtained from a calibrated strip-chart recorder or oscilloscope rather than from a digital display. Most digital displays are inaccurate because the unselective nature of time-based electrical sampling and averaging includes positive and negative breathing artifacts. The digital average then contains the very respiratory variations that can be excluded by visualizing the tracing and selecting the appropriate value.

The cardiac output is measured by the thermodilution technique, which correlates well with both the Fick and the dye dilution methods. Thermodilution represents an application of the indicator dilution principle in which a change in the heat content of the blood is induced at one point in the circulation, and the resulting change in temperature is detected at a point downstream. This change is produced by a rapid injection of a known volume of fluid at a known temperature (colder than the body) into the right atrium via the proximal port of the pulmonary artery catheter. The change in temperature is registered by a thermistor located 4 cm from the catheter tip. This lowered temperature decreases the electrical resistance of the thermistor and results in a thermodilution curve.

The measurement of CO is based on a modification of the Stewart- Hamilton equation:

CLICK HERE FOR EQUATION

The variables in the formula are essentially fixed before injection, except for the denominator. The denominator of the equation is the thermodilution curve produced by injection of the indicator. A computer integrates the area under this curve, and the resulting calculation is displayed as the cardiac output in liters per minute. The area under the curve is inversely proportional to the CO; that is, the larger the area under the curve, the lower is the CO. In actuality, right ventricular output is being measured: In the absence of intracardiac shunting, right and left ventricular cardiac outputs are equivalent.

The injectate solution can be either 5% dextrose in water or normal saline. A volume of 10 mL of iced or room-temperature injectate is recommended. The injection should be smooth, completed within 4 s, and timed with a specific phase of the respiratory cycle—i.e., injecting at peak inspiration or end-exhalation—rather than randomly. The measurement protocol should be consistent, and three measurements should be averaged, since a single measurement is not reliable. If for any reason the fluid bolus cannot be injected through the atrial port of the catheter (e.g., obstructed lumen), it can be administered through the venous infusion port, the right ventricular port, or the introducer side port.

Pitfalls in cardiac output measurement include injectate temperature different from the temperature used to determine the computer constant or that of the fluid being monitored by the reference probe, delivered volume less than the one entered in the computation constant, incorrect computer constant, rapid infusion of intravenous fluids during measurements, electrical noise created by electrocautery, faulty catheter lumens, improperly positioned catheter (e.g., if the catheter is in the wedge position or if the proximal lumen is above the atrium or within the introducer sheath), and presence of intracardiac shunts or tricuspid regurgitation.

A continuous thermodilution technique is now available for measuring CO. Pulmonary artery catheters are modified to locate a 10-cm thermal filament in the right ventricle during use. Without using any fluid injectate, the thermal filament continually transfers a safe level of heat directly into the blood in a random on-off fashion. The resulting temperature changes are detected at the distal thermistor located in the pulmonary artery. These data are collected by a computer, which then applies a complex formula to cross-correlate the temperature changes with the heat-input sequence to produce the familiar thermodilution curve. CO is then computed from the area under the curve by using an equation similar to the one used for standard bolus thermodilution. The continuous CO monitoring technique has been reported to be accurate and safe in critically ill patients.

Pulmonary artery catheters equipped with rapid-response thermistors and ECG electrodes have permitted the measurement of right ventricular ejection fraction at the bedside; however, the clinical utility of these systems remains unclear.

Catheter Insertion

The most commonly used pulmonary artery catheter is a 7 Fr 110-cm catheter with a distal pulmonary artery lumen, a proximal lumen 30 cm from the tip, a lumen for inflation of the balloon located at the catheter tip, and a thermistor for measurement of cardiac output by the thermodilution method (Fig. 12-3). Newer catheters may contain an additional lumen for fluid administration or for passing a pacing electrode, fiberoptic bundles for continuous measurement of the oxygen saturation of mixed venous hemoglobin (SV O₂), or a rapid-response thermistor to measure right ventricular ejection fraction.

Preparation of the electronic monitoring equipment and testing of the catheter components before insertion are essential because the displayed tracing is used to localize the position of the catheter tip during insertion. The pressure transducer must be calibrated and zeroed to the level of the left atrium. The catheter should be tested before insertion by (1) flushing the proximal and distal lumens to ensure that they are patent, (2) inflating the balloon (1.5 mL) to detect asymmetry or leaks, (3) testing the thermistor by connecting it to the cardiac output computer, and (4) shaking the catheter tip to verify that a tracing can be obtained on the oscilloscope.

Access to the central venous circulation for insertion of a pulmonary artery catheter is the same as for placement of a CVP catheter. Once an introducer sheath is in place, the pulmonary artery catheter is inserted and advanced until the tip reaches an intrathoracic vein (as evidenced by respiratory variations on the pressure tracing). The balloon is then inflated with 1.5 mL of air and the catheter advanced while the operator observes both the pressure waveform and the ECG tracing. After the right atrium is entered, the catheter is advanced through the right ventricle and into the pulmonary artery until a PAOP tracing is obtained (Fig. 12-4). Maneuvers often used to facilitate passage through the pulmonary valve include elevation of the head of the bed, turning the patient into the right lateral decubitus position, performance of the Valsalva maneuver, and increasing ventricular ejection in low-output states by the administration of inotropic drugs. To determine if the catheter is in the wedge position, the waveform needs to be inspected. The mean PAOP should be lower than the MPAP and lower than or equal to the PADP. In the wedged position, arterialized blood can be aspirated, or SV O₂ will increase to systemic arterial levels or above if an oximetric pulmonary artery catheter is used. The latter is not an absolute criterion because incomplete arterialization of the sample can occur if the tip of the pulmonary artery catheter lies wedged in a low ventilation- perfusion region.

Complications

There are risks to pulmonary artery catheterization, although they are typically infrequent and not usually life-threatening. In addition to the complications attributed to central venous cannulation, complications can occur during passage or after the catheter is in place.

The most common complication during passage of the pulmonary artery catheter is the development of dysrhythmias. They can occur in up to 50 percent of patients, but less than 1 percent of these are serious. The incidence of malignant dysrhythmias during catheterization seems to be lower when patients are in the head-up and right lateral tilt position. Transient right bundle branch block (RBBB) has been reported in 3 to 6 percent of catheterizations. Because of the rare but grave consequences of RBBB in patients with preexisting left bundle branch block, the use of standby external pacemakers and equipment for transvenous pacemaker insertion has been recommended in these patients during catheterization. Coiling, looping, or knotting in the right ventricle can occur during catheter insertion. This can be avoided if no more than 10 cm of catheter is inserted after a ventricular tracing is visualized and before a pulmonary artery tracing appears. Aberrant catheter location, such as pleural, pericardial, peritoneal, aortic, vertebral artery, renal vein, and inferior vena cava, also have been reported.

Complications that can occur after the catheter is in place include infections, thromboembolism, and rupture of the pulmonary artery. Infections from pulmonary artery catheters are directly related to the length and severity of illness. The incidence of microbial colonization of the catheter has been reported to be between 5.9 and 29.1 percent, but only 0 to 4.6 percent of catheters produce catheter-related bacteremia. Asymptomatic thrombotic complications are frequent, but symptomatic complications attributable to such thrombi are rare. Pulmonary infarction can occur due to emboli, distal migration of the pulmonary artery catheter tip, or prolonged balloon inflation occluding distal blood flow in the pulmonary artery. Pulmonary artery rupture and hemorrhage are the most serious of all the pulmonary artery catheter complications and are more likely in patients with pulmonary hypertension and in the elderly. Recurrent hemorrhage from a pulmonary artery pseudoaneurysm secondary to pulmonary artery catheter-induced perforation also can occur. Complications related to the peripheral migration of the catheter tip can be limited by continuous monitoring of the pulmonary artery tracing, avoiding prolonged balloon inflation, ensuring proximal catheter placement by review of daily x-rays, and the use of continuous heparin flush systems. Whenever the balloon is inflated, the tracing must be observed. Inflation must be stopped instantly when the waveform changes. If the catheter tip has drifted distally and is in a smaller artery, inflation with the usual 1.5 mL of air may be too much and may rupture the thin-walled pulmonary artery. Other complications that can occur after the catheter is in place include thrombocytopenia, cardiac valve injuries, catheter fracture, and balloon rupture.

In addition to the complications associated with catheter insertion and use, complications can result from delays in treatment due to time-consuming insertion problems and from inappropriate treatment based on erroneous information or erroneous data interpretation. Complications of pulmonary artery catheterization can be minimized by meticulous attention to detail and by careful evaluation of the data obtained.

Derived Hemodynamic Parameters

In addition to the information directly provided by arterial and pulmonary artery catheterization, many parameters can be calculated. The derived hemodynamic parameters (Table 12-2) aid the clinician by quantitating the relationships among heart rate, filling pressures, resistance, contractility, and cardiac output.

Cardiac output (CO) is the sum of all stroke volumes ejected in a given time. It is usually represented as the product of average stroke volume and heart rate (beats per minute),

where stroke volume is the amount of blood ejected by the heart with each contraction. The primary determinants of stroke volume are the ventricular preload, afterload, and contractility.

Preload is the passive load that establishes the initial muscle length of the cardiac fibers before contraction and therefore is not usually measured directly in critically ill patients. On the basis of the work by Otto Frank and others, Starling described the relationship between the resting fiber length of the myocardium and ventricular work. As resting fiber length increases, there is an increase in work performed on subsequent contraction. Beyond a certain point, however, further increases in fiber length will not increase external mechanical work, and work may decrease—a description of cardiac failure. The end-diastolic fiber length is proportional to the end-diastolic volume. If there is no change in ventricular compliance (the relationship between pressure and volume), LVEDV is proportional to LVEDP. Because in most clinical circumstances the PAOP provides a reliable measure of LVEDP, changes in PAOP frequently are used as an estimate of changes in left ventricular preload. In critically ill patients, however, changes in ventricular compliance may affect the relationship between LVEDP and LVEDV. Therefore, caution should be taken in interpretation of the PAOP as the sole measure of left ventricular preload. In clinical practice, judgments concerning preload adequacy are often best made empirically, by observing the responses of PAOP and indices of cardiac performance to a rapid alteration of intravascular volume.

The second determinant of stroke volume is afterload. Afterload is the sum of all the loads against which the myocardial fibers must shorten during systole, including the aortic impedance, the arterial wall resistance, the peripheral vascular resistance, the mass of blood in the aorta and great arteries, the viscosity of the blood, and the end-diastolic volume of the ventricle. In the clinical setting, the most commonly used measure of ventricular afterload is the peripheral or systemic vascular resistance (SVR). Changes in SVR usually reflect either altered blood viscosity or a change in the radius of the vascular circuit. SVR, however, does not necessarily reflect left ventricular loading conditions, since the true measure of ventricular afterload must consider the interaction of factors internal and external to the myocardium. Although it is not physiologically correct to speak of afterload in terms of SVR, it is clinically useful to relate changes in SVR to changes in ventricular afterload. Since sympathetic control of the circulation mediated by peripheral baroreceptors is designed to maintain blood pressure within relatively narrow limits, cardiac output is inversely proportional to SVR whenever this control is functioning. In the human circulatory system, however, additional factors are so often present that this relationship should not be assumed to be a substitute for direct measurements and repeated calculations.

Contractility, the final determinant of stroke volume, may be estimated in the laboratory by the maximum velocity of contraction of the cardiac muscle fibers. At the bedside, we only have inferences based on the stroke work performed by the ventricle as filling pressure (“preload”) changes. Plotting the work done by the ventricle for each beat—the left ventricular stroke work index (LVSWI) or right ventricular stroke work index (RVSWI)—against an estimate of preload and comparing that point with a normal range may be a useful means of assessing overall ventricular function (Fig. 12-5). An upward shift to the left has been interpreted as an improvement in ventricular performance. A shift downward and to the right has been considered as a declining ventricular performance. The “ventricular function curves” are influenced by changes in ventricular afterload and compliance and therefore do not reflect true contractility. At present, the method for assessing myocardial contractility most widely considered load-independent is the end-systolic pressure-volume relationship (ESPVR). The logistical difficulty of obtaining frequent ventricular volume measurements in the intensive care unit (ICU) limits the clinical usefulness of this method. Thus plotting PAOP and stroke work against normal curves is an appropriate use of data currently available in the ICU, but the underlying physiology is often better understood if it is considered in terms of the ventricular pressure-volume relation.

An appreciation of the determinants of stroke volume provides a rational approach in the management of patients with low-perfusion states. The first and most common intervention used to increase stroke volume is to increase preload by augmentation of intravascular volume. The level of PAOP that corresponds to optimal left ventricular preload can be determined only by sequentially assessing the effects of acute hemodynamic interventions on cardiac function and may vary over time in any particular patient. Fluid can be administered rapidly in predetermined increments while changes in PAOP and in the indices of cardiac performance are monitored. A major increase in PAOP during infusion suggests poor ventricular compliance, exhausted preload reserve, and increased risk of pulmonary edema with further volume loading. If the PAOP rises modestly, if indices of cardiac performance improve, and if PAOP returns to within several millimeters of mercury of the original value within 10 min of stopping the infusion, additional fluid can be given without high risk of exacerbating pulmonary venous congestion. After a brief observation period, this sequence can be repeated until the hemodynamic parameters are adequate or the PAOP shows an unacceptable rise. If tissue perfusion remains inadequate after volume optimization, augmentation of stroke volume may be accomplished by increasing myocardial contractility with inotropic drugs and/or decreasing ventricular afterload with vasodilators. Some authors have reported reduced complications and improved survival in perioperative patients when hemodynamic therapy was aimed at augmenting rather than simply normalizing hemodynamic and oxygen-transport parameters. Recent studies, however, found no advantage to the use of supranormal target values in a general population of critically ill patients.

RESPIRATORY MONITORING

Monitoring ventilation and gas exchange in critically ill surgical patients is of particular importance in deciding if mechanical ventilation is indicated, assessing response to therapy, optimizing ventilator management, and deciding if a weaning trial is indicated. In addition, gas monitoring permits an assessment of the adequacy of oxygen transport and calculation of derived parameters.

Ventilation Monitoring

Lung Volumes

Several lung volume measurements are useful for monitoring ventilatory function in the operating room and ICU. These include tidal volume, vital capacity, minute volume, and dead space.

Tidal volume (V_T) is defined as the volume of air moved in or out of the lungs in any single breath. If the tidal volume is depressed, the patient may have difficulty in both oxygenation and ventilation. Rapid, shallow breathing, as reflected by the respiratory frequency (f) to tidal volume ratio ($f/V_T > 100$), is an accurate predictor of failure, and its absence ($f/V_T < 80$) is an accurate predictor of success, in weaning patients from mechanical ventilation. V_T can be measured at the bedside using a hand-held spirometer (Wright respirometer). Because moisture impairs its performance, the instrument is most appropriate for intermittent monitoring. Continuous V_T monitoring is facilitated by the presence of pneumotachometers in the breathing circuit of modern ventilators. In order to obtain accurate V_T measurements, the spirometer must be located between the ventilator Y piece and the endotracheal tube. If the spirometer is instead positioned on the expiratory limb of the breathing circuit, the entire V_T delivered by a ventilator, not that actually received by the patient, is measured. Under conditions of decreased lung compliance or increased airway resistance, the higher peak inspiratory pressure (PIP) would result in an increase of gas volume compressed in the breathing circuit, with correspondingly less delivered to the patient. The product of $PIP\text{ (cmH}_2\text{O)} \times 5\text{ (mL/cmH}_2\text{O)}$ provides an estimate of the compression volume of most circuits.

Vital capacity (VC) is defined as the maximal expiration following a maximal inspiration. It can be readily measured at the bedside in a manner similar to the one used for V_T . The VC is reduced in diseases involving the respiratory muscles or their neural pathways, in obstructive and restrictive ventilatory impairment, and in patients who fail to cooperate fully. VC is normally 65 to 75 mL/kg, and a value of 10 mL/kg or greater is commonly considered a favorable predictor of weaning outcome. This value, however, is quite dependent on patient cooperation, and its predictive power is rather poor.

Minute volume (or total ventilation) ($V_{(dot)E}$) is the total volume of air leaving the lung each minute (product of V_T and f). Many ventilators display $V_{(dot)E}$, or it can be measured with a Wright spirometer. An increase in the minute volume required to maintain a normal arterial blood carbon dioxide tension (Pa_{CO_2}) suggests an increased dead space relative to V_T or an abnormally high carbon dioxide (CO_2) production. A resting $V_{(dot)E}$ of less than 10 L and the ability to double the resting $V_{(dot)E}$ on command have been associated with successful weaning from mechanical ventilation.

The physiologic (or effective, or total) *dead space* (V_D) is the portion of tidal volume that does not participate in gas exchange. Physiologic dead space may be divided into two components: the volume of gas within the conducting airways (the anatomic dead space) and the volume of gas within unperfused alveoli (the alveolar dead space). The ratio of physiologic dead space to tidal volume (V_D/V_T) is calculated from the Enghoff equation (modified from the Bohr equation) as follows:

$$V_D / V_T = Pa_{CO_2} - PE_{CO_2} / Pa_{CO_2}$$

where PE_{CO_2} is the mean partial pressure of exhaled CO_2 in the total exhaled volume of gas after thorough mixing. Normally, exhaled gas is collected in a bag over 3 min and the PE_{CO_2} is measured from the bag. The PE_{CO_2} should not be confused with P_{etCO_2} , the partial pressure of end-tidal CO_2 (discussed later). The V_D/V_T ratio provides a useful expression of the efficiency of ventilation. In healthy subjects, the ratio is between 0.33 and 0.45. The V_D/V_T ratio is increased in a number of disease states associated with regions of the lung possessing high ventilation-perfusion ratios, such as adult respiratory distress syndrome, emphysema, pulmonary embolism, shock with low cardiac output, and the employment of positive-pressure ventilation with high V_T or excessive (more than is needed) PEEP. Patients whose V_D/V_T exceeds 0.6 are usually not weanable from ventilatory support.

By measuring $V_{(dot)E}$ and calculating V_D/V_T , the alveolar (or effective) ventilation ($V_{(dot)A}$) also may be calculated:

$$V_{(dot)A} = V_{(dot)E} - (V_{(dot)E} \times V_D/V_T)$$

Pulmonary Mechanics

Various respiratory mechanical parameters also can be monitored in the operating room and ICU. These include maximal inspiratory pressure, static compliance, dynamic characteristic, and work of breathing.

Inspiratory force is measured as the maximal pressure below atmospheric that a patient can exert against an occluded airway. The measurement requires a connector to an endotracheal or tracheostomy tube and a manometer capable of registering negative pressure. A maximal inspiratory pressure (Pi_{max}) value more negative than -20 to 25 cmH₂O has been used as one of the clinical parameters to confirm recovery from neuromuscular block after general anesthesia. Pi_{max} values more negative than -30 cmH₂O have been used to predict successful weaning from mechanical ventilation. Studies have found that Pi_{max} has limited power in predicting weaning outcome, especially in patients receiving prolonged mechanical ventilation. These findings may be due in part to the fact that Pi_{max} assesses only the strength of the respiratory muscle pump without taking into account the demands placed on it.

Compliance, a measure of the elastic properties of the lung and chest wall, is expressed as a change in volume divided by a change in pressure (DV/DP). In patients receiving mechanical ventilation, a rough measure of total thoracic compliance (both the lungs and chest wall) can be obtained by dividing the delivered V_T by the inflation pressure displayed on the ventilator gauge during conditions of zero gas flow. These can be achieved by using the “inspiratory hold” option on the ventilator, during which period the airway pressure falls to a plateau. If the patient is receiving PEEP, this must be first subtracted from the plateau pressure before calculating static thoracic compliance, that is,

$$\text{Static compliance} = \text{volume delivered} / \text{plateau pressure} - \text{PEEP}$$

The usual range for adult patients receiving mechanical ventilation is 60 to 100 mL/cmH₂O. Decreased values are observed with disorders of the thoracic cage or a reduction in the number of functioning lung units (resection, bronchial intubation, pneumothorax, pneumonia, atelectasis, or pulmonary edema). When the static compliance is less than 25 mL/cmH₂O, as in severe respiratory failure, difficulties in weaning are common because of the increased work of breathing (see below).

The dynamic characteristic is calculated by dividing the volume delivered by the peak (rather than the plateau) airway pressure minus PEEP. It is not correct to call this value *dynamic compliance* because it is actually an impedance measurement and includes compliance and resistance components. The dynamic characteristic is normally about 50 to 80 mL/cmH₂O. It may be decreased by disorders of the airways, lung parenchyma, or chest wall; if it decreases to a greater extent than the static compliance, it suggests an increase in airway resistance (e.g., bronchospasm, mucous plugging, kinking of the endotracheal tube) or an excessive flow rate.

Work of breathing, which relates to the product of the change in pressure and volume, is a measure of the process of overcoming the elastic and frictional forces of the lung and chest wall. The work of breathing in the critically ill patient who requires ventilatory support (WOB_{PI}) can be divided into three components: normal physiologic work (WOB_{Phys}), work to overcome the pathophysiologic changes in the lung and chest wall (WOB_{Dis}), and work to overcome the imposed work of breathing (WOB_{Imp}) created by our methods of ventilatory support. The sum is total work. Physiologic work of breathing consists of three elements: elastic work, flow-resistive work, and inertial work. Elastic work is the work necessary to overcome the elastic forces of the lung and is inversely proportional to the compliance of the lung. If compliance becomes diminished, the work of breathing increases dramatically. The second element of physiologic work is flow-resistive work, or the work that is needed to overcome the resistance of the airways and parenchymal tissues. This may increase the pressure change necessary to inhale the same tidal volume but also adds another component of work during expiration, that necessary to expel the gas from the lungs through the narrow airways. The third component of physiologic work is the inertial work to overcome the tendency of gas volume to remain at rest. This element is negligible in comparison with the elastic and flow- resistive work. When a patient develops respiratory failure, in addition to the normal physiologic work, the patient must overcome the increased work of breathing associated with the disease. This is clinically manifest as a change from a relatively large tidal volume at a slow rate to a small tidal volume at a rapid rate. Finally, the patient must do additional work to breath spontaneously against a breathing apparatus that consists of the ventilator itself, demand valve, tubing, exhalation valves, and most important, the endotracheal tube. Banner and associates showed that the endotracheal tube acts as a resistor in series in the breathing apparatus, thereby causing an increase in work of breathing. Imposed work has been shown to exceed physiologic work of breathing by a factor of 6 under conditions of spontaneous breathing through a narrow-internal-diameter endotracheal tube at a high inspiratory flow rate demand during continuous positive airway pressure. Poor demand system sensitivity, ventilator dyssynchrony, malfunctioning demand valves, and inadequate inspiratory flows are also contributing factors.

The goal of ventilatory support is to carefully titrate the ventilator's contribution to minute ventilation so that the patient's effort remains a nonfatiguing work load. Failure to do so by supplying either too much or too little ventilatory support may result in unsuccessful weaning trials and increase the duration of mechanical ventilation. Normal range for WOB_{PI} is 0.3 to 0.6J/L.

Microprocessor-based respiratory monitors such as the CP-100 Pulmonary Monitor (Biocore Monitoring Systems, Irvine, CA) measure many mechanical ventilation and respiratory muscle parameters, including compliance, airway resistance, strength and endurance, and both patient and ventilator work of breathing. Physiologic data are accrued from a miniature pneumotachograph and airway pressure sensor positioned between the Y piece of the breathing circuit tubing and the endotracheal tube. A catheter with a distally annealed balloon is positioned in the distal esophagus to measure changes in intraesophageal pressure as an estimate of changes in intrathoracic pressure. We have employed this pulmonary monitor to evaluate unexplained tachypnea or respiratory distress, to guide endurance and strength reconditioning, and to avoid iatrogenic ventilator dependency caused by inappropriate ventilator settings in complex long-term ventilated patients.

Incorporation of work-of-breathing analysis into our preextubation trial allowed successful extubation in 97 (of 589) patients who remained on mechanical ventilation because of tachypnea (respiratory rate between 32 and 52 breaths per minute) secondary to excessive imposed work of breathing. Earlier extubation in this group of patients resulted in a projected net savings in excess of \$292,000. Incorporating work-of- breathing strategies also decreased the duration of ventilation in our trauma ICU from 8.8 to 4.2 days.

Gas Monitoring

Blood-Gas Analysis

Blood-gas measurements provide information about the efficiency of gas exchange, the adequacy of alveolar ventilation, and the acid-base status. Blood gas values are usually reported in terms of directly measured partial pressures (P_{O2} or P_{CO2}) and calculated hemoglobin oxygen saturations (S_{O2}). Calculated S_{O2} values are derived from the measured partial pressure and a nomogram of the oxyhemoglobin dissociation curve usually corrected for blood temperature, pH, and perhaps other factors. Because these assumptions may not be accurate in critically ill patients, actual measurements of S_{O2} by cooximetry are preferred. S_{O2} also can be measured continuously by using pulse oximeters or pulmonary artery catheters that incorporate oximetric fibers (see below).

Arterial blood gas tensions are determined by the composition of the alveolar gas and the efficiency of gas transfer between the alveoli and pulmonary capillary blood. Alveolar gas tensions depend on the mixture of inspired gas, ventilation, and blood flow in the lungs; the matching of ventilation and perfusion; and the composition of mixed venous blood gases. Pathophysiologic causes of arterial hypoxemia include ventilation- perfusion inequality or venous admixture from regional alveolar hypoventilation, true intrapulmonary or intracardiac shunt, and decreased mixed venous oxygen content. Although diffusion abnormalities may lead to hypoxemia if pulmonary end-capillary blood fails to equilibrate fully with alveolar gas, such conditions are uncommon. A decreased cardiac output in the presence of a constant oxygen consumption, an increased oxygen consumption in the presence of a constant CO, and a decreased CO and an increased oxygen consumption must all result in a lower mixed venous oxygen content and therefore also can produce arterial hypoxemia. Failure to recognize this nonpulmonary cause of hypoxemia may cause a clinician to falsely attribute a decreasing arterial blood oxygen tension (Pa_{O2}) to deteriorating pulmonary function. Thus pulmonary and cardiac function must be assessed to evaluate any given set of arterial blood gases accurately. A decreasing Pa_{O2} without a change in Pa_{CO2} suggests that blood oxygenation is deteriorating despite constant alveolar ventilation. In the acutely ill patient, this finding is usually attributable to ventilation-perfusion imbalance or intrapulmonary shunting. An important feature of shunting is that as it increases, supplemental oxygen has progressively less effect on Pa_{O2} because shunted blood bypasses ventilated alveoli. Intrapulmonary shunting usually does not result in elevation of the Pa_{CO2} because the central chemoreceptors sense any rise in Pa_{CO2} and respond by increasing ventilation.

The relation of P_{O2} to S_{O2} is described by the oxyhemoglobin dissociation curve (Fig. 12-6). The flat upper portion of the dissociation curve means that even if the P_{O2} in alveolar gas decreases somewhat, loading of oxygen will be little affected. The steep lower part of the curve means that the peripheral tissues can withdraw large amounts of oxygen for only a small decrease in capillary P_{O2}. The curve shifts as the affinity of hemoglobin for oxygen changes. A shift to the right (decreased affinity for oxygen) helps release oxygen into the tissue. A shift to the left (increased affinity for oxygen) causes less oxygen to be available to tissue. The curve can be shifted to the right by *increased* erythrocyte 2,3-diphosphoglycerate concentration, temperature, P_{CO2}, and concentration of hydrogen ion (decreased pH). Opposite changes shift it to the left. Other conditions such as carboxyhemoglobinemia and methemoglobinemia also can shift the oxyhemoglobin dissociation curve to the left and therefore interfere with peripheral oxygen unloading. The position of the oxyhemoglobin dissociation curve is defined by the P₅₀, that is, the P_{O2} at which hemoglobin is 50 percent saturated. Normal hemoglobin has a P₅₀ of 26.5 mmHg. When it is greater than this value, the curve is shifted to the right; when it is lower, the curve is shifted to the left. Despite a considerable amount of information, there is little evidence that shifts of the oxyhemoglobin dissociation curve are clinically significant in the majority of patients. Individuals with limited circulatory reserve, however, who cannot augment oxygen delivery by the usual compensatory mechanisms of increased cardiac output and organ blood flow, may develop local tissue hypoxia when an increased hemoglobin-oxygen affinity state (i.e., alkalemia) exists.

The Pa_{CO2} directly reflects the adequacy with which alveolar ventilation meets metabolic demands for CO₂ excretion. The relationship between Pa_{CO2}, CO₂ production (V_{(dot) CO2}), and alveolar ventilation (V_{(dot) A}) in normal lungs is given by the equation

$$Pa_{CO2} = V_{(dot) CO2} / V_{(dot) A} \times K$$

where K is a constant. In diseased lungs, the denominator V_{(dot) A} in this equation is less than the ventilation going to the alveoli because of alveolar dead space, that is, unperfused alveoli or those with high ventilation- perfusion ratios. For this reason, the denominator is sometimes referred to as the *effective alveolar ventilation*.

An increased Pa_{CO2} (hypercapnia) reflects the failure of the ventilatory system to eliminate the CO₂ produced during metabolism. This “ventilatory failure” is traditionally described as respiratory acidosis. Hypercapnia can occur because of hypoventilation (i.e., CNS depression), increased CO₂ production (e.g., hyperthermia, hyperthyroidism), or increased physiologic dead space resulting in inadequate alveolar ventilation. The mechanisms of hypocapnia are the reverse of those which produce hypercapnia, the most common being hyperventilation (respiratory alkalosis).

The oxygen tension of mixed venous blood (PV_{O2}) and the oxygen saturation of mixed venous hemoglobin (SV_{O2}) provide valuable diagnostic information and are necessary for the calculation of various parameters, such as arteriovenous oxygen content difference, intrapulmonary shunt, and oxygen consumption. Mixed venous blood is the mixture of all blood that has traversed the capillary beds capable of extracting oxygen. This venous effluent is thoroughly mixed, so its oxygen content is a flow- weighted representation of all the end-capillary contents of the body and as such will reflect the *total body balance* between oxygen delivery and oxygen consumption of *perfused* tissues. The mixed venous oxygen content (and therefore SV_{O2}) is determined by the variables in the Fick equation (see section on Continuous Mixed Venous Oximetry). The PV_{O2} is determined by the same factors and the position of the oxyhemoglobin dissociation curve.

In critically ill patients, sampling of mixed venous blood can be performed accurately only in the pulmonary artery. Sampling technique is important; blood should be withdrawn from the most proximal pulmonary artery location possible and at a very slow rate. A fast rate of blood withdrawal or a malpositioned catheter (distal migration or wedging) may cause a falsely elevated PV_{O2} and SV_{O2}. This is due to “contamination” of the mixed venous blood with arterialized pulmonary capillary blood and should be suspected if the CO₂ tension of mixed venous blood (PV_{CO2}) is equal to or lower than a simultaneously determined Pa_{CO2}.

Proper sample handling before arterial or venous blood-gas analysis is a prerequisite to accurate blood gas measurement. The two principal requirements are that the sample be obtained under strict anaerobic conditions and immediately placed on ice until analyzed. Because room air has a partial pressure of oxygen (P_{O2}) of about 150 mmHg and a partial pressure of carbon dioxide (P_{CO2}) of essentially zero, equilibration of a blood sample with air bubbles may significantly alter the results. Placement of the sample on ice is necessary to reduce the metabolic rate of the red blood cells and prevent continued oxygen consumption and CO₂ production if analysis is to be delayed beyond 15 to 20 min. The addition of excessive heparin also will alter the results, and therefore, after aspiration, the heparin in the syringe should be expelled, leaving only the heparin wetting the barrel. Alternatively, a commercially available syringe containing dry heparin can be used.

Because blood gas values can change rapidly in critically ill patients, intermittent sampling for blood-gas analysis might miss significant changes. Advances in fiberoptic and microprocessor technology have been combined with fluorescent dye chemistry to develop miniaturized intravascular gas sensors that permit continuous measurement of pH, P_{O2}, and P_{CO2}. The performance of fiberoptic continuous intraarterial blood gas monitors is comparable with that of modern blood gas analyzers. Further studies are needed, however, to determine the clinical utility of this new methodology.

Parameters Derived from Blood-Gas Analysis

Just as the derived hemodynamic parameters can be used to evaluate the choice and effects of hemodynamic interventions, parameters derived from blood-gas analysis (Table 12-3) yield information about the adequacy of cardiopulmonary function in meeting the tissue demands for oxygen.

The oxygen content of the blood is equal to the amount of oxygen bound to hemoglobin plus the amount dissolved in plasma. The amount of bound oxygen is directly related to the concentration of hemoglobin and to how saturated this hemoglobin is with oxygen (i.e., Sa_{O_2} or SV_{O_2}). The amount of oxygen dissolved in plasma depends on the oxygen tension (i.e., Pa_{O_2} or PV_{O_2}). Oxygen delivery ($\dot{V}_{O_2}$) is the volume of oxygen delivered from the heart each minute and is calculated as the product of cardiac output and arterial oxygen content (Ca_{O_2}). Oxygen consumption ($\dot{V}_{(dot) O_2}$) is the amount of oxygen that diffuses from the capillaries into all tissues and can be calculated according to the Fick principle as the product of CO and arteriovenous oxygen content difference [$C(a - PV)_{O_2}$]. If this equation is rearranged, the arteriovenous oxygen content difference relates oxygen consumption and CO ($\dot{V}_{(dot) O_2}/CO$). An increase in the arteriovenous oxygen content difference indicates that either consumption is too high or flow is too low. Finally, the oxygen utilization coefficient or extraction ratio (O_2UC), relates oxygen consumption and oxygen delivery ($\dot{V}_{(dot) O_2}/\dot{V}_{O_2}$). This parameter has been used in many ICUs to evaluate the adequacy of oxygen transport.

The adequacy of oxygen transport also must be assessed in relation to oxygen demand, which is the amount of oxygen *required* by the body tissues to use aerobic metabolism. Although oxygen demand cannot be measured clinically, the relative balance between consumption and demand is best indicated by the presence of excess lactate in the blood. Lactic acidosis means that demand exceeds consumption and anaerobic metabolism is present.

Although precise numerical end points cannot be defined, the parameters already listed provide a framework for testing a clinical hypothesis: If oxygen delivery or consumption is low, if utilization is high, or if lactic acidosis is present, arterial oxygen content might be augmented by increasing hemoglobin concentration or oxygen saturation, or cardiac output might be increased by manipulation of preload, afterload, or contractility. A response might be considered beneficial if oxygen consumption increases, if utilization returns to the normal range, or if lactic acidosis resolves.

Physiologic right-to-left shunt or venous admixture ($_{sp/t}$) estimates the fraction of total blood flow reaching the left side of the circulation without participating in gas exchange. Shunt may occur (uncommonly) in adults via intracardiac shunts. More commonly, increased venous admixture in critically ill patients is due to alterations in the balance of pulmonary ventilation and perfusion (lung areas that are perfused but not ventilated). Before calculating venous admixture, it is necessary to calculate arterial, mixed venous, and pulmonary end-capillary oxygen contents. The latter can be calculated by using the *alveolar* oxygen tension (Pa_{O_2}) to estimate pulmonary end-capillary oxygen tension and the oxyhemoglobin dissociation curve to estimate pulmonary end-capillary hemoglobin saturation (assume 100% if $Pa_{O_2} > 150$ mmHg) (see Table 12-3).

Other indices, such as the alveolar-arterial oxygen partial pressure difference ($Pa_{O_2} - Pa_{O_2}$) and the arterial-to-alveolar oxygen tension ratio (Pa_{O_2}/Pa_{O_2}), have been suggested for evaluating the efficiency of gas exchange. These oxygen tension–based indices, however, are inaccurate in predicting efficiency of gas exchange. The relationship between physiologic shunt and the oxygen tension–based indices is nonlinear and substantially influenced by changes in inspired oxygen concentration and arteriovenous oxygen content difference.

The collection of the measured and derived cardiopulmonary parameters has been called the *cardiopulmonary profile*. Normal values can be seen in Tables 12-2 and 12-3. The measured and derived data can be used to formulate a plan of interventions designed to improve oxygen delivery relative to myocardial and systemic needs. This analysis is a dynamic process that evolves as new data are obtained and response to therapy is incorporated. The process of generating a cardiopulmonary profile has been greatly simplified by the use of programmable calculators and microcomputers.

Capnography

Capnography is the graphic display of CO₂ concentration as a waveform. It should not be confused with *capnometry*, which refers to only the numerical presentation of the concentration without a waveform. Capnography includes capnometry when the capnographic display is calibrated.

Currently available systems for CO₂ analysis include infrared spectroscopy, mass spectrometry, and Raman scattering. In addition, a disposable, noninvasive, and inexpensive colorimetric device (Easy Cap; Nellcor Puritan Bennett, Inc., Pleasanton, CA) is available. This device permits a semiquantitative measurement of the end-tidal CO₂ concentration when it is attached between an endotracheal tube and a resuscitation bag.

In the majority of stand-alone capnographs, the CO₂ concentration is measured by infrared spectroscopy. A beam of infrared light is passed through the sampled gas. CO₂ molecules in the light path absorb some of the infrared energy. The capnograph compares the amount of infrared light absorbed by the patient gas in the sample cell with the amount absorbed either by gas in a reference cell or by the sample cell during a time of known zero-gas concentration. The capnograph then displays the instantaneous CO₂ concentration.

Gas for analysis of CO₂ may be aspirated from the airway (sidestream capnography) or may be analyzed as it flows through a sensor placed in the airway (mainstream capnography). Sidestream analyzers offer advantages in that gas is sampled close to the patient's mouth with the use of an inexpensive, lightweight connector, and they can be used in nonintubated patients. The major disadvantage of these systems is that analysis is delayed because gas is routed through a capillary tube to the capnograph.

Mainstream analyzers generate a capnogram practically instantaneously because the gas is analyzed as it passes through a sampling cuvette. The major disadvantage of these systems is the weight of the sensor and sampling cuvette. Because the sensor itself is a sophisticated instrument, it needs to be treated carefully: It is fragile, and replacements are expensive. The volume of the cuvette adds dead space to the system.

Normally, there is a fairly predictable relationship between the peak exhaled or end-tidal CO₂ (Pet_{CO_2}) and the Pa_{CO_2} . In healthy subjects with normal lungs, the Pa_{CO_2} is 4 to 6 mmHg higher than the Pet_{CO_2} . Patients with chronic obstructive lung disease and other derangements associated with increased dead space (see section on Lung Volumes) have an increased arterial to end-tidal CO₂ gradient [$P(a - et)_{CO_2}$]. This difference occurs because the exhaled gas from the alveolar dead space, which contains little or no CO₂, dilutes the CO₂-containing gas from the normally ventilated and perfused alveoli.

Measurement of Pet_{CO_2} and $P(a - et)_{CO_2}$ provides insight into several normal and pathologic processes. Pet_{CO_2} measurement is at present perhaps one of the most reliable means of determining proper endotracheal tube placement. Esophageal intubation may produce one or a few breaths containing CO₂ during expiration, but because there is no CO₂ in the stomach cavity, Pet_{CO_2} rapidly decreases to zero.

Pet_{CO_2} has been found to correlate with cardiac output and coronary perfusion pressure during cardiopulmonary resuscitation (CPR) and with successful resuscitation from and survival after cardiac arrest. Because circulatory arrest creates total dead space, if ventilation is continued, Pet_{CO_2} disappears. An increase in Pet_{CO_2} provides an immediate bedside validation of the efficacy of CPR, and if the increase is abrupt, it provides the earliest evidence of successful resuscitation. The use of Pet_{CO_2} to monitor resuscitation is predicated on maintaining a constant minute ventilation so that changes in Pet_{CO_2} result from changes in lung perfusion (and therefore cardiac output) and not ventilation.

Pet_{CO_2} monitoring is extremely useful as a diagnostic tool in several situations unique to the operating room. These include the detection of air emboli during neurosurgical procedures requiring the sitting position, the detection of increased CO₂ production in malignant hyperthermia, and the detection of disconnection or malfunction of the anesthesia breathing circuit.

In the ICU environment, Pet_{CO_2} monitoring also can be used as a ventilator disconnect alarm as well as a system to determine ventilator malfunction. Measurement of Pet_{CO_2} has been proposed as a substitute for arterial blood gas sampling during mechanical ventilation adjustment and weaning in critically ill patients. Pet_{CO_2} trends in these patients, however, are often misleading because the $P(a - et)_{CO_2}$ varies greatly in a single individual.

The $P(a - et)_{CO_2}$ is primarily a reflection of dead-space ventilation, and its size can serve as a gauge of physiologic aberration. Factors related to the instrumentation and the technique used, however, also may contribute to the $P(a - et)_{CO_2}$. For example, aspiration of room air through a loose connection or break in the circuit or sampling tube, a leak around the cuff, or aspiration of fresh gases will dilute the exhaled CO₂ and result in an increased $P(a - et)_{CO_2}$ and an altered waveform.

Analysis of the CO₂ waveform can provide valuable information. A detailed review of waveform analysis is outside the scope of this chapter but can be found elsewhere (Gravenstein and associates).

Pulse Oximetry

Pulse oximetry provides a reliable, real-time estimation of arterial hemoglobin oxygen saturation. This noninvasive monitoring technique has gained clinical acceptance in the operating room, recovery room, and ICU.

Pulse oximeters estimate arterial hemoglobin saturation by measuring the absorbance of light transmitted through well-perfused tissue, such as the finger or ear. The light absorbance is measured at two wavelengths—660 (red) and 940 nm (infrared)—to distinguish between two species of hemoglobin—oxyhemoglobin and deoxyhemoglobin. Oxyhemoglobin absorbs less red light than deoxyhemoglobin, accounting for its red color; at infrared wavelengths, the opposite is true (Fig. 12-7). Light absorbances at both wavelengths have two components: the pulsatile (or ac) component, which is attributed to the pulsating arterial blood, and the baseline (or dc) component, which represents the absorbances of the tissue bed, including venous blood, capillary blood, and nonpulsatile arterial blood (Fig. 12-8). The pulse oximeter first determines the ac components of absorbance at each wavelength and divides this by the corresponding dc component to obtain a pulse-added absorbance that is independent of the incident light intensity. It then calculates the ratio (*R*) of these pulse-added absorbances:

$$R = AC_{660}/DC_{660} / AC_{940}/DC_{940}$$

The ratio of the pulse-added absorbances at the two wavelengths is used to generate the oximeter's estimate of arterial saturation (Sp_{O_2}). The relationship between this ratio and Sp_{O_2} is empirical. The algorithm was created by measuring pulse-added absorbances in healthy, awake volunteers breathing hypoxic gas mixtures. These absorbances were then correlated with actual Sa_{O_2} as determined by a laboratory co-oximeter.

In practice, pulse oximeters use two light-emitting diodes (LEDs) and one photodiode as transmitting and sensing transducers, usually placed on opposite sides of a digit. The microprocessor of the pulse oximeter is programmed to distinguish arterial pulse waveforms; minimize the effects of ambient light, patient motion, and electrocautery; and vary the intensity of transmitted light required to obtain the waveforms.

In most studies of pulse oximetry accuracy, data have been collected only when the pulse oximeter heart rate equaled the ECG heart rate. It has been assumed that this is a necessary condition for accuracy because it implies that the pulse oximeter is detecting pulses produced by heartbeats.

Most manufacturers claim that their pulse oximeters are accurate within ±2 percent (SD) from 70 to 100 percent saturation. Although pulse oximetry may provide erroneous measurements when Sa_{o2} is less than 70 percent, these values occur quite rarely (or *should* occur quite rarely) in patients, because Pa_{o2} would be less than 40 mmHg. Sa_{o2} values in the range of perhaps 70 to 95 percent will reflect changes in Pa_{o2}; it is in this range that pulse oximetry finds great value in monitoring cardiorespiratory disease and directing therapy. High levels of saturation give no information about Pa_{o2}. Because of the sigmoid shape of the oxyhemoglobin dissociation curve (see Fig. 12-6), Sa_{o2} may not decrease despite a significant deterioration in pulmonary gas exchange, i.e., if Pa_{o2} fell from 200 to 100 mmHg. Since delivery of oxygen to the tissues is proportional to Sa_{o2}, however, pulse oximeters will detect changes before tissue oxygenation is impaired.

Various physiologic and environmental factors interfere with the accuracy of pulse oximetry. These include decreased amplitude of peripheral pulses (hypovolemia, hypotension, hypothermia, vasoconstrictor infusions), motion artifact, electrosurgical interference, backscatter from ambient light, and dyshemoglobinemias. Predictors of pulse oximeter data failure in the operating room include ASA physical status 3, 4, or 5; cardiac, vascular, and orthopedic surgery; hypotension; hypertension; and duration of procedure.

The pulse oximeter can only distinguish oxyhemoglobin and deoxyhemoglobin. If other hemoglobin species are present, an error is introduced. Laboratory co-oximeters, on the other hand, generally use more than two wavelengths and often can quantify other hemoglobin species directly. When dyshemoglobins such as carboxyhemoglobin and methemoglobin can be measured, it becomes meaningful to distinguish between functional saturation [100 × oxyhemoglobin/(oxyhemoglobin + deoxyhemoglobin)] and fractional saturation [100 × oxyhemoglobin/(oxyhemoglobin + deoxyhemoglobin + carboxyhemoglobin + methemoglobin)]. Barker and colleagues have shown that in the presence of elevated carboxyhemoglobin or methemoglobin levels, Sp_{o2} *overestimates* fractional saturation at all saturation values. Carboxyhemoglobinemia may occur in heavy smokers or in patients who suffer carbon monoxide inhalation. Methemoglobinemia may be induced by a large number of drugs, including local anesthetics (prilocaine, benzocaine), nitroglycerin, phenacetin, phenytoin, Pyridium, and sulfonamides.

Intravenously administered dyes, particularly methylene blue and indocyanine green, can temporarily induce artifactually low saturation readings. Deeply pigmented skin and opaque nail polish coatings may significantly decrease light transmission, rendering oximeters inoperative. The presence of fetal hemoglobin, hyperbilirubinemia, or moderate anemia (with hematocrits as low as 15 percent) does not affect the accuracy of pulse oximeters. Despite its limitations, pulse oximetry is generally acknowledged as one of the most significant advances in clinical monitoring.

Continuous Mixed Venous Oximetry

Measurement of the oxygen saturation of mixed venous hemoglobin (SV_{o2}) is helpful in the assessment of the oxygen supply-demand relationship in critically ill patients. The use of improved fiberoptic oximetry systems in conventional pulmonary artery catheters has permitted continuous monitoring of SV_{o2} and made bedside monitoring of this relationship practical.

SV_{o2} can be derived from the Fick equation (see Table 12-3) that relates oxygen consumption, cardiac output, and arteriovenous oxygen content difference. If the small quantity of physically dissolved oxygen in the blood is considered negligible, solving the Fick equation for SV_{o2} yields

SV_{o2} = Sa_{o2} - V_{(dot) O2} / CO × Hb × 1.39 × 10

Therefore, the determinants of SV_{o2} include the principal components of oxygen delivery [CO, hemoglobin (Hb), and Sa_{o2}] and oxygen consumption. There is a poor correlation between SV_{o2} and any single component of the equation (V_{(dot) O2}, CO, Sa_{o2}, Hb), as would be expected, because there are four separate determinants. There is, however, a good correlation between SV_{o2} and all these components acting at once.

With the preceding formula in mind, it is easy to understand that the SV_{o2} will decrease when there is an imbalance between oxygen consumption and delivery caused by an increase in V_{(dot) O2} or a decrease in CO, Hb, or Sa_{o2}. SV_{o2} will increase when the imbalance is due to changes in the opposite direction.

The normal range for SV_{o2} in healthy subjects is 0.65 to 0.80, with an average value of 0.75 corresponding to a PV_{o2} of 40 mmHg at a normal pH of 7.4. A rapid or prolonged fall from the normal range is indicative of a significant deterioration in the patient's clinical condition. Values below the normal range may be associated with increased oxygen consumption due to fever, shivering, seizures, exercise, and agitation or associated with decreased oxygen delivery due to low cardiac output, anemia, or arterial hemoglobin desaturation. Values of about 0.53 correspond to a PV_{o2} of about 28 mmHg; values at or below this level often have been associated with anaerobic metabolism, lactic acidosis, and death. Astiz and colleagues, however, were unable to identify the critical level of SV_{o2} associated with lactic acidosis in patients with sepsis or acute myocardial infarction.

Values above the normal range indicate an increase in oxygen delivery relative to consumption and are associated with the hyperdynamic phase of sepsis, cirrhosis, peripheral left-to-right shunting, general anesthesia (when V_{(dot) O2} is low), cellular poisoning such as cyanide toxicity (rare), marked arterial hyperoxia, or a technical malfunction of the system (e.g., wedged catheter). Normal or high SV_{o2} values do not *ensure* that the oxygen supply-demand balance is satisfactory because accurate interpretation assumes intact and consistent vasoregulation, which is not the case in some disease states (e.g., sepsis).

Pulmonary artery catheter oximetry differs from pulse oximetry in several ways. First, the pulmonary artery catheter measures *reflected* rather than *transmitted* light. Second, being immersed in blood, the pulmonary artery catheter has no need for the pulse-added signal analysis used by the pulse oximeter. Finally, one of these catheters, the Opticath (Abbott Critical Care Systems, Mountain View, CA) uses three wavelengths (670, 700, and 800 nm) rather than the two wavelengths (660 and 940 nm) employed by most pulse oximeters.

The Oximetrix 3 System used by the Opticath has three light-emitting diodes (LEDs) contained in an optical module that provide the light sources for the three selected wavelengths. Light from each of these diodes is sequentially transmitted through a single optical fiber to illuminate the blood flowing past the catheter tip. This illuminating light is absorbed, refracted, and reflected depending on the color and, therefore, oxyhemoglobin concentration of the blood. The reflected light is collected by a second fiber and returned through the catheter to a photodetector in the optical module (Fig. 12-9). Using the relative intensities of the signals representing the light levels at the various wavelengths, a computer calculates the oxygen saturation, and the average for the preceding 5 s is displayed.

The saturation measured by catheter oximetry correlates well with values obtained from a laboratory co-oximeter. The most common sources of error when measuring Sv_{o2} are incorrect calibration and catheter malposition. Carboxyhemoglobinemia, methemoglobinemia, and intravenous administration of methylene blue also can affect the measurements.

Continuous SV_{o2} monitoring serves three major functions. First, it serves as an indicator of the adequacy of the oxygen supply-demand balance of perfused tissues. In clinically stable patients, a normal and stable SV_{o2} may be considered an additional assurance of cardiopulmonary stability. Further assessments of cardiac output and arterial and mixed venous blood gas analyses are not necessary. Second, continuously measured SV_{o2} may function as an early warning signal of untoward events. In this situation, although an alert has been given, the cause of the change in SV_{o2} is not necessarily clear because the change in SV_{o2} is sensitive but not specific. It may be necessary to measure cardiac output, Sa_{o2}, and Hb in this setting to identify the etiology of the SV_{o2} change. Third, continuously monitored SV_{o2} may improve the efficiency of the delivery of critical care by providing immediate feedback as to the effectiveness of therapeutic interventions aimed at improving oxygen transport balance.

Finally, an important application of continuous venous oximetry must be one of cost containment in the ICU. The potential of cost savings lies in the decreased use of other modes for assessing oxygen transport balance, e.g., cardiac output measurements and venous blood gas analyses. The savings in some institutions are greater than the price of the catheter, and its use has been judged cost-effective.

GASTRIC TONOMETRY

General Considerations

Ischemia signifies failure to satisfy the metabolic needs of the cell secondary to either impaired oxygen delivery or impaired cellular oxygen extraction and utilization. Although hemodynamic and oxygen-transport variables document the severity of tissue hypoxia and oxygen debt, they fail to accurately portray the complex interactions between energy requirements and supplies at the tissue level. Gastric tonometry has been proposed as a relatively noninvasive monitor of the adequacy of aerobic metabolism in organs whose superficial mucosal lining is extremely vulnerable to low flow and hypoxemia and in which blood flow is sacrificed first in both shock and the systemic inflammatory response syndrome. The gastrointestinal tract therefore will display metabolic changes before other indices of oxygen utilization. In the anoxic cell, uncompensated adenosine triphosphate (ATP) hydrolysis is associated with the intracellular accumulation of adenosine diphosphate (ADP), inorganic phosphate, and hydrogen ions with resulting intracellular acidosis. These hydrogen ions lead to tissue acidosis, with unbound hydrogen ions combining with interstitial bicarbonate to form the weak acid carbonic acid that dissociates to produce CO₂ and water.

Measurements

A tonometer is comprised of a semipermeable balloon connected to a sampling tube. A dual-purpose product—a tonometer in combination with a standard vented gastric sump—is commercially available (TRIP NGS Catheter; Tonometrics Division, Instrumentarium Corp., Tewksbury, MA). The annealed balloon allows CO₂ generated in the superficial layers of the mucosa to equilibrate within the saline instilled into the balloon. The tonometer P_{CO2} is multiplied by an equilibration factor (based on the equilibration period) to derive the tissue P_{CO2} value. Using a modified version of the Henderson-Hasselbach equation, intramucosal pH (pHi) is then calculated as follows:

pHi = 6.1 + log₁₀ HCO₃ / 0.03 × P_{CO2 ss}
where HCO₃ represents the arterial bicarbonate concentration (from a contemporary sample of arterial blood) in milliequivalents per liter, P_{CO2 ss} represents the tonometrically measured steady-state carbon dioxide tension, and 0.03 represents a constant that converts carbon dioxide tension in plasma in milliequivalents per liter per millimeter of mercury to milliequivalents per liter of carbonic acid. The measurement of intramucosal pH depends also on the assumption that the bicarbonate concentration in the wall of the organ is the same as that which is delivered to it by arterial blood and that the pK (the dissociation constant) is the same as that in the plasma. The pK in plasma is not the same as that in the cytosol, but the value 6.1 is the best approximation of the pK within the interstitial fluid of the superficial layers of the mucosa.

Clinical Utility

Incomplete splanchnic cellular resuscitation has been associated with the development of multiple organ system failure, more frequent septic complications, and increased mortality in the critically ill patient. Many studies have demonstrated the utility of gut tonometry in various settings where the perfusion status of the intestinal mucosa has been of importance,

such as in patients undergoing elective cardiac or abdominal aortic operations. In critically ill patients, gastric tonometry has been used as a predictor of both organ dysfunction and mortality and has been shown to be a better predictor of mortality than base deficit, lactate, oxygen delivery, and oxygen consumption.

Failure of splanchnic resuscitation correlated with multiple organ system failure and increased length of ICU stay in the hemodynamically unstable trauma patient admitted to the ICU at the University of Miami/Jackson Memorial Hospital. The relative risk or likelihood of death among incompletely resuscitated patients (pHi < 7.32) at 24 h as compared with completely resuscitated patients (pHi ³ 7.32) was 4.5. In parallel, the relative risk of developing multiple organ system failure was 5.4 in patients with pHi < 7.32.

In a subsequent audit, we evaluated consecutive patients who met state trauma triage criteria, who were admitted to the Ryder Trauma Center at Jackson Memorial Hospital, and who had a nasogastric tube placed during resuscitation. The trauma patients judged after clinical evaluation to have minimal injury and who were treated in routine care areas were found to have a normal pHi on transfer from the resuscitation area. The mean pHi was 7.4 ± 0.11, >7.2 in 95 percent of patients and >7.25 in 88 percent of patients. All patients survived; none developed multiple organ system failure.

Measurement of pHi can provide clinicians with a metabolic end point of resuscitation. pHi therefore can be used to ensure the completion of resuscitation as judged by normalization of gastric pHi. Current methods reported in the literature focus only on increasing oxygen delivery using fluid therapy and vasoactive agents with a-adrenergic action. Because these induce splanchnic vasoconstriction, which leads to gastric intramucosal acidosis, our strategy employs splanchnic vasodilatory agents to optimize systemic and mesenteric blood flow (isoproterenol, dobutamine, nitroprusside, nitroglycerine, prostaglandin E₁), reserving those agents which cause splanchnic vasoconstriction (epinephrine, high-dose dopamine, phenylephrine, norepinephrine) to treat severe hypotension (MAP < 55 mmHg). If pHi falls unexpectedly, we search for intraabdominal catastrophe, intraabdominal hypertension, sepsis, tissue necrosis, line sepsis, nosocomial infection, unappreciated excess patient ventilatory work, hypovolemia, and hypoxemia.

Interventions to block and modify the ischemia-reperfusion injury and restore splanchnic perfusion can be incorporated into a resuscitation algorithm to reduce the incidence of bacterial translocation and systemic white cell priming before the ensuing systemic inflammatory response.

RENAL MONITORING

The primary reason for monitoring renal function is that the kidney serves as an excellent monitor of the adequacy of perfusion. The second major indication for monitoring kidney function is to prevent acute parenchymal failure. Finally, renal function monitoring is helpful in predicting drug clearance and proper dose management.

Urine output frequently is monitored but may be misleading. Although very low urine outputs, less than 0.5 mL/kg/h, are consistently associated with low *glomerular filtration rate (GFR)* values, levels greater than this also can be associated with low GFR values. Diuresis created by an osmotic load (radiographic contrast material or glucose), administration of diuretics, or nonoliguric renal failure may give the clinician a false sense of security while the patient has deteriorating renal function. Other methods of monitoring renal function are necessary. These include tests of glomerular function and tests of tubular function.

Glomerular Function Tests

Blood urea nitrogen (BUN) often has been used to estimate renal function. BUN is affected by GFR and urea production. Production may be increased if large amounts of nitrogen are administered during parenteral nutritional support, as a result of gastrointestinal bleeding, or in catabolic states induced by trauma, sepsis, or steroids. Urea production may be lowered during starvation and in advanced liver disease. Because these factors are often interrelated in an unpredictable manner in critically ill patients, BUN is not a reliable monitor of renal function.

The value of plasma creatinine as a measure of renal function far exceeds the value of the BUN. The serum creatinine level is directly proportional to the level of creatinine production and inversely related to the GFR. In contrast to BUN concentration, plasma creatinine levels are not influenced by protein metabolism or the rate of fluid flow through the renal tubules. When creatinine production is constant, the serum creatinine level reflects GFR. The plasma creatinine level will double with a 50 percent reduction in GFR, assuming that creatinine production remains constant. Acute reductions in the GFR rate are not immediately reflected, however, because it takes 24 to 72 h for equilibration to occur.

Creatinine production is directly proportional to the muscle mass and its metabolism. In rhabdomyolysis, creatinine formation exceeds its filtration rate, such that measured serum creatinine levels increase. In conditions where the skeletal muscle mass is reduced (e.g., advanced age, immobilization), the endogenous creatinine pool is diminished, and thus the serum creatinine concentration does not rise appropriately with impairment of renal function. Only with measurement of creatinine clearance can the severity of such renal function loss be determined.

Serial determination of creatinine clearance is currently the most reliable method for clinically assessing GFR and the most sensitive test for predicting the onset of perioperative renal dysfunction. Although measurements traditionally are performed using a 24-h urine collection, measurements using a 2-h collection are reasonably accurate and easier to perform. When urine output is measured in clearance studies, special efforts should be made to ensure complete bladder emptying at the beginning and end of collection periods. Creatinine clearance (C_{cr}) is calculated from the formula

$$C_{cr} = U_{cr} \times V / P_{cr} \times 1.73 / BSA$$

where U_{cr} = urinary creatinine concentration, mg/dL

V = urinary volume flow, mL/min

P_{cr} = plasma creatinine concentration, mg/dL

BSA = body surface area, m²

Clearance results are expressed as milliliters per minute per 1.73m².

An alternative approach to creatinine clearance estimation is to forgo urine collection and utilize criteria formulated by Cockcroft and Gault. The Cockcroft-Gault equation for males is

$$C_{cr} = (140 - \text{age}) \times \text{weight in kg} / 72 \times P_{cr}$$

For females, the derived C_{cr} is multiplied by 0.85. Using the lower of ideal or total body weight and the higher of actual serum creatinine or corrected serum creatinine concentration to 1 mg/dL in this formula results in more accurate predictions of GFR. Although the Cockcroft-Gault equation provides an inexpensive and accurate estimate of GFR in patients with stable renal function, it is unlikely that it would detect early renal deterioration because the serum creatinine, on which the equation depends, does not begin to increase above normal until the GFR decreases to less than 50 mL/min/1.73 m².

Tubular Function Tests

Tests that measure the concentrating ability of the renal tubules are used primarily in the differential diagnosis of oliguria to differentiate a prerenal cause unresponsive to judicious fluid therapy from intrinsic renal failure due to tubular dysfunction. Table 12-4 summarizes the differential diagnosis based on tubular function studies. The utility of each depends on the ability of the renal tubular cells to physiologically respond to a decreased extracellular fluid volume. Thus, with prerenal azotemia, the tubules can appropriately reabsorb sodium and water. In intrinsic renal failure, tubular function is markedly compromised, and the ability to reabsorb sodium and water is impaired. Multifactorial renal dysfunction is most common; rarely is an isolated etiology discovered.

Tubular function tests are useful in oliguric patients (urine output < 500 mL/day) because nonoliguric individuals typically have less severe tubular damage and their laboratory findings are likely to show more overlap with the values of patients with prerenal azotemia.

The fractional excretion of sodium (FE_{Na}) appears to be the most reliable of the laboratory tests for distinguishing prerenal azotemia from acute tubular necrosis. This test requires only simultaneously collected “spot” urine and blood samples. FE_{Na} can be calculated as follows:

$$FE_{Na}(\%) = U_{Na}/P_{Na} / U_{cr}/P_{cr} \times 100\%$$

where U_{Na} is the urinary sodium concentration (in milliequivalents per liter), and P_{Na} is the plasma sodium concentration (in milliequivalents per liter).

The FE_{Na} value is normally less than 1 to 2 percent. In an oliguric patient, a value of less than 1 percent is usually due to a prerenal cause. A value greater than 2 to 3 percent in this setting suggests compromised tubular function. When the value ranges between 1 and 3 percent, the test is not discriminating.

Although the FE_{Na} is very useful, it is now apparent that a number of causes of acute renal failure other than prerenal disease can, on occasion, be associated with a FE_{Na} value of less than 1 percent. These include nonoliguric acute tubular necrosis, acute tubular necrosis superimposed on chronic prerenal disease (e.g., advanced cardiac or liver disease), administration of radiocontrast media or release of heme pigments (hemoglobin or myoglobin), and renal allograft rejection.

Thus the FE_{Na} test must be interpreted in light of the specific clinical setting and other laboratory data to be useful in patient management. Correct interpretation of FE_{Na} or any of the other urinary indices is not possible if the patient had received diuretics in the 6 to 12 h preceding the test.

NEUROLOGIC MONITORING

Monitoring the function of the central nervous system may permit early recognition of cerebral dysfunction and facilitate prompt intervention in situations in which aggressive early treatment favorably influences outcome. In the perioperative and trauma settings, several methods have been used to evaluate brain function and the effects of therapy. These include intracranial pressure monitoring, electrophysiologic monitoring, transcranial Doppler ultrasonography, and jugular venous oximetry.

Intracranial Pressure Monitoring

Physical findings are often unreliable to ascertain the presence of increased *intracranial pressure (ICP)*. Thus the only direct assessment of ICP is obtained by measurement.

Measuring ICP permits calculation of cerebral perfusion pressure (CPP), which is defined as the difference between the MAP and ICP. Thus isolated increases in ICP or decreases in MAP will result in a reduction in CPP. The CPP may be insufficient if ICP increases to more than 20 mmHg. Although in the past one of the end points of central nervous system monitoring was felt to be the control of ICP within safe levels, emphasis has shifted to following CPP itself. Maintaining cerebral blood flow appears to require using an elevated minimal CPP threshold when treating the injured brain. A CPP level of at least 70 mmHg has been suggested.

The most common indication for ICP monitoring is severe head injury. Patients with a Glasgow Coma Scale (GCS; see Chap. 40, Table 40-1) score of 8 or less or a GCS motor score of 5 or less (i.e., not following commands) should be strongly considered for ICP monitoring. Narayan and associates demonstrated that patients with a GCS score of 8 or less who have a normal CT scan have a very low probability of developing intracranial hypertension if they have fewer than two of the following features: (1) prior episodes of hypotension, (2) age greater than 40 years, or (3) motor posturing. They suggested that these patients may be managed initially without monitoring ICP. In such instances, however, any deterioration should prompt immediate reconsideration of ICP monitoring and reimaging.

Other conditions for which ICP monitoring has been recommended include subarachnoid hemorrhage, hydrocephalus, postcraniotomy, and Reye's syndrome. Although some investigators also have advocated this form of monitoring in patients with massive strokes, encephalitis, and post–cardiac arrest states, little evidence has been generated to suggest a beneficial impact.

Several methods of ICP measurement are available (Fig. 12-10). A ventricular catheter connected to a standard strain-gauge transducer via fluid-filled lines offers excellent waveform characteristics and permits withdrawal of cerebrospinal fluid (CSF). This catheter, however, may be difficult to insert when cerebral edema or hematoma causes shifting or collapse of the lateral ventricle system. A subarachnoid bolt is easily inserted under any circumstance, although at times it may give erroneous readings, depending on its placement relative to the site of injury. Compared with ventricular catheters, the waveforms obtained are not as good, and CSF drainage is usually not possible. Epidural bolts have a lower risk of complications but are less accurate than ventricular catheters or subarachnoid bolts and do not permit withdrawal of CSF. Non-fluid-coupled systems utilizing fiberoptic or catheter-tip strain-gauge technology can be placed into ventricular, subdural, and parenchymal sites. These devices appear to offer advantages over conventional ICP monitors, especially in their ability to measure brain parenchymal pressures. Without an associated ventricular catheter, however, catheter-tip strain-gauge or fiberoptic devices cannot be recalibrated after insertion. Consequently, if the device measurement drifts, there is potential for inaccurate measurements, especially if the ICP monitor is used for several days.

Complications of ICP devices include infection, hemorrhage, malfunction, obstruction, and malposition. Bacterial colonization of ICP devices increases significantly after 5 days of implantation; however, significant intracranial infections are uncommon.

Electrophysiologic Monitoring

The electroencephalogram (EEG) reflects spontaneous and ongoing electrical activity recorded on the surface of the scalp. Intraoperative EEG recording has been used primarily for monitoring the adequacy of cerebral perfusion during carotid endarterectomy. Other procedures in which EEG recording has been used include cerebrovascular surgery, open heart surgery, epilepsy surgery, and induced hypotension for a variety of surgical procedures. In the ICU, EEG recordings can detect the existence of subclinical epileptic seizures, which can cause a decreased level of consciousness. Standard EEG recording, however, is not used routinely in the ICU because of insufficient technical personnel, the volume of data generated in a short period, the difficulty of on-line EEG interpretation, numerous electrically induced artifacts, and drug-induced suppression of electrical activity. To simplify EEG recording and to make it more useful for clinical application in the operating room and ICU, some monitors process the raw data automatically. The compressed spectral array (CSA) is the most commonly used method of visually displaying such processed EEG information.

Sensory-evoked potentials (SEPs) are minute electrophysiologic responses elicited by a stimulus and extracted from an ongoing EEG by signal averaging. They reflect the functional integrity of specific sensory pathways and serve to some extent as more general indicators of function in adjacent structures. Somatosensory evoked potentials (SSEPs) reflect the integrity of the dorsal spinal columns and the sensory cortex and may be useful for monitoring during resection of spinal cord tumors, spine instrumentation, carotid endarterectomy, and aortic surgery. Brainstem auditory-evoked potentials (BAEPs) reflect the integrity of the eighth cranial nerve and the auditory pathways above the pons and are used for monitoring during surgery of the posterior fossa. Visual evoked potentials (VEPs) may be used to monitor the optic nerve and upper brainstem during resections of large pituitary tumors. Several studies have promoted the prognostic value of SEPs in the ICU, primarily in head trauma patients. Continuous monitoring in the ICU is still rare.

Transcranial Doppler Ultrasonography

Transcranial Doppler ultrasonography (TCD) is one of a number of methods currently being used to monitor cerebral blood flow. This technique employs the Doppler principle to record the blood flow velocity in the basal cerebral arteries and can provide valuable information regarding the integrity of the intracranial circulation. TCD also may detect vasospasm following spontaneous or traumatic subarachnoid hemorrhage and can help identify hyperemic or low-flow states.

The most commonly used TCD technique is “free-handing” the transmitter so that the ultrasound beam penetrates one of several natural “acoustical windows.” Modified probes are now available that may be kept in place for prolonged periods using a headband or other device. TCD is a valuable monitoring tool that is noninvasive, portable, and reproducible. It requires, however, that the studies be performed by a technician familiar with its use.

Jugular Venous Oximetry

Jugular venous oximetry is an invasive method of continuously monitoring jugular venous bulb oxyghemoglobin saturation ($S_{jv\ O_2}$). Changes in $S_{jv\ O_2}$ provide a measure of the relationship between total cerebral blood flow and total cerebral oxygen consumption. When arterial oxygenation remains constant, a decrease in $S_{jv\ O_2}$ from control values reflects a decrease in cerebral blood flow or an increase in cerebral metabolic oxygen consumption unmatched by an increase in flow. Jugular venous oximetry offers the potential to minimize secondary insults after severe head injury by facilitating recognition of cerebral ischemia. Other applications include monitoring during neurosurgical procedures, carotid endarterectomy, and cardiopulmonary bypass.

The technique for determination of $S_{jv\ O_2}$ is analogous to the determination of $SV\ O_2$ on a systemic basis (see above). A fiberoptic catheter is placed retrograde into the jugular bulb, and the $S_{jv\ O_2}$ is recorded in a continuous fashion. The range of $S_{jv\ O_2}$ in normal subjects is 55 to 71 percent. In head-injured patients, the range for $S_{jv\ O_2}$ is considerable wider, but most investigators agree that a sustained desaturation of 50 to 55 percent warrants evaluation. $S_{jv\ O_2}$ levels below 50 percent are indicative of cerebral ischemia. Artifactual recordings, however, are not uncommon and usually are caused by catheter malposition from head and neck movement.

METABOLIC MONITORING

General Considerations

The need to substitute artificial feedings during recovery from surgery and trauma turns the simple everyday function of ingestion of food into a complicated area involving mathematics, suppositions, high technology, and the potential for creating harmful side effects.

Assessment of Caloric Expenditure

Energy requirements depend on a number of factors including the body surface area, age, and sex. Basal energy expenditure (BEE) can be predicted with reasonable accuracy (± 5 percent) by the Harris-Benedict equation:

Men: $66.47 + (13.75 \times W) + (5.00 \times H) - (6.76 \times A) = \text{BEE kcal/day}$

Women: $655.10 + (9.56 \times W) + (1.85 \times H) - (4.68 \times A) = \text{BEE kcal/day}$

where W
= body weight, kg

H = height, cm

A = age, years

Resting energy expenditure (REE) can be approximated from the BEE by increasing it by 10 percent, adding the calories necessary to compensate for the specific dynamic action of food.

The stress of illness, the change in hormonal milieu relating to the stress state, alterations in substrate utilization, and fever all can be predicted to increase REE. While early studies emphasized increases of 25 percent for multiple-trauma patients and 50 percent for burn patients, even an increase of 60 percent above BEE would only result in a need for 40 kcal/kg/day, or less than 3000 kcal in a 70-kg patient. Excessive caloric administration is potentially detrimental. Carbohydrates given in excess of requirements are turned into fat, resulting in an increase in CO_2 production. The liver may develop fatty infiltration resulting in hepatic dysfunction.

In addition, many patients may have protein malnutrition despite being overweight due to obesity or overhydration. In a study of critically ill patients, Makk and colleagues found that 58 percent were overweight. Forty-two percent had a kwashiorkor-like pattern of malnutrition, and 11 percent had a marasmus-like pattern. Protein malnutrition necessitates administration of higher levels of protein to avoid or reverse end-organ dysfunction and impaired acute-phase protein synthesis, which sets the stage for nosocomial infections and the multiple organ system failure syndrome.

Assessment of Oxygen Consumption

If caloric expenditures can be considered the fuel, then another aspect of metabolic monitoring is the flame. Oxidative metabolism or oxygen consumption reflects metabolic activity. Cuthbertson described the ebb- and-flow phase of the metabolic response to injury. In his terms, the ebb phase, lasting 24 to 48 h, is a period of “diminished circulatory vitality”—decreased cardiac output, diminishing oxygen delivery, and resulting decreased oxygen consumption. The recognition of the association between diminished oxygen delivery in the ebb phase with the late occurrence of sepsis and multiple organ system failure has focused attention on improving oxygen delivery in an attempt to forestall this seemingly inevitable progression. Thus monitoring and optimizing oxygen delivery as described previously have been given additional significance in terms of minimizing long-term detrimental metabolic effects secondary to stress and injury.

Measurements

Oxygen delivery or cardiac output times arterial oxygen content and oxygen consumption (normally about 150 mL/min/m²) assess oxidative metabolism. CO₂ production is a measure of a by-product of oxidative metabolism. The ratio of CO₂ production to oxygen consumption is termed the *respiratory quotient* (RQ). During normal oral dietary intake, carbohydrates, protein, and fat are ingested, giving an average RQ of approximately 0.8. During prolonged starvation, the body adapts to fat metabolism, and the RQ may fall to as low as 0.6 to 0.7. On the other hand, during excessive carbohydrate administration, the transformation into fat releases additional CO₂, and the RQ rises above 1.0. Thus monitoring oxygen consumption and CO₂ production and calculating the RQ provide inferences into the adequacy of total calories as well as the mixture of substrates.

In the past few years, clinicians, recognizing the number of variables affecting REE and total energy expenditure (TEE), have attempted direct measurements in order to meet an individual patient's needs at a particular time in the course of the illness. It is clear that estimates of REE from the Harris-Benedict equation, even modified by estimates to adjust for increased metabolic activity, still lead to inaccurate assessment of TEE. Direct calorimetry measures body heat production and is correlated with energy use. The subject, however, must be placed in a closed chamber for measurement of heat production, and clearly, this is not applicable to patient care. Indirect calorimetry permits derivation of caloric requirements from the measurements of oxygen consumption and CO₂ production. Methods used include Douglas bag collection of expired gases, use of the Fick equation, and computerized open circuit measurements.

Douglas Bag Collection

Samples of inspiratory and mixed expiratory gases are analyzed for oxygen and CO₂ concentrations. Unfortunately, this technique is prone to many errors. Gas leakage, incomplete emptying, inaccurate measurement of total expired gases, and incomplete mixing or leakage through incompletely sealed connections are but a few of the problems encountered. Finally, it is a tedious method that does not lend itself to frequent or routine clinical use.

The Reverse Fick Method

The classic Fick equation (see Table 12-3) relates oxygen consumption to the product of cardiac output and arterial venous oxygen content difference. Repeated cardiac output determinations have an accuracy in the range of ±5 percent but are not as important as variability in the patient's physiology, in which changes may exceed ±10 percent in a short time. Pulse oximetry and mixed venous oximetry show that minute-to-minute variations in these values also occur. Smithies has shown that these factors result in a 10 percent variation between measurements and lower total values compared with validated spirometric techniques. This latter difference reflects the oxygen consumption of the lung, which is included in the spirometric method but not in the reverse Fick technique.

Computerized Open Circuit Indirect Calorimetry

Several commercial indirect calorimeters have been evaluated by Makita and associates. The Datex Deltatrac Metabolic Monitor (Datex Instrumentarium, Helsinki, Finland), the Engström Metabolic Computer (Gambro Engström A.B., Bromma, Sweden), and the SensorMedics MMC Horizon (SensorMedics, Anaheim, CA) use slightly different techniques and formulas. Mean relative errors of measurement for all three monitors, however, were in the range of 1.4 to 6 percent for oxygen consumption, CO₂ production, and respiratory quotient.

In routine clinical use, however, there are many potential errors. Instruments lose calibration easily. High inspired oxygen concentrations (Fi_{O2} > 0.6) render measurements inaccurate. Fluctuations in hospital gas- line pressure cause 1 to 2 percent fluctuations in inspired Fi_{O2}, sufficient to impair the accuracy of overall oxygen consumption. Leaks at joints and connections occur frequently and must be eliminated before and during studies. Measurements of CO₂ production depend on a steady-state relationship between CO₂ production at the cellular level, transport to lungs, and CO₂ elimination via the lungs and measured in exhaled gases. Therefore, changes in the metabolic rate, changes in cardiac output, and changes in ventilation all affect the inherent supposition that CO₂ production and CO₂ elimination measured by the metabolic monitor are equal. Minute-to-minute variations may be minimized by ensuring that measurements encompass at least 15 min twice daily.

Clinical Utility

First, metabolic monitoring can be used to judge the end result of the interactions of all the unknowns (including degrees of illness, activity, and metabolism) in producing the current levels of oxygen consumption and CO₂ production. Second, either too little or incorrect proportions of nutritional support are undesirable. If we remember that approximately 32 to 35 kcal/day should suffice in most clinical circumstances and that there are no outcome data suggesting that ±10 or even 20 percent has a demonstrable adverse effect on any known variable, we should ensure that we stay within this range of measured values. Third, we could identify patients with significant hypometabolism or hypermetabolic states because clinical estimates are often incorrect. Adjusting nutritional support in these circumstances seems to be common sense, although it has not been proved that these adjustments make any real difference. Fourth, given the wide variations in caloric requirements and in substrate utilization, metabolic monitoring, allowing more appropriate partitioning of the substrates, seems more desirable than administering quantities by rote or formula. Fifth, measurements of respiratory quotient may be useful in patients who have respiratory failure and CO₂ retention while receiving high carbohydrate loads. Identification of a high respiratory quotient due to excessive carbohydrate administration is one correctable factor.

Because commercially available indirect calorimeters now can measure oxygen consumption and CO₂ production and calculate other parameters of metabolic cellular activity, the demonstrated accuracy in measurement can be used to support more precise prescription of nutritional therapy. Questioning whether this degree of obtainable precision alters outcome is relevant; at the same time, inaccurate prescriptions based on invalid estimates would not seem to be a viable objective. The costs of this approach are not inconsiderable, and while they may be offset by savings in wasted nutritional support, more data reflecting improvement in outcome are necessary before mandatory or daily usage even in critically ill patients can be supported.

TEMPERATURE MONITORING

Temperature, along with heart rate, blood pressure, and respiratory rate, remains one of the traditional four cardinal vital signs. Temperature is usually taken rectally in ill patients or orally when significant elevations are not expected.

Temperatures of the periphery of the body measured at the mouth, skin, or axilla are often unreliable and may be influenced by factors such as mouth breathing, temperature of recently ingested food, and/or environmental temperature. Therefore, it is recommended that deeper core temperatures be taken in the critically ill.

Core temperatures, which are relatively resistant to external influences, more accurately reflect the mean temperature of the body's vital organs. Core temperature can be measured by placing a thermistor probe into either the esophagus or the rectum. Esophageal wires are uncomfortable and invasive and are used exclusively in patients under general anesthesia. Rectal probes are used commonly in the operating room and ICU but may be extruded from the rectum and have a recognized risk of bowel wall perforation.

Three devices are available commercially for bedside measurement of core temperature in ICU patients: pulmonary artery thermistor catheters, urinary bladder thermistor catheters, and infrared auditory canal probes. Measurement of pulmonary artery blood temperature by the pulmonary artery thermistor catheter has been used increasingly as a reliable indicator of core temperature. The need for a catheter is an obvious disadvantage of this approach. Urinary bladder catheters have the advantage of giving both exact measurements of urine output and continuous urine temperature. Infrared probes noninvasively measure tissue temperature in the ear canal; however, their measurements have more variability than bladder readings. Moderate variation is probably inherent in infrared ear thermometry due to differences in ear canal anatomy and probe positioning.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 13 - Skin and Subcutaneous Tissue

David M. Young
Stephen J. Mathes

INTRODUCTION

The skin is the largest and one of the most complex organs of the body. Its uniform appearance belies its great variation from region to region of the body and the complex organization and interaction of the many different cells and matrices of the skin. Although the skin functions simply as a protective barrier and interface with our environment, its structure and physiology are complex.

The skin is an extremely good interface with our environment. It is protective against most of the noxious agents, such as chemicals (by the impermeability of the epidermis), solar radiation (by means of pigmentation), infectious agents (through efficient immunosurveillance), and physically deforming forces (by the durability of the dermis). The skin is the major organ responsible for thermoregulation, having an efficient ability to conserve or disperse heat. To direct all these functions, the skin has a highly specialized nervous structure.

These various functions are better served by different components of skin, so teleologically there developed regional variation. The palms and soles are particularly thick, to bear weight. The fingertips have the highest density of sensory innervation and allow for intricate tasks. Even the lines of the skin, first described by Langer, are oriented perpendicularly to the long axis of muscles to allow the greatest degree of stretching and contraction without deformity.

The relative ease of observing and obtaining skin specimens for examination and experiments has made skin one of the best-studied tissues of the human body. Thus skin is not just the subject of the field of dermatology, but also the study of the skin launched the fields of immunology, transplantation, and wound healing. Although this chapter emphasizes surgically treated diseases of the skin, it is important for students of surgery to be familiar with the basic physiology and structure of skin since many of the future advances in medicine will come from these studies.

ANATOMY AND PHYSIOLOGY

The skin has been traditionally divided into three layers: the epidermis, the basement membrane, and the dermis (Fig. 13-1). The epidermis is composed mainly of cells, with very little extracellular matrix. Each cell type serves a specific barrier function. Keratinocytes provide a mechanical barrier, melanocytes a radiation barrier, and Langerhans' cells an immunologic barrier. The dermis contains mostly extracellular matrix, providing support for nerves, vasculature, and adnexal structures. The dermis allows skin to resist deforming forces and return to its resting state, thus providing durability. The basement membrane is a specialized structure that anchors the epidermis to the dermis.

The main cell type in the epidermis is the keratinocyte. The deep, mitotically active, basal cells are a single cell layer of the least-differentiated keratinocytes. Some multiplying cells leave the basal layer and begin to travel upward. In the spinous layer they lose the ability to undergo mitosis. These differentiated cells start to accumulate keratohyalin granules in the granular layer. Finally, in the horny layer, the keratinocytes age, the once- numerous intercellular connections disappear, and the dead cells are shed. Using radioactive and fluorescence labeling, experiments have shown the keratinocyte transit time to be from 40 to 56 days. The control of keratinocyte multiplication and subsequent maturation is an area of active study and may clarify the complex mechanism of cellular differentiation.

Melanocytes migrate to the epidermis from precursor cells in the neural crest. They lie scattered beneath basal cells and have dendritic processes that reach out to surrounding keratinocytes. They number approximately one for every 35 keratinocytes. The melanocytes produce a pigment, melanin, from tyrosine and cysteine. The pigment is packaged in melanosomes and transported to the tips of the dendritic processes. The tips are sheared off (apocoptation) and then phagocytized by the keratinocyte, thus transferring the pigment to the keratinocyte. Once in the keratinocyte they aggregate on the superficial side of the nucleus in an umbrella shape. The density of melanocytes is constant between individuals of different skin color. The rate of melanin production, transfer to keratinocytes, and melanosome degradation determine the degree of skin pigmentation. These activities are influenced by genetically activated factors as well as ultraviolet radiation and hormones such as estrogen, adrenocorticotrophic hormone, and melanocyte-stimulating hormone.

The Langerhans' cells migrate from the bone marrow and function as the skin's macrophages. The Langerhans' cells constitutively express class II major histocompatibility antigens and have antigen-presenting capabilities. These cells play a crucial role in immunosurveillance against viral infections and neoplasms of the skin, and may initiate skin allograft rejection.

The dermis consists mostly of several structural proteins. Collagen constitutes 70 percent of the dry weight of dermis and is responsible for its remarkable tensile strength. Tropocollagen consists of three polypeptide chains (formed mainly of hydroxyproline, hydroxylysine, and glycine) wrapped in a helix. These long molecules are then cross-linked to one another to form collagen fibers. Of the seven structurally distinct collagens, the skin contains mostly type I. Early fetal dermis contains mostly type III (reticulin fibers) collagen, but this remains only in the basement membrane zone and the perivascular regions in postnatal skin. Elastic fibers are highly branching proteins that are capable of being reversibly stretched to twice their resting length. This allows skin to return to its original form after stretching. Ground substance is an amorphous material that fills the remaining spaces. It consists of various polysaccharide-polypeptide (glycosaminoglycans) complexes. The nonsulfated form is mostly hyaluronic acid, and the sulfated forms are heparin sulfate, dermatan sulfate, and chondroitin-6-sulfate. Glycosaminoglycans, which can hold up to 1000 times their own volume in water, constitute most of the volume of dermis.

Fibroblasts are scattered throughout the dermis and are responsible for production and maintenance of the protein matrix. Recently proteins that control the proliferation and migration of fibroblasts have been isolated. The study of fibroblast activity by these growth factor interactions is crucial to our understanding of wound healing and organogenesis.

The basement membrane zone of the dermoepidermal junction is a highly organized structure of proteins that anchors the epidermis to the dermis. Mechanical disruption or a genetic defect in the synthesis of this structure (epidermolysis bullosa) results in separation of the epidermis from the dermis.

In the dermis are situated the remaining structures of the skin. An intricate network of blood vessels regulates body temperature. Two horizontal plexuses, one at the dermal-subcutaneous junction and one in the papillary dermis, are interconnected by vertical vascular channels. Glomus bodies are tortuous arteriovenous shunts that allow a tremendous increase in blood flow to the skin when open. This large amount of blood flow in excess of its nutritional needs allows skin to dissipate a vast amount of body heat when needed.

Thermoregulation is carried out by autonomic fibers that synapse to sweat glands, the hair erector muscles, and control points in the vasculature. Sensory innervation follows a dermatomal distribution from segments of the spinal cord. These fibers connect to corpuscular receptors (pacinian, Meissner's, and Ruffini's) that respond to pressure, vibration, and touch, to "unspecialized" free nerve endings associated with Merkel cells of the basal epidermis, or to hair follicles. These nerves are stimulated by temperature, touch, pain, and itch.

The skin has three main adnexal structures. The eccrine glands, which produce sweat, are located all over the body but are concentrated on the palms, soles, axillae, and forehead. The hair follicles consist of a mitotically active germinal center that produces a cylinder of tightly packed cornified epithelial cells. Control of the growth cycle of the hair is little understood. The sebaceous glands produce an oily substance that coats the skin. Together these two structures form a pilosebaceous unit. The apocrine glands are found primarily in the axillae and the anogenital region. In lower mammals these glands produce scent hormones (pheromones).

TRAUMA

Penetrating Injuries

Disruption of the continuity of the skin allows the entry of organisms that can lead to wound infection. Sharp lacerations, bullet wounds, "road rash" (injury from scraping against road pavement), and degloving injuries should be treated by gentle cleansing, debridement of all foreign debris and necrotic tissue, and application of a proper dressing. Dirty or infected wounds should be left open to heal by secondary intention or delayed primary closure. Clean lacerations may be closed primarily. Road rash injuries are treated as second-degree burns and degloving injuries as third-degree burns. The degloved skin can be placed back on the wound like a skin graft and assessed for survival prospects. If the skin becomes necrotic it is removed and the wound is covered with split-thickness skin grafts.

Pressure Ulcers (Decubitus Ulcers)

Pressure ulcers, as the name implies, are caused by excessive, unrelieved pressure. In animal studies, 60 mmHg pressure applied to the skin for 1 h produces histologically identifiable injuries such as venous thrombosis, muscle degeneration, and tissue necrosis. The average human being exerts 60 to 70 mmHg pressure on such body areas as the sacrum, occiput, and heels while lying in bed or on the ischia while sitting in a chair. Healthy people, however, regularly shift their body weight, even while asleep. Sitting in one position causes pain in areas of increased pressure, thus stimulating movement. Patients unable to sense pain or to shift their body weight, such as paraplegics or bedridden individuals, develop prolonged elevated tissue pressures and, eventually, necrosis. Muscle tissue is more sensitive to ischemia than the overlying skin. That is why the necrotic area is always wider and deeper than it appears on first inspection (Fig. 13-2).

Treatment of pressure sores requires relief of pressure with special cushions and beds and nutritional support to promote healing. The necrotic tissue should be removed, often along with the underlying bony prominence. Shallow ulcers may close by secondary intention, but deeper wounds with involvement of the underlying bone require surgical debridement and soft tissue and skin coverage. To prevent future breakdown of the area, stable coverage should be obtained with local myocutaneous or fasciocutaneous flaps. Prevention of ulcers is best achieved by close attention to susceptible areas and frequent repositioning of paralyzed patients. Air flotation mattresses and gel seat cushions redistribute pressure, decrease the incidence of pressure ulcers, and are cost effective in the care of patients at high risk. The addition of growth factors to these wounds has been found to increase healing and offers promising therapy in the future.

Keloids and Hypertrophic Scars

All wounds heal by scar formation. Hypertrophic scars are raised, red, and nodular, but remain within the limits of the original incision or trauma. Keloids are much bulkier, and their nodularity and firmness extend beyond the wound (Fig. 13-3). Although there are distinct histologic and biochemical differences between keloids and hypertrophic scars, the distinction is largely clinical. Dark-skinned individuals are more likely to develop keloids and hypertrophic scars, and a genetic component has been found. The location of a wound, such as across a joint and on the sternum, is an important factor in development of these lesions. Fibroblasts isolated from patients with keloids synthesize increased amounts of extracellular matrix proteins and growth factors.

Medical treatment includes intradermal injections of corticosteroids (triamcinolone acetonide) to reduce the itching and flatten the scar. Mechanical pressure (as with pressure garments) also can soften and flatten raised scars. Excision of keloids and hypertrophic scars remains the most effective treatment. These lesions are prone to recur unless steroid injections or pressure therapy with silicone gel sheeting is begun soon after surgery. Laser treatment and radiation therapy of these lesions have been found to be effective. With a better understanding of the pathogenesis of these abnormal states of wound repair, more effective therapy with local growth factors or their blockers is being developed.

INFECTIONS

Bacterial Infections

Folliculitis, Furuncles, and Carbuncles

Folliculitis is infection and inflammation of a hair follicle. The causative organism is usually *Staphylococcus* and occasionally a gram-negative organism. A furuncle (boil) begins as folliculitis but progresses to form a nodule that eventually becomes fluctuant. The abscess eventually ruptures and usually resolves. Deep-seated infections that result in multiple draining cutaneous sinuses are called carbuncles.

Folliculitis usually resolves with time and adequate hygiene. Warm soaks to a furuncle may hasten liquefaction, speed drainage, and encourage healing. Occasionally antibiotics are used to manage surrounding cellulitis. Carbuncles are more difficult to treat and require incision and drainage or wide excision of the infected tissue and sinuses.

Hidradenitis Suppurativa

Bacterial infection of a plugged apocrine gland occurs most commonly in the axillae and inguinal and perianal regions. An abscess forms with subsequent drainage and sinus formation. Repeated infections create a wide area of inflamed and scarred tissue that is foul-smelling and painful (Fig. 13-4). Treatment of acute infections includes application of warm compresses, antibiotics, and open drainage. Proper hygiene and discontinuation of deodorants may prevent recurrence. Chronic hidradenitis requires excision of the entire area of infection and closure with skin grafts.

Pilonidal Disease

Infected pilonidal cysts of the sacrococcygeal region occur primarily in young adults and are four times more common in males. The pathogenesis of the disease is much debated, but sweaty activity and buttock friction, such as occurred in jeep drivers in World War II (jeep driver's disease), is associated with a high incidence of pilonidal disease. The infection probably begins in a pilosebaceous unit in the natal cleft. Recurrent trauma causes obstruction of a hair follicle and leads to infection. The localized folliculitis spreads into the surrounding soft tissue and produces an abscess. This eventually drains to the surface and produces a sinus that is usually located lateral to the midline. The sinus is lined with granulation tissue but over time can epithelialize. Constant movement and friction of the buttocks causes hair and loose debris to enter the tract, inciting a foreign-body reaction.

Acute pilonidal abscesses should be drained. Without further therapy, many will recur. Use of perioperative antibiotics does not affect the outcome of the disease. There are many different ways of treating the chronic sinus tract, including tract curettage, local excision and closure, wide excision and marsupialization, and wide excision and flap closure. Another method currently gaining favor is fistulotomy and marsupialization. Each method has its drawbacks. Patients undergoing primary closure stay in the hospital longer but return to work sooner than patients with the wound left open after excision. Nonsurgical treatment of a pilonidal lesion by natal- cleft shaving and perineal hygiene requires the least amount of hospitalization but may have a higher recurrence rate. The most cost- effective method remains to be determined.

Toxic Epidermal Necrolysis and Staphylococcal Scalded Skin Syndrome

These two diseases create a similar clinical picture, which includes erythema of the skin, bullae formation, and, eventually, wide areas of skin loss. Staphylococcal scalded skin syndrome (SSSS) is caused by an exotoxin produced during a staphylococcal infection of the nasopharynx or middle ear in the pediatric population. Toxic epidermal necrolysis (TEN) is thought to be an immunologic reaction to certain drugs, such as sulfonamides, phenytoin, barbiturates, and tetracycline. Diagnosis can be made with a skin biopsy examination because SSSS produces a cleavage plane in the granular layer of the epidermis, whereas TEN occurs at the dermoepidermal junction. The injury is similar to a second-degree burn. Treatment involves fluid and electrolyte replacement and wound care as in a burn injury. In Stevens-Johnson syndrome epithelial sloughing of the respiratory and alimentary tracts occurs, with resultant respiratory failure and intestinal malabsorption.

Patients with TEN should be treated in burn units to decrease the morbidity from the wounds. The skin slough has been successfully treated with cadaveric or porcine skin or semisynthetic biologic dressings (Biobrane). Temporary coverage with a biologic dressing allows the underlying epidermis to regenerate spontaneously. Corticosteroid therapy has not been efficacious.

Actinomycosis

Actinomycosis is a localized inflammatory mass, usually of the jaw area, that spreads by multiple fistulas and abscesses into the neck and face. The underlying bone also can become infected, as can the apex of the lung. The causative agent is *Actinomyces*, an organism of the Actinomycetaceae family, in the Actinomycetales order. Other actinomycetes, including *Nocardia*, *Actinomadura*, and *Streptomyces*, cause mycetomas, which are deep cutaneous infections that present as nodules and spread to form draining tracts to the skin and surrounding soft tissue. Chronic disease causes fibrosis and contractures. The most common site for infection is the foot (Madura foot).

The gram-negative bacteria that cause these infections were once believed to be fungi because they grow slowly as branched filaments and chains. Diagnosis depends on the presence of characteristic sulfur granules on microscopic examination. Special stains should be used to exclude fungal infection. Penicillin and sulfonamides are effective against these infections. Abscesses and areas of chronic scarring may require surgical therapy.

Lymphogranuloma Venereum

Chlamydia trachomatis is a sexually transmitted, intracellular, gram- negative bacterium. After infection and a 2-week incubation period, an inconspicuous ulcer appears on the penis or labia, although in more than half the cases this lesion is not noticed or does not appear. A few weeks later, inguinal lymphadenopathy erupts. The nodes become very large and painful (buboes) and are occasionally confused with an incarcerated inguinal hernia. Adenopathy can occur above and below the inguinal ligament, forming a characteristic groove. The matted nodes may suppurate, and occasionally they rupture. Surgical drainage of unruptured abscesses is not recommended because a chronic draining sinus often develops. Active infection is treated with doxycycline for 1 week or azithromycin in one dose. Inflammation from infection can lead to lymphatic obstruction and chronic lower extremity edema. Rectal strictures also can occur.

Viral Infections

Warts are epidermal growths associated with human papillomavirus infection. Histologically they are characterized by hyperkeratosis (hypertrophy of the horny layer), acanthosis (hypertrophy of the spinous layer), and papillomatosis. Koilocytes, large keratinocytes with eccentric nuclei, are present. Different morphologic types have a propensity to occur on different parts of the body. The common wart (verruca vulgaris) is found on the fingers and toes, and has a rough, gray-brown surface. Plantar warts (verruca plantaris) occur on the soles and palms, and may look like a callus. Flat warts (verruca plana), which are flat but slightly raised, appear on the face, legs, and hands. Venereal warts (condylomata acuminata) grow in the moist areas around the vulva, anus, and scrotum.

Warts can be removed by a number of chemicals, including formalin, podophyllum, and phenol-nitric acid. Curettage with electrodesiccation also can be used for scattered lesions. Treatment of extensive areas of skin requires surgical excision under general anesthesia. Because of the infectious etiology, recurrences are common, and repeated excisions often are necessary to eliminate lesions. Some warts (especially human papillomavirus types 5, 8, and 10) are associated with squamous cell cancers, and therefore lesions that grow rapidly or ulcerate should be biopsied.

Condylomata acuminata are sexually transmitted and can be particularly bothersome. Patients with human immunodeficiency virus (HIV) infection are more likely to develop clinically significant venereal warts. The lesions often are multiple and can grow to large size (Buschke-Lowenstein tumor). Small lesions can be treated with podophyllotoxin cream. Larger lesions have a significant risk of malignant transformation and should be excised. The lesions often recur. Adjuvant therapy with interferon, isotretinoin, or autologous tumor vaccine decreases recurrence rates.

BENIGN TUMORS

Cysts (Epidermal, Dermoid, Trichilemmal)

Epidermal cysts are the most common type of cutaneous cyst. They occur anywhere on the body, as a single firm nodule. On the scrotum they are often multiple and can calcify. Trichilemmal (pillar) cysts, the next most common, occur more often in females and usually on the scalp. When ruptured these cysts have a characteristic strong odor. Dermoid cysts are present at birth and may result from epithelium trapped during midline closure in fetal development. Dermoids are most often found in the midline of the face (e.g., on the nose or forehead) and are also common on the eyebrow (Fig. 13-5).

On gross examination, it is difficult to distinguish one type of cyst from another. They are all subcutaneous, thin-walled nodules containing a white, creamy center. Histologic examination is needed to differentiate them. The walls of all these cysts consist of a layer of epidermis oriented with the basal layer superficial and the more mature layers deep (i.e., with the epidermis growing into the center of the cyst). The desquamated cells (keratin) collect in the center and form the creamy substance of the cyst. Epidermal cysts have a completely mature epidermis containing a granular layer. Trichilemmal cyst walls do not contain a granular layer but do have a distinctive outer layer resembling the outer root sheath of the hair follicle (trichilemmoma). Dermoids have a squamous epithelium, eccrine glands, pilosebaceous units, and, occasionally, bone, tooth, or nerve tissue. Surgeons often refer to cutaneous cysts as sebaceous cysts because they appear to contain sebum; this is a misnomer because the substance is actually keratin.

Cysts usually are asymptomatic and ignored until they rupture and cause local inflammation. The area becomes infected and an abscess forms. Incision and drainage is recommended for an acutely infected cyst. After resolution of the abscess the cyst wall must be excised or the cyst will recur. Similarly, when excising an unruptured cyst, care must be taken to remove all of the wall in order to prevent recurrence.

Keratoses (Seborrheic, Solar)

Seborrheic keratoses commonly occur on the chest, back, and abdomen of older individuals. The lesions are light brown or yellow and have a velvety, greasy texture. They are rarely mistaken for other lesions, so biopsy and treatment are seldom needed. Sudden eruptions of multiple lesions in elderly patients may be associated with internal malignancies.

Solar (or actinic) keratoses are also found in the older age group. They arise in sun-exposed areas of the body, such as the face, the forearms, and the back of the hands.

Histologically they contain atypical-appearing keratinocytes and evidence of solar damage in the dermis. These are thought to be premalignant lesions, and squamous cell carcinoma may develop over time. Treatment is by local removal or application of topical 5- fluorouracil. Malignancies that do develop rarely metastasize.

Nevi (Acquired, Congenital)

Acquired melanocytic nevi are classified as junctional, compound, or dermal, depending on the location of the nevus cells. This classification does not represent different types of nevi but rather different stages in the maturation of nevi. Initially, nevus cells accumulate in the epidermis (junctional), migrate partially into the dermis (compound), and finally rest completely in the dermis (dermal). Eventually most lesions undergo involution.

Congenital nevi are much more rare, occurring in only 1 percent of neonates. These lesions are larger and may contain hair. Histologically they appear similar to acquired nevi. Congenital giant lesions (giant hairy nevus) most often occur in a bathing trunk distribution or on the chest and back (Fig. 13-6). These lesions are a major cosmetic problem. In addition, they develop malignant melanoma in 5 percent of the cases. Excision of the nevus is the treatment of choice, but often the lesion is so large that closure of the wound with autologous skin grafts is not possible because of the lack of adequate donor sites. Serial excisions over several years with either primary closure or skin grafting is the present mode of therapy. Tissue expansion of normal surrounding skin is now also used to accelerate the rate of nevus excision and avoid the use of skin grafts.

Vascular Tumors

Hemangiomas (Capillary, Cavernous)

Hemangiomas are benign vascular neoplasms that arise soon after birth. They undergo rapid cellular proliferation initially and slowly involute through early childhood. Capillary (strawberry) hemangiomas are soft, compressible papular lesions with sharp borders located mostly on the shoulders, face, and scalp. Cavernous hemangiomas are bright red or purple and have a spongy consistency. Histologically capillary hemangiomas are composed of endothelial cells seen primarily in fetal veins. Cavernous lesions contain large, blood-filled spaces lined by normal-appearing endothelial cells.

Hemangiomas can enlarge during the first year of life, and more than 90 percent of them involute over time. Allowing lesions to regress spontaneously usually gives optimal cosmetic results (Fig. 13-7). Acute treatment is limited to lesions that interfere with bodily functions, such as vision, feeding, and urination, or lead to systemic problems, such as thrombocytopenia and high-output cardiac failure. The growth of these rapidly enlarging lesions can be stopped with a course of prednisone or interferon-alpha-2a. Hemangiomas that remain after early adolescence will probably not involute further. Surgical excision is recommended.

Vascular Malformations (Port Wine Stains, Arteriovenous Malformations, Glomus Tumors)

Vascular malformations are a result of structural abnormalities formed during fetal development and hence are not neoplasms. Unlike hemangiomas, vascular malformations do not undergo rapid growth and involution but rather grow in proportion to the body. Histologically they contain enlarged vascular spaces lined by nonproliferating endothelium, and not the mitotically active endothelial cells of a hemangioma.

The port wine stain (nevus flammeus) is a flat, dull red capillary malformation that can be located on the trunk, extremities, and, most commonly, along a trigeminal distribution on the face. Histologically these nevi are composed of ectatic capillaries lined by mature endothelium. They may be part of the Sturge-Weber syndrome (leptomeningeal angiomatosis, epilepsy, and glaucoma). Unsightly lesions can be covered with cosmetics, treated with pulsed dye laser, or surgically excised.

Arteriovenous malformations are high-flow lesions. They appear as a mass under the skin with locally elevated temperature, a dermal stain, and a thrill and bruit. Overlying ischemic ulcers, adjacent bone destruction, or local hypertrophy may occur. Large malformations can cause cardiac enlargement and congestive heart failure.

Complications of arteriovenous malformations, such as pain, hemorrhage, ulceration, cardiac effects, and destruction of surrounding structures, should be treated by elimination of the lesion. Therapy consists of angiography with selective embolization or complete surgical resection. Embolization is particularly useful for lesions not accessible to surgery or in cases in which resection would cause too much mutilation. Embolization also can be used preoperatively to reduce blood loss during surgery. Occasionally hypothermia and cardiac bypass are required in order to minimize blood loss during surgical excision of large lesions.

Glomus tumors are blue-gray nodules that are extremely tender. They can occur anywhere on the body, but the most common location is subungual. The tumor arises from a glomus body and histologically resembles the arterial portion of the glomus. Excision of the tumor relieves the pain.

Soft Tissue Tumors (Acrochordons, Dermatofibromas, Lipomas)

Acrochordons (skin tags) are fleshy, pedunculated masses located on the axillae, trunk, and eyelids. They are composed of hyperplastic epidermis over a fibrous connective tissue stalk. These lesions usually are small and are always benign.

Dermatofibromas are usually solitary nodules measuring approximately 1 to 2 cm in diameter. They are found primarily on the legs and sides of the trunk. The lesions are composed of whorls of connective tissue containing fibroblasts. The mass is not encapsulated, and vascularization is variable. Dermatofibromas can be diagnosed by clinical examination. When lesions enlarge to 2 to 3 cm, excisional biopsy is recommended to assess for malignancy.

Lipomas are the most common subcutaneous neoplasm. They are found mostly on the trunk but may appear anywhere. They may sometimes grow to a large size. Microscopic examination reveals a lobulated tumor containing normal fat cells. Excision is performed for diagnosis and to restore normal skin contour.

Neural Tumors (Neurofibromas, Neurilemmomas, Granular Cell Tumors)

Benign cutaneous neural tumors arise primarily from the nerve sheath. Neurofibromas can be sporadic and solitary, but they are more commonly noted in multiple formations associated with café-au-lait spots and an autosomal dominant inheritance (von Recklinghausen's disease). The lesions are firm, discrete nodules attached to a nerve. Histologically there is proliferation of perineurial and endoneurial fibroblasts and Schwann cells embedded in collagen. Neurilemmomas are solitary tumors found along peripheral nerves of the head and extremities. They are discrete nodules that may be locally painful or radiate along the distribution of the nerve. Microscopically the tumor contains Schwann cells with nuclei packed in palisading rows.

Granular cell tumors usually are solitary lesions of the skin or, more commonly, the tongue. They consist of granular cells derived from Schwann cells that often infiltrate the surrounding striated muscle.

MALIGNANT TUMORS

The most common cancers of the skin arise from the cells of the epidermis; they are, in order of frequency, basal cell carcinoma, squamous cell carcinoma, and melanoma. Malignancies arising from cells of the dermis or adnexal structures are much less common.

Environmental influences and concomitant diseases are associated with an increased incidence of epidermal malignancies. These factors have been extensively studied and form some of our best understanding about the causes of cancer.

Epidemiology

Increased exposure to ultraviolet radiation is associated with an increased development of all three of the common skin malignancies. Epidemiologic studies have shown that people with outdoor occupations have skin malignancies more often than people who work indoors. Squamous cell cancer is much more common on the lower lip than the upper. People with fair complexions are more prone to skin cancer. These same people also are more likely to develop malignancies if they live in areas of the world that receive more sunlight, such as New Zealand, as compared to Great Britain. Albino individuals of dark-skinned races are prone to develop cutaneous neoplasms that usually are rare in the nonalbino members, suggesting that melanin has a large role in protection from carcinogenesis.

Other factors associated with skin malignancies also have been identified. Chemical carcinogens have long been known. In the eighteenth century Sir Percival Pott noted the association of soot and scrotal cancer in chimney sweeps. Tar, arsenic, and nitrogen mustard are known carcinogens. Human papillomavirus has been found in certain squamous cell cancers and may be linked with oncogenesis. Radiation therapy in the past for skin lesions such as acne vulgaris, when it resulted in radiation dermatitis, is associated with an increased incidence of basal and squamous cell cancers in the treated areas. Any area of skin subjected to chronic irritation, such as burn scars (Marjolin's ulcers), repeated sloughing of skin from bullous diseases, and decubitus ulcers, all have an increased chance of developing squamous cell cancer. A variant of this type of lesion develops on skin that has suffered repeated burns.

Systemic immunologic dysfunction is related to an increase in cutaneous malignancies. Immunosuppressed patients receiving chemotherapy for other malignancies or immunosuppressants for organ transplants have an increased incidence of basal cell and squamous cell cancers and malignant melanoma. The acquired immunodeficiency syndrome (AIDS) is associated with an increased risk of developing skin neoplasms. Patients with HIV infection should be monitored vigilantly for early diagnosis of skin cancer.

Basal Cell Carcinoma

Basal cell carcinomas contain cells that resemble the basal cells of the epidermis. It is the most common skin cancer and is subdivided into several types by gross and histologic morphology. The nodulocystic or noduloulcerative type accounts for 70 percent of basal cell carcinomas. It is a waxy, cream-colored lesion with rolled, pearly borders (Fig. 13-8). It often contains a central ulcer. When these lesions are large they are called "rodent ulcers." Pigmented basal cell carcinomas are tan to black in color and should be distinguished by biopsy examination from melanoma. Superficial basal cell cancers occur more commonly on the trunk and form a red, scaling lesion sometimes difficult to distinguish grossly from Bowen's disease. A rare form of basal cell carcinoma is the basosquamous type, which contains elements of basal cell and squamous cell cancer. These lesions can metastasize more like a squamous cell carcinoma and should be treated aggressively. Other types include morpheaform, adenoid, and infiltrative carcinomas.

Basal cell carcinomas usually are slow growing, and patients often neglect these lesions for years. Metastasis and death from this disease is extremely rare, but the lesions can cause extensive local destruction. The majority of small (less than 2 mm), nodular lesions may be treated by dermatologists with curettage and electrodesiccation or laser vaporization. A major drawback to these procedures is that no pathologic specimen can be obtained to confirm the diagnosis. Larger tumors, lesions that invade bone or surrounding structures, and more aggressive histologic types (morpheaform, infiltrative, and basosquamous) are best treated by surgical excision with a 2- to 4-mm margin of normal tissue. Histologic confirmation that the margins of resection do not contain tumor is required. Because nodular lesions are less likely to recur, the smaller margin may be used, whereas the other types need a wider margin of resection. Alternative methods of treatment, such as radiation therapy and Mohs' surgery, are discussed later.

Squamous Cell Carcinoma

Squamous cell carcinomas arise from keratinocytes of the epidermis. It is less common than basal cell carcinoma but is more devastating because it can invade surrounding tissue and metastasize more readily. In situ lesions have the eponym of Bowen's disease, and in situ squamous cell carcinomas of the penis are referred to as erythroplasia of Queyrat. Contrary to previous reports, Bowen's disease is not a marker for other systemic malignancies.

Tumor thickness correlates well with its biologic behavior. Lesions that recur locally are more than 4 mm thick and lesions that metastasize are 10 mm or more. The location of the lesion also is important. Tumors arising in burn scars (Marjolin's ulcer), areas of chronic osteomyelitis, and areas of previous injury metastasize early. Lesions on the external ear frequently recur and involve regional lymph node basins early. Squamous cell cancers in areas with solar damage behave less aggressively and usually require only local excision.

Although small lesions can be treated with curettage and electrodesiccation, most surgeons recommend excision of the tumor. Lesions should be excised with a 1-cm margin, if possible, and histologic confirmation that the margins are tumor-free is mandatory. Tumor invading bone should be excised if recurrence is to be avoided. Regional lymph node excision is indicated for clinically palpable nodes (therapeutic lymph node dissection). Lesions arising in chronic wounds behave aggressively and are more likely to spread to regional lymph nodes. For these lesions lymphadenectomy before the development of palpable nodes is indicated (prophylactic lymph node dissection). Metastatic disease is a poor prognostic sign, with only 13 percent of patients alive after 10 years.

Alternative Therapy

Alternatives to surgical therapy for squamous and basal cell cancers consist of radiation therapy or topical 5-fluorouracil for patients unable or unwilling to undergo surgery. Radiation therapy for small and superficial lesions obtains cure rates comparable to surgical excision. Radiation damage to surrounding normal skin with inflammation and scarring can be a problem. Also the development of cutaneous malignancies in irradiated skin is a serious long-term risk with this method.

For lesions on the face near the nose or eye, resection of a wide rim of normal tissue to remove all the tumor can cause significant functional and cosmetic problems. These lesions can be removed by Mohs' micrographic surgery. Mohs' technique, developed in 1932, is a method to serially excise a tumor by taking small increments of tissue until the entire tumor is removed. Each piece of tissue removed is frozen and immediately examined microscopically to determine whether tumorous tissue has been resected. The advantage of this method over that of standard histologic examination after wide surgical resection is that the entire margin of resection is evaluated. The major benefit is the ability to remove a tumor with the least sacrifice of uninvolved tissue. This technique is effective for treating carcinomas around the eyelids and nose, where tissue loss is most conspicuous. The procedure is extremely lengthy (up to several days) since complete excision may require multiple attempts; this remains its major drawback. Cure rates are comparable to those of wide excision.

Patients with basal cell carcinomas have been treated with intralesional injection of interferon. The majority of the lesions were eliminated or controlled by the injections. The lesions that did not respond required surgical excision. When lesions respond to injections, operation is avoided and no reconstruction of the defect is required. The major disadvantages of this treatment are the need for multiple office visits over several weeks for injections, the systemic side effects of interferon, and a potential need for surgery if the lesions do not respond to injections. Clinical trials with combinations of retinoids (vitamin A derivatives) and interferon have demonstrated good response rates in patients with advanced, inoperable squamous cell carcinomas. These results suggest that interferon is likely to have a greater role in therapy of cutaneous neoplasms in the future.

Malignant Melanoma

What was a relatively rare disease 50 years ago has now become alarmingly more common. The rise in the rate of melanoma is the highest of any cancer in the United States. In 1935 the annual incidence of the disease was 1 per 100,000 people. By 1991 the incidence had risen to 12.9 per 100,000. There were 32,000 new cases of melanoma in 1991. The case fatality rate has fallen over the years, probably as the result of earlier detection and treatment.

Because melanoma is becoming so common, it is important for all physicians to be familiar with this disease. The important clinical features of a melanoma include a pigmented lesion with an irregular, raised surface and irregular borders. About 5 to 10 percent of melanomas are not pigmented. Lesions that change in color and size and ulcerate over a few months' time are suspicious and should be biopsied. Surgery is still the mainstay of therapy for melanoma, so it is imperative that surgeons be aware of the latest methods of diagnosis, staging, and therapy.

Pathogenesis

Melanoma arises from transformed melanocytes and can arise anywhere that melanocytes have migrated during embryogenesis. The eye, central nervous system, gastrointestinal tract, and even the gallbladder have been reported as primary sites of the disease. Over 90 percent are found on the skin; however, 4 percent of melanomas are discovered as metastases without an identifiable primary site. Many melanomas, especially in the early phases of growth, are found to contain areas of tumor regression on histologic examination. Regression represents a host immune response to the tumor. Metastatic melanomas with unknown primary sites probably arise from completely regressed lesions that are difficult to locate.

Nevi are benign melanocytic neoplasms found on the skin of most people. Dysplastic nevi are much rarer and contain a histologically identifiable focus of atypical melanocytes. This type of nevus may represent an intermediate between a benign nevus and a true malignant melanoma. It is well documented that patients with melanoma have significantly more nevi and dysplastic nevi than matched controls. The relative risk of developing melanoma increases with the number of dysplastic nevi that a patient develops. The relationship is similar to that between the number of colonic polyps and the development of colon cancer. Patients with dysplastic nevi and family members with dysplastic nevi and melanoma are at increased risk for developing melanoma, suggesting that in these patients there is a genetic component to the risk of developing the malignancy.

Once the melanocyte has transformed into the malignant phenotype, the growth of the lesion is radial in the plane of the epidermis. Even though microinvasion of the dermis can be observed during this radial growth phase, metastases do not occur. Only when the melanoma cells form nests in the dermis are metastases observed. The transformed cells in the vertical growth phase are morphologically different and express different cell-surface antigens than those in the radial phase or cells of the dysplastic nevus. In addition, these cells behave differently in cell culture. They can grow in a less enriched media and have a longer life span.

Types

There are four common distinct types of melanoma. These are, in order of decreasing frequency, superficial spreading, nodular, lentigo maligna, and acral lentiginous. Each has distinct characteristics and behaviors.

The most common type, representing 70 percent of melanomas, is the superficial spreading type. These lesions occur anywhere on the skin except the hands and feet. They are flat, commonly contain areas of regression, and measure 1 to 2 cm in diameter at the time of diagnosis (Fig. [13-9](#)). There is a relatively long radial growth phase before vertical growth begins.

The nodular type accounts for 15 to 20 percent of melanomas. These lesions are darker and raised. The histologic criterion for a nodular melanoma is the lack of radial growth peripheral to the area of vertical growth; hence all nodular melanomas are in the vertical growth phase at the time of diagnosis. Although it is an aggressive lesion, the prognosis for a patient with a nodular-type lesion is the same as that for a patient with a superficial spreading lesion of the same depth of invasion.

The lentigo maligna type, accounting for 5 to 10 percent of melanomas, occurs mostly on the neck, the face, and the back of the hands of elderly people. These lesions are always surrounded by dermis with heavy solar degeneration. They tend to become quite large before a diagnosis is made but also have the best prognosis because invasive growth occurs late.

The rarer acral lentiginous type is distinctly different. It occurs on the palms and soles and in the subungual regions. Although melanoma among dark-skinned people is relatively rare, the acral lentiginous type accounts for a higher percentage in dark-skinned people than in people with less pigmented skin. The subungual lesions appear as blue-black discolorations of the posterior nail fold and are most common on the great toe or thumb.

Prognostic Factors

The original staging system classified melanoma into local (Stage I), regional lymph node (Stage II), and metastatic (Stage III) disease. This staging system had the disadvantage of lumping most patients into Stage I disease, therefore limiting its usefulness in prognostic studies. The most current staging system, from the American Joint Committee on Cancer (AJCC), contains the best method of interpreting clinical information in regard to prognosis of this disease (Table [13-1](#)).

The T classification of the lesion comes from the original observation by Clark that prognosis is directly related to the level of invasion of the skin by the melanoma. Whereas Clark used the histologic level (I, superficial to basement membrane [in situ]; II, papillary dermis; III, papillary/reticular dermal junction; IV, reticular dermis; and V, subcutaneous fat), Breslow modified the approach to obtain a more reproducible measure of invasion by the use of an ocular micrometer. The lesions were measured from the granular layer of the epidermis or the base of the ulcer to the greatest depth of the tumor (I, 0.75 mm or less; II, 0.76 to 1.5 mm; III, 1.51 to 3.0 mm; IV, 3.0 mm or more). These levels of invasion have been subsequently modified and incorporated in the AJCC staging system (Fig. [13-10](#)).

Evidence of tumor in regional lymph nodes is a poor prognostic sign. This is accounted for in the staging system by advancing any T classification from Stage I or II to Stage III. The 10-year survival rate drops precipitously with the presence of lymph node metastasis. The number of positive lymph nodes also is correlated with survival rates.

The presence of distant metastasis is a grave prognostic sign (Stage IV). The median survival ranges from 2 to 7 months, depending on the number and site of metastases, but survival up to a few years has been reported (Fig. [13-11](#)).

Other independent prognostic factors have been identified:

(1) Anatomic location. Independent of histologic type and depth of invasion, people with lesions of the extremities do better than people with melanomas of the trunk or face (82 percent 10-year survival rate for localized disease of the extremity, compared to 68 percent survival with a lesion of the face).

(2) Ulceration. Presence of ulceration in a lesion carries a worse prognosis. For unknown reasons these melanomas act more aggressively than nonulcerated ones. The 10-year survival rate for patients with local disease (Stage I) and an ulcerated melanoma was 50 percent, compared to 78 percent for the same stage lesion without ulceration.

(3) Sex. Women have melanomas in more favorable anatomic sites, and these lesions are less likely to contain ulceration. Despite correction for these factors, females have a higher survival rate than men (80 percent 10-year survival for women versus 61 percent for men with Stage I disease).

(4) Histologic type. Nodular melanomas have the same prognosis as superficial spreading types when lesions are matched for depth of invasion. Lentigo maligna types, however, have a better prognosis even after correcting for thickness, and acral lentiginous lesions have a worse prognosis.

Treatment

The treatment of melanoma is primarily surgical. The indication for procedures such as lymph node dissection, superficial parotidectomy, and resection of distant metastases have changed somewhat over time, but the only hope for cure and the best treatment for regional control and palliation remains surgery (Fig. [13-12](#)). Radiation therapy, regional and systemic chemotherapy, and immunotherapy are effective in a limited set of circumstances, but none is a first-line option.

All suspicious lesions should undergo excisional biopsy. A 1-mm margin of normal skin is taken if the wound can be closed primarily. If removal of the entire lesion creates too large a defect, then an incisional biopsy of a representative part is recommended. Biopsy incisions should be made with the expectation that a subsequent wide excision of the biopsy site may be done. Once a diagnosis of melanoma is made, the biopsy scar and any remains of the lesion need to be removed to eradicate any remaining tumor. For in situ lesions a 0.5- to 1-cm margin of normal skin is adequate for cure. A T1 melanoma (less than 0.76 mm deep) requires a 1-cm margin to prevent local recurrence. For thicker lesions a 2- to 3-cm margin is recommended. Any wider margin of resection does not decrease local recurrence rates. The surrounding tissue should be removed down to the fascia to remove all lymphatic channels. If the deep fascia is not involved by the tumor, removing it does not affect recurrence or survival rates, so the fascia is left intact. If the defect cannot be closed primarily, a skin graft or flap is used.

All clinically positive lymph nodes should be removed by regional nodal dissection. If possible, the lymphatics between the lesion and the regional nodes are removed in continuity. Leaving tumor behind results in recurrence of lesions that cause great morbidity. When groin lymph nodes are removed, the deep (iliac) nodes must be removed along with the superficial (inguinal) nodes, or disease will recur in that region. For axillary dissections the nodes medial to the pectoralis minor muscle must also be resected. For lesions on the face, anterior scalp, and ear, a superficial parotidectomy to remove parotid nodes and a modified neck dissection is recommended. Disruption of the lymphatic outflow does cause significant problems with chronic edema, especially of the lower extremity.

Treatment of regional lymph nodes that do not obviously contain tumor in patients without evidence of metastasis (Stage I and II) is determined by considering the possible benefits of the procedure as weighed against the risks. In patients with thin lesions (less than 0.75 mm) the tumor cells are still localized in the surrounding tissue, and the cure rate is excellent with wide excision of the primary lesion; for these patients treatment of regional lymph nodes is not beneficial. With very thick lesions (more than 4 mm), it is highly likely that the tumor cells have already spread to the regional lymph nodes and distant sites. Removal of the lymph nodes has no effect on survival. Most of these patients die of metastatic disease before developing problems in regional nodes. Because there are significant morbid effects of lymphadenectomy, most surgeons defer the procedure until clinically evident disease appears. Approximately 40 percent of these patients eventually develop disease in the lymph nodes and require a second palliative operation. Elective lymphadenectomy is sometimes performed in these patients as a staging procedure before entry into clinical trials.

In patients with intermediate-thickness tumors (T2 and T3, 0.76 to 4.0 mm) and no clinical evidence of nodal or metastatic disease, the use of prophylactic dissection (elective lymph node dissection on clinically negative nodes) is controversial. To date, prospective, randomized studies have not definitively demonstrated that elective lymph node dissection improves survival in patients with intermediate-thickness melanomas. Careful examination of specimens in patients undergoing elective lymph node dissection have found that in 25 to 50 percent of the cases, specimens contain micrometastases. Among patients who do not have an elective lymph node dissection, 20 to 25 percent eventually develop clinically evident disease and require lymphadenectomy. More evidence suggests that there may be improved survival with elective lymph node dissection in patients with higher risk of developing metastasis (i.e., lesions with ulceration or on the trunk, head, and neck). The most compelling argument for the potential benefits of elective lymph node dissection comes from evidence in large clinical trials; patients with intermediate-thickness melanomas without elective node dissection, continue to die of the disease 10 years later, whereas patients who had an elective lymph node dissection do not. Although not yet statistically significant, these differences may become significant in the future.

One surgical modality gaining acceptance is the use of intraoperative methods to locate the pattern of lymphatic drainage from the primary lesion. Vital blue dye or a radioisotope is injected at the site of the melanoma. This permits identification of the first (sentinel) lymph node draining the tumor. The node is removed, and if micrometastases are identified in frozen-section examination, a complete lymph node dissection is performed.

When the sentinel node can be identified, it serves as an accurate indication of the status of the rest of the nodes in the region. This method may be used to identify patients who would benefit from lymph node dissection while sparing others an unnecessary operation. Whether this procedure actually improves survival in these patients awaits the results of clinical trials.

When patients develop distant metastases surgical therapy may be indicated. Solitary lesions in the brain, gut, or skin that are symptomatic should be excised when possible. Although cure is extremely rare, the degree of palliation can be high and asymptomatic survival prolonged. A decision to operate on metastatic lesions must be made after careful deliberation with the patient.

The most promising area of melanoma treatment is in the use of immunologic manipulation. The only adjuvant therapy known to influence survival so far is the use of intravenous interferon alpha-2b (INF alpha-2b) in patients with lesions of 4 mm or more (T4) or nodal metastasis (N1). In these patients both the relapse-free interval and overall survival are improved with use of INF alpha-2b. Toxicity while under therapy was a problem; the majority of the patients required modification of the initial dosage, and 24 percent discontinued treatment. These recent findings are encouraging because trials of adjuvant therapy in the past have never demonstrated a beneficial effect.

Vaccines have been developed with the hope of stimulating the body's own immune system against the tumor. Melanoma cells contain a number of distinctly different cell-surface antigens. Monoclonal antibodies have been raised against these antigens. These antibodies have been used alone or linked to a radioisotope or cytotoxic agent in an effort to selectively kill tumor cells. All treatments are currently investigational.

Although initially thought to be ineffective in the treatment of melanoma, radiation therapy has been shown to be useful. High-dose-per-fraction radiation produces a better response rate than low-dose, large-fraction therapy. Radiation therapy is the treatment of choice for patients with symptomatic multiple brain metastases.

Hyperthermic regional perfusion of the limb with a chemotherapeutic agent (e.g., melphalan) is the treatment of choice for patients with local recurrence or in-transit lesions (local disease in lymphatics) on an extremity and not amenable to excision. The goal of regional perfusion therapy is to increase the dosage of the chemotherapeutic agent to maximize tumor response while limiting systemic toxic effects. While difficult to perform and associated with complications, it does produce a high response rate (greater than 50 percent). The introduction of tumor necrosis factor alpha and interferon-g to the limb perfusate may increase the tumor response rate without increasing toxicity. Prospective clinical trials are under way to evaluate the use of regional perfusion for melanoma of the limbs as adjuvant therapy for patients with Stage I disease. In addition, regional perfusion therapy for metastatic disease to the liver is under investigation.

Pathologic Conditions Associated with Skin Malignancies

There are several well-recognized diseases associated with an increased incidence of skin malignancies. Some are associated with a specific neoplasm, whereas others appear to have the less specific effect of leaving the patient susceptible to a variety of neoplasms.

Diseases linked with basal cell carcinoma include the basal cell nevus syndrome and nevus sebaceus of Jadassohn. Basal cell nevus syndrome is an autosomal dominant disorder characterized by the growth of hundreds of basal cell carcinomas during young adulthood. Palmar and plantar pits are a common physical finding and represent foci of neoplasms. Treatment is limited to excision of only aggressive and symptomatic lesions. Nevus sebaceus of Jadassohn is a lesion containing several cutaneous tissue elements that develops during childhood. This lesion is associated with a variety of neoplasms of the epidermis, but most commonly basal cell carcinoma.

Diseases associated with squamous cell carcinoma may have a causative role in the development of carcinoma. Skin diseases that cause chronic wounds, such as epidermolysis bullosus and lupus erythematosus, are associated with a high incidence of squamous cell carcinoma. Epidermodysplasia verruciformis is a rare autosomal recessive disease associated with infection with human papillomavirus. Large verrucous lesions develop early in life and often progress to invasive squamous cell carcinoma in middle age.

Xeroderma pigmentosum is an autosomal recessive disease associated with a defect in cellular repair of DNA damage. The inability of the skin to correct DNA damage from ultraviolet radiation leaves these patients prone to cutaneous malignancies. Squamous cell carcinomas are most frequent, but basal cell carcinomas, melanomas, and even acute leukemias are seen.

Dysplastic nevi may represent a precursor to melanoma. Familial dysplastic nevus syndrome is an autosomal dominant disorder. Patients develop multiple dysplastic nevi, and longitudinal studies have demonstrated an almost 100 percent incidence of melanoma. Familial dysplastic nevus syndrome is similar to familial polyposis coli and the association with colon cancer. While the development of colon cancer can be arrested with total proctocolectomy, a similar solution is not possible with familial dysplastic nevi. Close surveillance and frequent biopsy of all suspicious lesions constitutes the best therapy.

Other Malignancies

Merkel Cell Carcinoma (Primary Neuroendocrine Carcinoma of the Skin)

Originally thought to be a variant of squamous cell carcinoma, it has only recently been demonstrated by immunohistochemical markers that Merkel cell carcinomas are of neuroepithelial differentiation. These tumors are associated with a synchronous or metasynchronous squamous cell carcinoma 25 percent of the time. These tumors are very aggressive, and wide local resection with 3-cm margins is recommended. Local recurrence rates are high, and distant metastases occur in one-third of patients. Prophylactic regional lymph node dissection and adjuvant radiation therapy are recommended. Overall, the prognosis is worse than for malignant melanoma.

Extramammary Paget's Disease

This tumor is histologically similar to the mammary type. It is a cutaneous lesion that appears as a pruritic red patch that does not resolve. Biopsy demonstrates classic Paget's cells. Paget's disease is thought to be a cutaneous extension of an underlying adenocarcinoma, although an associated tumor cannot always be demonstrated.

Adnexal Carcinomas

This group includes apocrine, eccrine, and sebaceous carcinomas, all rare tumors. They are locally destructive and can cause death by distant metastasis.

Angiosarcomas

Angiosarcomas may arise spontaneously, mostly on the scalp, face, and neck. They usually appear as a bruise that spontaneously bleeds or enlarges without trauma. Tumors also may arise in areas of prior radiation therapy or in the setting of chronic lymphedema of the arm, such as after mastectomy (Stewart-Treves syndrome). The angiosarcomas that arise in these areas of chronic change occur decades later. The tumors consist of anaplastic endothelial cells surrounding vascular channels. While total excision of early lesions can provide occasional cure, the prognosis usually is poor, with 5-year survival rates under 20percent. Chemotherapy and radiation therapy are used for palliation.

Kaposi's Sarcoma

Kaposi's sarcoma (KS) appears as rubbery bluish nodules that occur primarily on the extremities but may appear anywhere on the skin and viscera. These lesions are usually multifocal rather than metastatic. Histologically the lesions are composed of capillaries lined by atypical endothelial cells. Early lesions may resemble hemangiomas, while older lesions contain more spindle cells and resemble sarcomas.

Classic KS is seen in people of Eastern Europe or sub-Saharan Africa. The lesions are locally aggressive but undergo periods of remission. Visceral spread of the lesions is rare, but a subtype of the African variety has a predilection for spreading to lymph nodes. A different variety of KS has been described for people with AIDS or with immunosuppression from chemotherapy. For reasons not yet understood, AIDS-related KS occurs primarily in male homosexuals and not in intravenous drug abusers or hemophiliacs. In this form of the disease, the lesions spread rapidly to the nodes, and the gastrointestinal and respiratory tract often are involved. Development of AIDS-related KS may be associated with concurrent infection with a herpes-like virus.

Treatment for all types of KS consists of radiation to the lesions. Combination chemotherapy is effective in controlling the disease, although most patients develop an opportunistic infection during or shortly after treatment. Surgical treatment is reserved for lesions that interfere with vital functions, such as bowel obstruction or airway compromise.

Dermatofibrosarcoma Protuberans

Dermatofibrosarcoma protuberans consists of large nodular lesions located mainly on the trunk. They often ulcerate and become infected. With enlargement the lesions become painful. Histologically the lesions contain atypical spindle cells, probably of fibroblast origin, located around a core of collagen tissue. Sometimes they are mistaken for an infected keloid. Metastases are rare, and surgical excision can be curative. Excision must be complete because local recurrences are common.

Fibrosarcoma

Fibrosarcomas are hard, irregular masses found in the subcutaneous fat. The fibroblasts appear markedly anaplastic with disorganized growth. If they are not excised completely, metastases usually develop. The 5-year survival rate after excision is about 60 percent.

Liposarcoma

Liposarcomas arise in the deep muscle planes and, rarely, from the subcutaneous tissue. They occur most commonly on the thigh. An enlarging lipoma should be excised and inspected to distinguish it from a liposarcoma. Wide excision is the treatment of choice, with radiation therapy reserved for metastatic disease.

FUTURE DEVELOPMENTS IN SKIN SURGERY

The major challenge in surgical therapy for diseases of the skin is in the lack of replacement for diseased or damaged tissue. The development of autologous skin grafting for treatment of skin defects was a tremendous advancement. However, technical limitations, such as graft contraction and donor site problems, and biologic limitations, such as the limited amount of autologous skin available, make autografts less than a universal solution. The future of surgical therapy for diseases of the skin lies in the development of skin replacement. Current research is directed at tissue expansion, cell culture expansion, and neogenesis of skin.

Techniques for tissue expansion have been reported since 1982. During skin expansion with subcutaneous balloon implants (Fig. 13-13), new epidermis and some collagen is produced. Much of this new tissue, however, is rearrangement of the old tissue. Expansion of skin produces a limited amount of tissue for use.

The expansion of epidermis by the growth and maturation of keratinocytes in culture can be performed. The use of a postage-stamp-sized biopsy specimen to produce enough autologous epithelium to cover a burn area of more than 90 percent of total body surface area has been reported (Fig. 13- 14). Although this was a major advancement in covering large wounds, the final results are less than optimal. The cultured epidermis often blisters and sloughs, and wound contractures are common. Skin (comprising dermis, vasculature, adnexal structures, and pigmentation) is much more complex than just epidermis, and replacement of these other structures is under investigation.

Dermal replacements from synthetic materials or cadaveric sources are in clinical use. A bilaminar collagen and proteoglycan dermis (Integra) has been approved by the Food and Drug Administration for clinical use. This prosthetic dermis, available in ready-to-use form, can be used to cover large surface areas, decreasing fluid losses through the wound, which is crucial in burn patients. Vascularization of this dermis takes 2 weeks, and final epidermal coverage of the wound requires a thin skin graft. The final result is functionally and aesthetically quite good. Despite its limitations, it is the first promising dermal replacement to be widely used. Autologous skin grafts revascularize in 2 to 4 days because they already contain a network of capillaries in the dermis. For synthetic dermis to survive as well as an autologous graft, the problem of delayed vascular ingrowth must be solved. Pigmentation problems also may be addressed by repopulating skin replacements with cultured melanocytes; however, melanocyte biology is very complex, and pigmentation control is still rudimentary. With more sophisticated methods of tissue culture, a more complex skin replacement will become available.

As investigators learn more about the protein factors that control wound healing and tissue growth, the replacement for damaged skin will eventually come from complete neoorganogenesis of tissue. Characterization of these growth factors on a structural and functional level is just beginning, but the information obtained to date has been substantial. Factors have been isolated that cause specific mesenchymal cells to proliferate (fibroblast growth factor), migrate (epidermal growth factor), and organize into structures such as capillaries (transforming growth factor-beta) or even rudimentary organoid tissue. This may allow generation of new tissue in situ for skin replacement.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 14 - Breast

Kirby I. Bland
Michael P. Vezeridis
Edward M. Copeland III

INTRODUCTION

The breast or mammary gland is a distinguishing feature of the class Mammalia. From puberty to death, the breast is subjected to constant physical and physiologic alterations that are related to menses, pregnancy, gestation, and menopause. The impact of breast disease in Western societies assumes greater importance as the incidence of breast cancer continues to increase steadily. One of every two women will consult her physician for breast disease, approximately one of every four women will undergo breast biopsy, and one of every nine American women will develop some variant of breast carcinoma.

EMBRYOLOGY

The breast is a highly modified sudoriferous gland that develops as ingrowths from ectoderm form the alveoli and ducts. Supporting vascularized connective tissue is derived solely from mesenchyme. At the fifth or sixth week of fetal development, two ventral bands of thickened ectoderm (mammary ridges, “milk lines”) are evident in the embryo. In the majority of the class Mammalia, paired glands develop along these ridges and extend from the base of the forelimb (future axilla) to the region of the hind limb (inguinal area). These ridges are not prominent in the human embryo and disappear shortly thereafter, except for a small portion that may persist in the pectoral region. Accessory mammary glands (polymastia) or accessory nipples (polythelia) may occur along the original mammary ridge or milk line (Fig. 14-1) if the normal regression fails.

Each mammary gland develops as an ingrowth of ectoderm and initiates a primary bud of tissue in underlying mesenchyme. Each primary bud initiates the development of 15 to 20 secondary buds or outgrowths. In the fetus, epithelial cords develop from the secondary buds and extend into the surrounding connective tissues of the chest wall. Lumina develop in the outgrowths to form lactiferous ducts with prominent branches. By birth, lactiferous ducts open into shallow epithelial depressions referred to as the mammary pit. In infancy, the pit becomes elevated and transformed into the nipple as a consequence of proliferation of mesenchyme. If there is failure of the pit to elevate above skin level, a congenital malformation, recognized in 2 to 4 percent of patients as inverted nipples, is evident.

In newborn females or males, transient enlargement of the breast bud may be evident, and the bud may produce a secretion referred to as “witch’s milk.” These transitory changes occur in response to maternal hormones that cross the fetal-maternal circulation of the placenta.

At birth the breasts appear essentially identical in both sexes and demonstrate only the presence of major lactiferous ducts. The gland remains undeveloped in the female until puberty. Thereafter the organ enlarges rapidly in response to estrogen and progesterone secretion by the ovaries. Hormonal stimulation initiates proliferation of glandular tissue as well as fat and connective tissue elements associated with breast support. Glandular tissues remain incompletely developed until pregnancy occurs. With parturition, the intralobular ducts undergo rapid development and form buds that become alveoli.

Unilateral absence of the breast (amastia) is more common than bilateral amastia; both conditions occur more commonly in females. This rare congenital anomaly occurs as a result of an arrested mammary ridge around the sixth week of fetal development. Typically no other abnormalities are associated with bilateral absence of nipple and breast tissue. By contrast, Alfred Poland described the absence of musculature (pectoralis major and minor) of the shoulder girdle and malformations of the ipsilateral upper limb with unilateral amastia. Hypoplasia or complete absence of the ipsilateral breast or nipple, costal cartilage and rib defects, hypoplasia of subcutaneous tissues of the chest wall, and brachysyndactyly is referred to as Poland’s syndrome.

Breast hypoplasia also may be induced iatrogenically. The failure of complete development of the vestigial male or female breast (developmental hypomastia) may be initiated by therapeutic manipulation or injury to the mammary anlage in infancy or in the prepubertal state. Recognized iatrogenic mechanisms that initiate hypoplasia of the organ include trauma, abscess, incisions, infectious lesions, and radiation therapy.

Symmastia is the term, coined relatively recently, for medial confluence of the breast. This rare anomaly is recognized as webbing across the midline in breasts that usually are symmetrical. The presternal blending (confluence) of tissue that is associated with macromastia is more common.

Accessory or supernumerary nipples, or polythelia, is a relatively common, minor congenital anomaly that occurs in both sexes with an estimated frequency of 1:100 to 1:500 persons. Polythelia may be associated with abnormalities of the urinary tract (renal agenesis and carcinoma), abnormalities of the cardiovascular system (conduction disturbance, hypertension, congenital heart anomalies), and other conditions (pyloric stenosis, epilepsy, ear abnormalities, and arthrogryposis).

Supernumerary nipples or breasts may occur in any size or configuration along the mammary milk line, usually between the nipple and the symphysis pubis. Turner syndrome (ovarian agenesis and dysgenesis) and Fleischer syndrome (lateral displacement of nipples to the midclavicular line with bilateral renal hypoplasia) may have polymastia as a component. Accessory or ectopic axillary breast tissue is relatively uncommon but, when present, usually is bilateral.

ANATOMY AND DEVELOPMENT

Located within the superficial fascia of the anterior thoracic wall, the breast is composed of 15 to 20 lobes of glandular tissue of the tubuloalveolar type (Fig. 14-2). Fibrous connective tissues connect the lobes; adipose tissue is abundantly interposed between the lobules. Subcutaneous connective tissues surround the gland and extend as septa between lobes and lobules, providing structural support for glandular elements. The deep layer of superficial fascia lies on the posterior surface of the breast adjacent to and at some points fusing with the deep (pectoral) fascia of chest wall. The retromammary bursa may be identified surgically on the posterior aspect of the breast between the deep layer of superficial fascia and deep investing fascia of the pectoralis major and contiguous muscles of the thoracic wall. Fibrous bands of connective tissue interdigitate between parenchymal tissue to extend from the deep layer of the superficial fascia (hypodermis) and attach to the dermis of the skin. These suspensory ligaments (of Cooper) insert perpendicular to the delicate superficial fascial layers of the dermis and permit mobility of the breast while providing structural support.

The mature breast of the female extends from the level of the second or third rib to the inframammary fold at approximately the sixth or seventh rib. Transversely, it extends from the lateral border of the sternum to the anterior axillary or midaxillary line. The deep or posterior surface rests on portions of the deep investing fascia of the pectoralis major, serratus anterior, and external oblique abdominal muscles and the upper extent of the rectus sheath. The axillary tail (of Spence) extends superolaterally into the anterior axillary fold. The upper half of the breast, and particularly the upper outer quadrant, contains a greater volume of glandular tissue than do other sectors.

At maturity, glandular components of the breast take a protuberant conical form. The base of the cone is roughly circular, measuring 10 to 12 cm in diameter and 5 to 7 cm in thickness. Tremendous variations in size, contour, and density of the breast are evident at maturity. The nulliparous breast has a typical hemispheric configuration with distinct flattening above the nipple. By contrast, with multiparity and the hormonal stimulation that accompanies pregnancy and lactation, the organ assumes a larger and more pendulous form and increases in volume and density. With senescence, the aging breast assumes a flattened, flaccid, and more pendulous configuration with decreased volume.

Nipple and Areola

The epidermis of the nipple and areola is highly pigmented and variably corrugated. The complex is covered by keratinized stratified squamous epithelium. During puberty, the skin becomes increasingly pigmented and the nipple assumes an elevated, prominent configuration. During pregnancy, the areola enlarges and pigmentation is enhanced. Smooth muscle bundle fibers arranged radially and circumferentially in the dense connective tissue and longitudinally along the lactiferous ducts extend upward into the nipple. These muscle fibers are responsible for erection of the nipple, which occurs with various sensory and thermal stimuli.

The areola contains sebaceous glands, sweat glands, and accessory areolar glands. These accessory glands produce small elevations on the surface of the areola (Montgomery tubercles). The tip of the nipple contains numerous sensory nerve cell endings and Meissner’s corpuscles in the dermal papillae; the areola contains few of these structures. The rich sensory innervation of the breast, particularly the nipple and areola, is of great functional importance because the sucking infant initiates a chain of neurohumoral events that result in “milk letdown.”

Inactive Mammary Tissue

The tubuloalveolar glands derived from modified sweat glands of the epidermis lie in the subcutaneous tissues. Each of the 15 to 20 irregular lobes of branched tubuloalveolar glands in the adult terminates in a lactiferous duct (2 to 4 mm in diameter), which opens into a constricted orifice (0.4 to 0.7 mm in diameter) with entry into the ampulla of the nipple (see Fig. 14-2). Immediately under the areola, each duct has a dilated portion, the lactiferous sinus. Lined with stratified squamous epithelium, these ducts show a gradual transition to two layers of cuboidal cells, which then become a single layer of columnar or cuboidal cells in the remaining duct system. Myoepithelial cells of ectodermal origin reside between surface epithelial cells in the basal lamina. In the secretory portion of the gland and in the larger ducts, these cells contain myofibrils and are microscopically similar to smooth muscle cells.

In the inactive gland, the glandular component is sparse and consists chiefly of duct elements (Fig. 14-3). During menstruation, the breast undergoes cyclical changes. In early phases of the cycle, ductules appear as cords with sparse or absent lumina. With estrogen stimulation at or about the time of ovulation, secretory cells increase in height, lumina appear, and a small volume of secretions accumulates. Thereafter, fluid and lipids accumulate in connective tissue. In the absence of prolonged hormonal stimulation, the glandular components regress to a more inactive state throughout the remainder of the menstrual cycle.

Active Mammary Gland—Pregnancy and Lactation

With pregnancy and preparation for lactation, glands undergo marked proliferative and developmental maturation. As the breast enlarges in response to hormonal stimulation, lymphocytes, plasma cells, and eosinophils infiltrate and accumulate within fibrous components of connective tissue.

Development of glandular tissue is asymmetric; variation in degree of development may occur within a single lobule. With cellular division following mitotic phases, ductules branch and alveoli begin to develop. In the third trimester of pregnancy, alveolar development becomes more prominent (Fig. 14-4). With termination of pregnancy, proliferation declines and subsequent enlargement of the breasts occurs via hypertrophy of alveolar cells and accumulation of secretory products in the lumina of the ductules.

Secretory cells contain abundant endoplasmic reticulum, moderate numbers of large mitochondria, Golgi complexes, and a number of dense lysosomes. Two distinct substances

are produced by the cells and are released by different mechanisms: (1) the protein component of milk, synthesized in the granular endoplasmic reticulum; and (2) the lipid, or fatty, component of milk, which forms as free lipid droplets in the cytoplasm. These components of milk are formed by merocrine secretion (protein) and apocrine secretion (lipid).

Milk released in the first few days after parturition is termed *colostrum*. Colostrum has a low lipid content but contains considerable quantities of antibodies that are passively transferred from the mother to the fetus via the placenta. Lymphocytes and plasma cells infiltrate the stroma of the breast during proliferation and development and are believed to be the source of the components of colostrum. With reduction of these cellular structures, the production of colostrum is terminated and lipid-laden milk is released.

Blood Supply

The gland receives its principal blood supply from: (1) perforating branches of the internal mammary artery; (2) lateral branches of the posterior intercostal arteries; and (3) various branches from the axillary artery, including the highest thoracic, lateral thoracic, and pectoral branches of the thoracoacromial artery (Fig. 14-5). The second, third, and fourth anterior perforating arteries give off branches that arborize in the breast as medial mammary arteries. The lateral thoracic vessel gives off branches to the serratus anterior, pectoralis major and minor, and subscapularis muscles. It also gives rise to lateral mammary branches that invest lateral portions of the pectoralis major.

Veins of the breast follow the course of the arteries; primary venous drainage is toward the axilla. Three principal groups of veins for drainage of the thoracic wall and breast are: (1) perforating branches of the internal thoracic vein; (2) tributaries of the axillary vein; and (3) perforating branches of the posterior intercostal veins. Lymphatics usually parallel the course of blood vessels.

The vertebral venous tributaries (Batson's plexus) may provide a secondary route for metastases of breast cancer. This plexus invests the vertebrae and extends from the base of the skull to the sacrum. Venous channels exist between this plexus and veins associated with thoracic, abdominal, and pelvic organs. These potential pathways explain metastases to the vertebrae, skull, pelvic bones, and central nervous system in the absence of pulmonary metastases.

Innervation of the Breast

Lateral and anterior cutaneous branches of the second through sixth intercostal nerves provide sensory innervation. Nerves of the breast are principally derived from the fourth, fifth, and sixth intercostal nerves. A limited area of skin over the upper portion of the breast is supplied by nerves that arise from the cervical plexus, specifically the anterior or medial branches of the *supraclavicular nerve*.

Lateral branches of the *intercostal nerves* exit the intercostal spaces through slips of the serratus anterior muscle. These nerves supply the anterolateral thoracic wall; the third through sixth branches, also known as *lateral mammary branches*, supply the breast. The *intercostal brachial nerve* is the lateral branch of the second intercostal nerve and is commonly visualized during surgical dissection of the axilla. Resection of the intercostal brachial nerve causes loss of sensation from the upper medial aspect of the arm and axilla.

Lymphatic Drainage

The boundaries of lymphatic drainage of the axilla are not well demarcated. There is also considerable variation in the positions of regional nodes. Although anatomists usually define five groups of axillary nodes, surgeons identify six primary groups (Figs. 14-6 and 14-7). (1) The *axillary vein group*, or *lateral group*, consists of four to six nodes medial or posterior to the vein; these receive most of the lymph drainage from the upper extremity. (2) The *external mammary group* (anterior or pectoral group) harbors five or six nodes along the lower border of the pectoralis minor muscle contiguous with the lateral thoracic vessels. This group receives the majority of lymphatic drainage from the lateral breast. (3) The *scapular group* (*posterior* or *subscapular*) consists of five to seven nodes from the posterior wall of the axilla at the lateral border of the scapula and is contiguous with the subscapular vessels. These nodes receive lymph principally from the lower posterior neck, posterior trunk, and posterior shoulder. (4) The *central group* consists of three or four large groups of nodes that are embedded in the fat of the axilla immediately posterior to the pectoralis minor muscle. This group receives lymph from the three preceding groups but may receive lymphatics directly from the breast. (5) The *subclavicular group* (*apical*) consists of 6 to 12 nodal groups posterior and superior to the upper border of the pectoralis minor muscle. This group receives lymph from all groups of axillary nodes and unites with efferent vessels from the subclavicular nodes to form the *subclavian trunk*. (6) The *interpectoral* (*Rotter's*) *group* consists of one to four nodes interposed between the pectoralis major and minor muscles. Lymph from these nodes passes directly into the central and subclavicular groups.

As indicated in Fig. 14-7, these nodal groups are assigned levels according to their relationship to the pectoralis minor muscle. Nodes located lateral to or below the lower border of the pectoralis minor are referred to as *Level I nodes* and include the *external mammary*, *axillary vein*, and *scapular groups*. Nodes located deep to or behind the pectoralis minor are referred to as *Level II nodes* and include the *central group*. Nodes located medial to or above the upper border of the pectoralis minor are *Level III nodes* and include the *subclavicular lymph node group*.

Lymph Flow

Metastatic disease of the breast occurs predominantly by routes that are extensive and arborize in multiple directions through skin and mesenchymal lymphatics. Lymphatic flow is unidirectional except in the pathologic state and has preferential flow from the periphery toward the right side of the heart. Preferential lymphatic flow toward the axilla is observed in lesions of the upper anterolateral chest. The lymphatics of the dermis are intimately associated with deeper lymphatics of underlying fascial planes; this fact explains the multidirectional potential for drainage of superficial breast neoplasms. Two accessory directions for lymphatic flow from breast parenchyma to nodes of the apex of the axilla are the *transpectoral* and *retropectoral* routes. *Interpectoral* (*Rotter's*) *nodes*, between the pectoralis major and minor, receive lymph that terminates in the apical (Level III) group. The retropectoral pathway drains the superior and internal aspects of the breast and similarly terminates at the apex of the axilla in the apical group.

Accessory pathways provide major lymphatic drainage by way of the external mammary and central axillary node groups (Levels I and II, respectively). Internal mammary lymphatic trunks eventually terminate in subclavian node groups (see Figs. 14-6 and 14-7). The presence of supraclavicular nodes (stage IV disease) results from lymphatic permeation and subsequent obstruction of the inferior and deep cervical groups of the jugular-subclavian confluence. The supraclavicular node group is the termination of efferent trunks from subclavian nodes of the internal mammary nodal group. Central and medial lymphatics of the breast pass medially and parallel to the course of major blood vessels to perforate the pectoralis major muscle and terminate in the internal mammary nodal chain. This also is a major pathway for metastatic spread of carcinoma into the systemic circulation.

Cross-communication from the interstices of connecting lymphatic channels for each breast provides ready access of lymphatic flow to the opposite axilla. Communicating dermal lymphatics to the contralateral breast account for occasional metastatic involvement of the opposite breast and axilla.

Lymphatic vessels that drain the breast occur in three interconnecting groups: (1) within the gland in interlobular spaces that parallel lactiferous ducts; (2) within glandular tissue and overlying skin of the central part of the gland beneath the areola (subareolar plexus); and (3) on the posterior surface of the breast, communicating with minute vessels that parallel the perimysium in deep fascia. Lymphatic vessels from deeper structures of the thoracic wall drain principally into parasternal, intercostal, or diaphragmatic nodes.

More than 75 percent of lymph from the breast passes to the axillary lymph nodes (see Figs. 14-6 and 14-7); the remainder flows into parasternal lymphatics. Although it has been suggested that parasternal nodes receive lymph principally from the medial aspects of the breast, vital- dye flow studies report that both the axillary and the parasternal lymphatic groups receive lymph from all quadrants of the breast.

PHYSIOLOGY

Mammary development and function are initiated by a variety of hormonal stimuli, including estrogen and progesterone, prolactin, oxytocin, thyroid hormone, cortisol, and growth hormone. Estrogen, progesterone, and prolactin have profound trophic effects that are essential to normal breast development and function. Estrogen initiates ductal development; progesterone is primarily responsible for differentiation of epithelial cells and lobular development. Progesterone also may reduce estrogen binding in mammary epithelium and limit tubular system proliferation. Prolactin is the primary hormonal stimulus for lactogenesis in late pregnancy and in the postpartum period. Prolactin increases the number of estrogen receptors and stimulates epithelial cells to act synergistically with ductular and lobuloalveolar development.

Figure 14-8 depicts secretion of neurotrophic hormones from the hypothalamus that are responsible for regulation of secretion of mammogenic hormones. Secretion of the gonadotropins—luteinizing hormone (LH) and follicle-stimulating hormone (FSH)—regulates the ovarian secretory release of estrogen and progesterone. Release of LH and FSH from basophilic cells of the anterior pituitary gland are further regulated by secretion of *gonadotropin-releasing hormone (GnRH)* from the hypothalamus. Secretion of LH, FSH, and GnRH is regulated by positive and negative feedback effects of circulating levels of estrogen and progesterone.

The secretion of these mammogenic hormones throughout the life of the normal female is responsible for alterations in the hormonal milieu and for development, function, and maintenance of lobuloalveolar tissues (Fig. 14- 9). In the human female *neonate*, plasma estrogen and progesterone levels decrease after birth. Throughout childhood these levels remain low as a result of the regulatory sensitivity of the hypothalamic-pituitary axis to the negative feedback effects of sex steroids. With onset of *puberty* an increase in the central drive of the hypothalamus occurs, with a concurrent decrease in sensitivity to negative feedback by estrogen and progesterone. Thereafter, an increase in sensitivity to positive feedback by estrogen is evident. These physiologic events thereby initiate an increase in GnRH secretion, an increase in FSH and LH secretion, and ultimately an increase in ovarian estrogen and progesterone secretion. With the development of positive feedback by estrogen, the *menses* are initiated.

Cyclic Changes During the Menstrual Cycle

There are great variations in breast volume during the menstrual cycle. Volume is greatest in the second half of the cycle, after a premenstrual increase in size, nodularity, density, and sensitivity. Progesterone may stimulate glandular growth in the luteal phase. Changes in the mitotic rate of glandular components are greater in the luteal phase than in the follicular phase. The premenstrual increase in volume occurs as a consequence of the increase in size of the lobule without any evidence of epithelial proliferation. Thereafter, engorgement of the stroma, lobules, and ducts is evident, with an increase in the size of ducts and acini as the lumina dilate. Parenchymal engorgement and edema subside with onset of menses.

Pregnancy

A dramatic increase in secretion and release of circulating ovarian and placental estrogens and progestins is evident with pregnancy. These hormones initiate striking alterations in the form and substance of the breast (Fig. 14-9 B). The gland enlarges, the areolar skin darkens, and the areolar glands become prominent as ducts and lobules proliferate. In the *first trimester*, lobuloalveolar formation is initiated as ducts branch to form multiple alveoli. With increases in lobular size, proliferating glandular epithelium replaces connective tissue and the components of adipose tissue. In the *second trimester*, proliferation of ductular elements increases after stimulation by estrogens and progestins secreted by the placenta. These sex steroids cause arborization of glandular structures to further develop alveoli. As this glandular system enlarges, the secretory capacity of the epithelium increases, as is evident by the accumulation of colloid in the alveoli.

During the *third trimester*, fat droplets accumulate in the alveolar cells and colostrum fills the alveolar and ductular spaces. Mammary blood flow increases and myoepithelial cells hypertrophy. The mammogenic action of prolactin requires the presence of cortisol, insulin, growth hormone, and epidermal growth factor. In late pregnancy, limited synthesis of milk fats and proteins is initiated. This process is stimulated by the lactogenic effects of prolactin on breast lobular tissue; other pituitary lactogenic hormones may also have trophic effects.

Postpartum Lactation

After delivery of the placenta, progesterone and estrogen levels diminish. These quantitative decreases in the plasma estradiol and progesterone levels allow full expression of the lactogenic action of prolactin. Maintenance of lactation requires regular removal of milk and stimulation (“milk letdown”) of the neural reflexes to prolactin secretion. The magnitude of the suckling-induced surge of prolactin decreases with time, probably as a consequence of the decreased duration and frequency of nursing. Milk production and ejection in nursing women are controlled by neural reflex arcs that originate in free nerve endings of the nipple-areolar complex.

Oxytocin initiates contraction of smooth muscle components of myoepithelial cells that surround the alveoli; compression of the alveoli occurs, and expulsion of milk under pressure into the lactiferous sinuses is evident. Oxytocin release can result from auditory, visual, olfactory, or other stimuli associated with nursing. Maintenance of lactation requires an intact hypothalamic-pituitary axis, adequate diet and nutrition, regular suckling release of milk, and the absence of psychologic stresses that interfere with normal control of prolactin and oxytocin release.

After the weaning of the infant, the gland returns to an inactive, nonsecretory state. Prolactin and oxytocin release subside. The secretory activity of the lactogenic epithelium decreases, and unremoved dormant milk increases pressure within the ductular and alveolar structures. The lobular structure thereafter atrophies and the secretory cells degenerate (Fig. [14-9 C](#)).

Postmenopausal Breast

After menopause and the concomitant decrease in ovarian secretion of estrogen and progesterone, there is a progressive involution of ductular and glandular components. A decrease in the number and size of glandular elements is evident; the epithelium of the lobules and ducts becomes atrophic or hypoplastic. Surrounding fibrous tissue increases in density, and the parenchyma is replaced with adipose and stromal tissue rather than supporting glandular structure. With aging there is loss of fat content and the supporting stroma, thereby initiating loss of lobular structure, density, form, and contour (Fig. [14-9 D](#)).

Gynecomastia

Gynecomastia implies the presence of a female-type mammary gland in the male. Most examples of gynecomastia should not be considered a disease, because enlargement of the male breast is common. *Physiologic gynecomastia* occurs mostly during three phases of life: (1) neonatal period, (2) adolescence, and (3) senescence. Common to each is an excess of estrogens in relation to circulating testosterone. Neonatal physiologic gynecomastia is caused by the action of placental estrogens on neonatal breast parenchyma. In adolescence there is an excess of estradiol relative to testosterone. With aging, the plasma testosterone level falls, and senescent gynecomastia is caused by a relative hyperestrinism.

Pathology

The few ductal structures of the male breast enlarge, elongate, and branch with an ensheathing connective tissue. There is a combined increase in glandular and stromal elements, with regular distribution of each element throughout the enlarged breast. In the pubertal male, the condition often is unilateral and typically occurs between the ages of 12 and 15 years. By contrast, senescent gynecomastia usually is bilateral, although there may be asymmetry.

In the nonobese patient, at least 2 cm of subareolar breast tissue must be present before gynecomastia can be confirmed. Mammography and ultrasonography are used to differentiate indistinguishable or ill-defined contiguous fatty tissue from male breast lesions and soft-tissue structures. Dominant nontender masses and local areas of firmness, irregularity, or asymmetry suggest the possibility of an early male breast cancer in the aging patient. Gynecomastia does not predispose the male breast to the development of cancer. By contrast, the hypoandrogenic state of primary testicular failure in Klinefelter’s syndrome (47,XXY) is associated with a higher risk for breast cancer in men.

Pathophysiology

Table [14-1](#) identifies the pathophysiologic mechanisms that may initiate gynecomastia. *Estrogen excess states* result from an increase in the secretion of estradiol from the testicles or from nontesticular tumors. Endocrine disorders, such as hyperthyroidism or hypothyroidism, or hepatic disease (nonalcoholic and alcoholic cirrhosis) may initiate estrogen excess. Estrogen excess states also may be induced by nutritional alterations such as protein and fat starvation. “Refeeding gynecomastia” is perhaps related to the resumption of pituitary gonadotropin secretion after pituitary shutdown.

Androgen deficiency states such as aging initiate gynecomastia. Concurrent with decreased plasma testosterone levels is an elevation in the level of plasma testosterone-binding globulin, resulting in a reduction of unbound testosterone. Senescent hypertrophy occurs most commonly in men between the ages of 50 and 70 years. Klinefelter’s syndrome of 47,XXY karyotype is manifested by gynecomastia, hypergonadotropic hypogonadism, and azoospermia. Other causes of primary testicular failure include ACTH deficiency, hereditary defects of androgen biosynthesis, and eunuchoidal males (congenital anorchia).

Secondary testicular failure as a cause of gynecomastia may result from trauma, orchitis, cryptorchidism, abdominal or genital irradiation, hydroceles, varicoceles, and spermatoceles. Renal failure, regardless of cause, may initiate gynecomastia.

Drugs with estrogenic or estrogen-related activity (digitalis, estrogens, anabolic steroids, marijuana) may be causative. Drugs that inhibit the action or synthesis of testosterone (cimetidine, ketoconazole, phenytoin, spironolactone, antineoplastic agents, diazepam) also may be implicated. Drugs that enhance estrogen synthesis (human chorionic gonadotropin) or drugs with idiopathic mechanisms may induce gynecomastia (reserpine, theophylline, verapamil, tricyclic antidepressants, furosemide).

Treatment

Medical therapy of gynecomastia is rarely of value except when a specific diagnosis has been established. For disorders of androgen deficiency, testosterone administration may effect breast regression. For large, progressive gynecomastia refractory to drug discontinuance or therapy of an endocrine defect, the most effective therapy, especially in the young adult, is transareolar mastectomy. Surgical therapy is reserved, however, for idiopathic causes of gynecomastia in which exhaustive attempts to define endocrine, metabolic, or drug-related causes have failed. Attempts to reverse gynecomastia with danazol have been successful, although side effects from the androgenic properties of the drug are significant. Tamoxifen citrate, as therapy for benign breast disorders including gynecomastia, has had encouraging initial results.

DIAGNOSIS

Presentation

A lump in the breast is a common premenopausal and postmenopausal physical finding in the female. Up to one-half of patients presenting with breast complaints have no evidence of breast pathology; 65 percent or more of all breast lumps are discovered by the patient. In patients who commonly perform breast self-examination, more than 85 percent of definable lesions are detected by the patient. The patient also may note breast pain, but this symptom more commonly represents a proliferative benign breast disorder rather than carcinoma. Other presenting symptoms of breast cancer that occur less frequently include enlargement, nipple discharge, changes of the nipple, retraction or alterations of symmetry, ulceration, erythema, axillary mass, and infrequently bone or musculoskeletal discomfort. While many women recognize these symptoms, the delay in seeking medical attention persists.

Examination

The technique for examination of the breast should include inspection and palpation of the entire breast and draining lymph node sites. The clinician, standing in front of the patient, should first inspect the breast with the patient’s arms by her side (Fig. [14-10 A](#)); with her arms straight up in the air (Fig. [14-10 B](#)); and with her hands on her hips with and without pectoral muscle contraction. Symmetry, size, and shape of the breast should be recorded, as well as any evidence of edema (*peau d’orange*), nipple inversion or change, skin retraction, or erythema. With the arms extended forward and the patient in a sitting position, a forward lean accentuates skin retraction.

Palpation

All regions of concern in the breasts that were identified by inspection should be recorded and the entire breast mass should be carefully palpated.

Examination of the patient in the supine position (Fig. [14-10 C](#)) is best performed with the benefit of a pillow supporting the ipsilateral hemithorax. The examiner should gently palpate the breast from the ipsilateral side, making certain to examine all quadrants of the breast from the sternum to the clavicle, laterally to the latissimus dorsi muscle and inferiorly to the upper rectus sheath. The physician should perform the examination with the palmar aspects of the fingers; a grasping or pinching motion should be avoided. The breast may be cupped or molded in the examiner’s hands to check for retraction.

A systematic search for lymphadenopathy is essential if breast cancer is suspected. The position for examination of the axilla is indicated in Fig. [14-10 D](#). The shoulder girdle is stabilized by supporting the upper arm and elbow. Using gentle discrete palpation, all three levels of potential lymphatic enlargement in the axillary sites are assessed; this technique allows bimanual palpation of disease at the level of the pectoralis minor muscle. Careful palpation of subclavicular, supraclavicular, cervical, and parasternal sites is performed. A diagram of the chest and contiguous nodal sites is useful for recording location, size, and characteristics of any palpable disease (Fig. [14-11](#)). Size, consistency, shape, mobility, and fixation of any palpable breast mass or nodal site should be recorded.

Mammography

Mammography has been used in North America since the 1960s. Techniques for mammography have been modified and improved to enhance image quality, and proper use of mammography requires special techniques and film and a radiologist skilled in interpretation. Conventional mammography uses a low-dose film/screen technique that delivers as little as 0.1 cGy per study. By comparison, each chest roentgenogram delivers one-quarter of this radiation volume. There is no proven escalation of breast cancer risk related to low-dose irradiation with screening mammography. The benefits of mammography’s ability to detect a small cancer that is often curable far outweigh any theoretic risk.

Mammography should not be considered a substitute for biopsy; rather, this technique is an adjunctive, complementary study that supplements history and physical examination. The technique is useful for (1) examination of an indeterminate mass that presents as a solitary lesion that may be a neoplasm; (2) examination of an indeterminate mass that cannot be considered a dominant nodule, especially when multiple cysts or other vague masses are present and the indication for biopsy is uncertain; (3) follow-up examination of

breast cancer treated by segmental mastectomy and radiation therapy; (4) follow-up examination of the contralateral breast after segmental or total mastectomy; and (5) evaluation of the large, fatty breast in the symptomatic patient in whom nodules are not palpable (Fig. 14-12).

Additional value of diagnostic mammography lies in the early detection of an occult cancer before it reaches 5 mm. When abnormalities have been detected by the patient or the clinician, mammography can more precisely define the abnormality, detect multicentric disease, and identify the presence of synchronous cancers.

The presence of fine, stippled calcium in the radiogram of an occult or suspicious lesion is suggestive of cancer. Calcification occurs in one-third to one-half of nonpalpable cancers, but fewer than one-half of these calcific foci demonstrate the classical appearance suggestive of malignancy (Fig. 14-13). Fewer than one-half of the dominant mass lesions have spiculated or irregular margins, and approximately one-fifth of cancers are discovered by secondary signs such as architectural distortions, duct dilatation, asymmetry, and fibronodular densities. Microcalcification as a sign of malignancy assumes greater importance in younger women, in whom it may be the sole mammographic feature. The importance of calcification as a sign of breast cancer diminishes with age.

Careful analysis of direct and indirect signs is essential to document that the benign:malignant biopsy ratio is surgically acceptable. The positive predictive value of certain mammographic signs dramatically increases when parenchymal distortion, poorly defined mass lesions, typical malignant-type calcifications, and stellate opacities are evident. It is estimated that a skilled radiologist can detect cancer of the breast with a false-positive rate of approximately 10 percent and a false-negative rate of 6 to 8 percent.

The clinical impetus for screening mammography came from the Health Insurance Plan (HIP) study and the Breast Cancer Detection Demonstration Project (BCDDP). The HIP study demonstrated a reduction of 33 percent in mortality in patients who were screened by mammography; these data have been verified by the BCDDP. The BCDDP confirmed that mammography conducted in an optimal environment provided a true-positive rate that exceeded 90 percent and was significantly greater in accuracy than clinical examination for detection of occult or early tumors. In both studies, 80 percent of the patients with mammographically detected carcinomas had no axillary nodal metastases. These findings contrast significantly with patients whose breast cancer was detected clinically, in whom more than 50 percent have positive axillary nodes. Reports suggest that screening of breast cancers in women under 50 years of age allows earlier diagnosis and treatment of breast cancer. Disease-free, 5-year survival in this younger cohort exceeded 90 percent.

The current guidelines of the American Cancer Society recommend that all women initiate breast self-examination at the age of twenty and that a “baseline” mammographic examination be obtained at about 35 years of age after consultation with a physician. The patient should consult her physician regarding the need for regular mammographic screening between the ages of forty and fifty; annual mammographic examination should be conducted thereafter.

Prospective randomized studies of routine mammographic screening confirm a 40 percent reduction for stage II disease and more advanced cancers in the screened population; a corresponding 30 percent increase in survival was evident in patients found to have cancer. The enlarging population of patients with occult (nonpalpable) cancers detected in annual screening programs often necessitates needle-localization biopsy techniques. Postexcision confirmation of the suspicious mass or calcifications using specimen radiography is essential (Fig. 14-14); otherwise, histologic confirmation of benignity of the lesion cannot be assured. With the increasing availability of stereotactically guided biopsy for cytology and histology, the requirement for open breast biopsy may be reduced.

Xeromammography

This technique is identical with mammography except that the image is recorded on a xerographic plate rather than a conventional transparency. The image produced is positive rather than negative (see Fig. 14-12 B). Edge enhancement and wide recording latitude allow details of the soft tissues of the breast, chest wall, and thinner peripheral portions of the breast to be recorded with one exposure. Current opinion favors film screening techniques for routine breast radiographic examination.

Magnification Mammography

This technique enhances the sharpness of detail and increases diagnostic accuracy for breast cancer. The optimal magnification is 1.5 times life size; margins of breast masses and the degree and specificity of microcalcifications are clearly defined. This technique may significantly reduce the number of patients referred for biopsy.

Ultrasonography

Ultrasonography has no ionizing radiation, it is highly reproducible, and it has high patient acceptability. The importance of ultrasonography lies in the resolution of equivocal mammography, the diagnosis of cystic disease, and the demonstration of solid abnormalities with specific echogenic features. The resolution of ultrasound is inferior to high-resolution mammography, and lesions 1 cm in diameter, unless cystic, will not be detected. In the presence of a normal physical examination and mammogram, ultrasonographically demonstrated abnormalities are, in the majority of cases, not significant.

Ultrasonography is also useful for guiding the aspiration of cysts to provide cytologic specimens. Cysts, on ultrasound examination, are always well circumscribed, with smooth margins, and have an echo-free center irrespective of the sensitivity setting (Fig. 14-15). There are reported to be criteria to distinguish benign from malignant lesions on sonography but these lack specificity. Benign solid masses usually show smooth contours, round or oval shapes, with weak internal echoes and well-defined anterior and posterior margins. Malignant lesions have characteristically jagged, irregular walls; malignant lesions, however, may have smooth margins with acoustic enhancement (Fig. 14-16).

Doppler Flow Studies

Blood flow in malignant breast lesions is enhanced. Doppler flow signals can be used to detect increased flow and may further distinguish benign from malignant lesions. Malignant lesions produce signals of high frequency and amplitude with continuous flow through diastole. Although Doppler flow studies may play a role in distinguishing benign from malignant lesions, this technique is not of adequate sensitivity to dictate treatment.

Thermography

Transmission of detectable heat from the breast is nonspecific, and in malignant lesions results from the hypervascularity that frequently accompanies carcinoma. Three thermographic methods are used: telethermography, contact thermography, and computed tomography. Using special heat scanners it is possible to delineate these “hot” perfusion sites on film. Results are so variable and inaccurate, however, that its use was terminated in the BCDDP. Sensitivity is less than 50 percent and it is not advocated as a routine screening method, because it is unable to detect minimal breast cancer.

Light Scanning of the Breast

This is a noninvasive and relatively inexpensive technique, and therefore it has attracted considerable attention as a diagnostic screening modality. This technique utilizes electromagnetic waves that impinge on a transparent medium; thereafter, the wave is scattered and absorbed. Light attenuation varies considerably, depending on the biologic characteristics and the hemoglobin content of the tissue studied. The most consistent and important sign of breast cancer is light absorption with a decrease in luminescence displayed in a black and white mode. In the normal breast there should be symmetrical absorption of light. The sensitivity of light scanning for detection of breast cancer remains limited, however. With lesions smaller than 1 cm in diameter, sensitivity varies from 19 to 44 percent.

Magnetic Resonance Imaging (MRI)

MRI has great value in detection of vertebral body metastases and musculoskeletal pathology related to breast cancer. However, its value as a potential screening method in breast cancer is questionable. The breast is examined using conventional spin-echo T₁- and T₂-weighted sequences, preferably in the sagittal plane, but coronal and axial images can be obtained. Malignant tissue may be identified on MRI using the same morphologic criteria as mammography. Irregularly marginated spiculated masses, secondary skin changes, and enlarged glandular tissue are signs of malignancy. MRI spectroscopy may have a role in determining the treatment rationale for patients.

Interventional Techniques

Ductography

Ductograms are performed by injecting radiopaque contrast media into one or more of the mammary ducts and performing subsequent mammographic imaging. The primary indication for this technique is discharge from the nipple, particularly when the fluid is serosanguineous or bloody. The duct is gently enlarged with a dilator. A small, blunt cannula is inserted under sterile conditions into the nipple ampulla, and with the patient in a supine position, 0.1 to 0.2 mL of dilute contrast media is injected until the patient reports a sense of fullness. Craniocaudal and mediolateral mammographic views are obtained without compression. Intraductal papillomas can be demonstrated as small filling defects surrounded by contrast media (Fig. 14-17). Cysts may opacify after injection of the contrast when they communicate with ducts. Carcinomas may appear as irregular masses; invasive tumors appear as multiple intraluminal filling defects.

Localization of Nonpalpable Breast Masses

An integral part of the management of breast disease includes localization of nonpalpable (occult) breast lesions. Although surface localization and spot method identification are in use in many institutions, mammographically controlled placement of a hooked wire is a more accurate, state-of-the-art technique. Under local anesthesia, the stylet is accurately placed parallel to the chest wall using acrylic plastic compression plates that contain multiple holes, each of which allows passage of a stylet needle. After needle withdrawal, the lesion is accurately localized on mammogram, and the tiny wire hook is left in position (see Fig. 14-13 D). Thereafter, the surgeon can adequately excise the suspicious lesion, sampling minimal breast tissue. Great advances in localization and diagnosis have been made with the use of stereotactic needle placement with cytologic aspiration. This technique has allowed localization of occult breast cancer successfully in more than 90 percent of patients, and the sensitivity and specificity of the cytologic aspirate for occult breast cancer exceed 95 percent.

In all cases, needle localization of suspicious occult masses and excision of the abnormal tissue must be evaluated by specimen radiography to ensure that mammographically detected abnormalities have been removed (see Fig. 14-14). Before dissection of the specimen, a radiograph of the excised tissue should be obtained using conventional mammographic equipment or, for better resolution, a dedicated specimen radiographic unit. A specimen radiograph further directs the pathologist to the precise location of the abnormality in the tissues to ensure that appropriate sampling is obtained.

Stereotactic core needle biopsy has gained acceptance as a diagnostic alternative to needle localization biopsy in appropriate cases. The advantages of this technique in comparison to needle localization biopsy are lower incidence of complications, avoidance of breast disfigurement, shorter recovery period, and lower cost. In addition, when the mammographically detected lesion is a cancer and the diagnosis is made by stereotactic core needle biopsy, the definitive surgical management can be accomplished with only one visit to the operating room. A recent study showed that in patients who underwent excisions of cancers after establishment of the diagnosis with stereotactic core needle biopsy the surgical margins were negative in all patients, while in patients who had needle location biopsies the surgical margins were positive in 55 percent of cases. Thus reexcision was required in a substantial proportion of patients after needle localization biopsies in the study.

INFLAMMATORY AND INFECTIOUS DISORDERS

Bacterial Infection

Staphylococcus aureus and streptococci are the organisms most frequently recovered from nipple discharge in an active infection of the breast. Abscesses often are related to lactation and typically occur within the first few weeks of breast feeding. Figure [14-18](#) depicts the progression of an inflammatory process that may result in diffuse breast cellulitis with localized subcutaneous, subareolar, interlobular (periductal), retromammary, or unicentric and multicentric abscesses. Streptococcal infections initiate diffuse cellulitis without localization until a more advanced stage. *S. aureus* abscesses tend to be more localized, deeply invasive, and suppurative. The multilocular abscess evident in Fig. [14-18](#) is typically seen in staphylococcal infections. Diffuse cellulitis of streptococcal origin with lymphatic permeation is adequately treated with local wound care, including focal heat compresses and administration of intravenous antibiotics (e.g., penicillin or cephalosporin derivatives).

All superficial or deep abscesses with overgrowth of any bacterial organism present with point tenderness, erythema, and hyperthermia. This presentation necessitates immediate and adequate operative drainage of fluctuant areas to avoid sepsis. Immediate surgical drainage is essential for the advanced breast abscess. Thorough debridement of the abscess via circumareolar or multiple nonradial incisions paralleling Langer's lines is recommended.

Breast infections may be chronic, with recurrent abscess formation. Appropriate cultures must be taken for acid-fast bacteria, fungi, and anaerobic and aerobic bacteria. Uncommon organisms might be encountered, and long-term antibiotic therapy could be indicated. In extreme cases, simple mastectomy may be required to eradicate a severe chronic infection.

Puerperal (Lactational) Mastitis

Hospital-acquired puerperal infections of the breast are much less common with isolated deliveries and standardization of breast and infant hygiene. Nursing women may present with milk stasis, noninfectious inflammation, or infectious mastitis. Epidermic mastitis is initiated by highly variant strains of penicillin-resistant *S. aureus* that are transmitted via the suckling neonate. This variant of mastitis is associated with other neonatal staphylococcal infections and, in the untreated patient, may result in substantial morbidity and, occasionally, death. The patient may have pus expressed from the nipple of a very tender hyperemic breast. The infant must be rapidly weaned from breast feeding.

Nonepidemic (sporadic) puerperal mastitis refers to involvement of the interlobular connective tissue of the breast parenchyma. Such patients present with nipple fissuring and milk stasis, which initiate the secondary invasive retrograde bacterial infection. Emptying of the breast has been shown to shorten duration of symptoms and improve outcome, with a remarkable reduction in recurrence of the infectious mastitis. The addition of antibiotics will result in an excellent outcome for 96 percent of cases treated.

Discontinuance of lactation is essential to enable resolution of the inflammatory process. The use of breast suction pumps helps to empty stagnant milk ducts and central abscess collections. Continuation of lactation after removal of the suckling reflex may necessitate the use of intramuscular injections of stilbestrol or testosterone/estradiol derivatives (e.g., Deladumone).

Superficial and Deep Mycoses

Fungal infestations of breast parenchyma are rare and most commonly include blastomycosis and sporotrichosis. Preoperative diagnosis is rarely established.

Fungal infections are most commonly initiated by intraoral fungi inoculated into the breast parenchyma by the suckling infant. These fungal infections present as recurrent mammary abscesses juxtaposed to the nipple or areola. Pus may be expressed from the sinus tracts near the nipple and discharge may be mixed with blood.

Diagnosis of *mammary blastomycosis* is confirmed by collecting material from the abscess and demonstrating round budding organisms in a potassium hydroxide mount or a McMannus-stained smear. Staining of biopsy smears with periodic acid-Schiff reaction after digestion with diastases commonly identifies the mammary fungi of *sporotrichosis*. Serologic and sensitivity reactions may assist in the diagnosis.

Amphotericin B and stilbamidine derivatives are the most effective antifungal agents for the treatment of blastomycosis. Iodides continue to provide specific chemotherapy for the typical cutaneous variant of sporotrichosis but are ineffectual for systemic fungal disease. Amphotericin B is essential therapy for the disseminated and noncutaneous forms. Antifungal therapy, continued for months, may eliminate the necessity of surgical intervention. Drainage of the abscess, simple mastectomy, or quadrantectomy may be necessary to eradicate the fungal disorder.

Candida

Candida albicans affecting the breast often presents as a relatively innocuous variant of systemic mycoses. Patients demonstrate large quantities of the fungus in scrapings from the lesions or in the purulent discharge from elevated lesions that have scalloped borders with a definable margin of sodden scales. Scrapings contain an abundant quantity of filaments and budding cells. Skin tests and serologic reactions are of little value. Therapy requires removal of the predisposing factors causing maceration. Water-miscible combinations of nystatin or clotrimazole are recommended. Antibacterials and steroids should not be used.

Hidradenitis Suppurativa

Hidradenitis suppurativa of the breast areola or axilla occurs infrequently. This chronic inflammatory state originates within the large sebaceous glands (apocrine glands of Montgomery or the axilla) that are located on the epithelial surface. Patients with chronic acne have a propensity to develop hidradenitis.

When located in and about the nipple-areola complex, this chronic inflammatory state may mimic invasive carcinoma, Paget's disease of the nipple-areola, or other chronic benign inflammatory states. When confined to the nipple, the breast parenchyma is well preserved without demonstrable abscess or other lesions on mammogram. With compression of sinus tracts, scars, or pustules, purulent discharge can be expressed from sebaceous glands. The patient frequently is ill.

Therapy is directed at control of the inflammatory process by elimination of the infection in the apocrine glandular system. Excision may be accomplished under local or general anesthesia. Losses of larger areas of skin may be covered with advancement flaps from the ipsilateral breast or split-thickness skin grafts from noncontiguous sites.

Mondor's Disease

This variant of thrombophlebitis involves the superficial veins of the anterior chest wall and breast. In 1939 the classic description was provided by the French surgeon Henry Mondor. Typically this disorder occurs in an area contiguous with the breast and is detected as a thrombosed vein presenting as a tender, cordlike structure ("string phlebitis"). The cause is unknown. Surgical procedures, infectious processes, and stress-related exercise of the upper extremity, especially repetitive movements, may initiate the syndrome. Superficial veins of the anterior chest wall and abdomen that are commonly involved include the lateral thoracic vein, the thoracoepigastric vein, and, more rarely, the superficial epigastric vein.

This benign self-limited disorder is not indicative of a neoplasm. Typically the patient presents with acute pain in the lateral half of the breast or the anterior chest wall. A tender, firm cord follows the distribution of one of the three major superficial veins of the chest or abdominal wall. Only rarely does bilateral presentation occur, and most patients have no evidence of thrombophlebitis in other anatomic sites.

When the clinical diagnosis is uncertain, or if a contiguous mass is present near the fibrous cord, confirmation must be established by excisional biopsy. Therapy includes liberal use of salicylates and heat compresses along the distribution of symptomatic venous involvement. Restriction of motion of the ipsilateral extremity and shoulder and brassiere support are encouraged. The process usually resolves over 2 to 6 weeks. When symptoms persist or are refractory to therapy, division of the vein above and below the area of involvement or excision of smaller lesions is appropriate.

BENIGN LESIONS

Nonproliferative Lesions

Fibrocystic disease, preferably termed *fibrocystic disorder*, is an ill-defined entity. Patients present with diffuse, often bilateral breast pain. Palpation reveals multiple irregularities. In biopsy examination the specimens are found to contain "fibrocystic elements." Most lesions are not risk factors for development of cancer of the breast. The risk for cancer is increased only when there is associated dysplasia.

Patients presenting with discomfort or pain associated with multiple cystic lesions of the breast defined by palpation or mammography are managed conservatively. Pain is generally accentuated in the second half of the menstrual period and is diminished with the onset of menses. Analgesics usually control the pain, and, in some instances, diuretics are helpful to reduce the extent of fluid accumulation.

A single dominant cyst is usually identifiable by palpation of a smooth, rounded mass. The cystic nature can be substantiated by ultrasonography, but this is generally not required. The cyst should be aspirated, and if the aspirate is clear or cloudy greenish-gray in color, it can be discarded without sending a sample for cytopathologic evaluation. If the fluid is bloody, the cyst should be excised. Recurrent cysts can be treated by recurrent aspiration, and in some patients excision is carried out if the cyst persists after multiple aspirations and is symptomatic.

Cysts

Cysts are considered foremost among all of the benign histologic changes in the breast, as are apocrine lesions that commonly accompany cysts. The size of cysts varies from 1 mm to as great as several centimeters. Most cysts are lined by cells that harbor multiple mitochondria with secretory granules that appear pink by usual eosin staining. Cysts originate as lobular lesions in which the individual acini or terminal ductules dilate or unfold to produce solitary lobules that enlarge as a cystic mass. No consistent relationship between cysts and breast cancer risk has been established. The studies of Dupont and Page demonstrate a slightly higher risk for women with a family history of cysts and breast cancer compared to women with a family history of breast cancer alone. *Apocrine cytoplasmic alterations* assume minimal importance with regard to breast cancer risk. Wellings and Alpers suggested that apocrine changes in breasts are associated with cancer risk. Nonetheless, this is not considered an indicator of cancer risk in a predictive manner. Neither cysts nor apocrine changes are indicators of increased cancer risk in the absence of other established factors. *Epithelial hyperplasia*, which is related to an increased cancer risk, may coexist with cysts, and either change could be present without the other in individual biopsy specimens. It is for this reason that cysts and hyperplastic epithelial lesions should be pathologically distinguished.

Proliferative Lesions

The relationship between extensive *hyperplasia* and associated carcinoma is supported in many prospective studies. Proliferative breast disease is distinguished by epithelial hyperplasia, in which there is an increased number of cells (≥2 cell layers) above the basement membrane. *Atypical ductal hyperplasia* (ADH) lesions must be differentiated from carcinoma in situ. Mild hyperplasia is characterized by three or more cells above the basement membrane in a lobular unit or duct. Such lesions commonly are an “inflammatory” type, with separation of the epithelial cells by inflammatory components. *Moderate* and *florid hyperplasia* are found in more than 20 percent of biopsy specimens. Moderate and florid degrees of hyperplasia are clinically important because these lesions imply a slightly higher risk (1.5 to 2 times) for subsequent invasive carcinoma.

Atypical Hyperplasia

Table 14-2 summarizes the risks for development of invasive breast carcinoma on the basis of histologic findings. Atypical hyperplasia (AH) indicates a specific pattern that has atypia with an increased disposition for development of subsequent breast cancer. The presence of atypical hyperplasia indicates a moderately higher risk (4 to 5 times) above that for apocrine changes, ductal ectasia, or mild epithelial hyperplasia. The presence of atypical *ductal* hyperplasia or *lobular* hyperplasia indicates a higher risk for breast cancer than the base population. These lesions have many features of carcinoma in situ (Fig. 14-19).

Localized Sclerosing Lesions

Sclerosing adenosis mimics invasive carcinoma. It is characterized by lobulocentric changes causing distortion and enlargement of lobular units; increased numbers of acinar structures are accompanied by fibrous changes. The lesions often contain foci of microcalcifications and, when present in an aggregate form, may be detectable mammographically (Table 14-3). Hutter and colleagues consider this lesion to indicate a slightly higher cancer risk (1.5 to 2 times) over that of the base population (see Table 14-2). These lesions carry a risk probability that is similar to that of moderate or florid hyperplastic lesions.

Radial Scar and Complex Sclerosing Lesions

These histopathologic entities are similar to sclerosing adenosis and may mimic carcinoma histologically or clinically. The lesions are not lobulocentric but incorporate various deformed lobular units that possibly originate from the area in which terminal ductules branch from the major duct. The lesions are characterized by a central scar from which elements radiate, with a full array of histologic presentations, including cystic dilatation with units that demonstrate hyperplasia and lobulocentric sclerosis, similar to sclerosing adenosis. The combination of cystic and apocrine changes, as well as hyperplasia, is evident as the lesion matures. Bilaterality and multifocality may be evident. Although radial scars may assume a diverse spectrum of histologic appearances, they are not premalignant.

Ductal Ectasia

This term is reserved for conditions in which the clinical presentation includes palpable lumpiness in the region of the breast beneath the areola. Ducts are involved in a segmental fashion; nipple discharge is a common feature, with periductal scarring and inflammation an attendant finding in later stages of the process. The process is initiated with periductal inflammation and progresses to destruction and dilatation of the ductular system and, eventually, periductal fibrosis and ectasia. These lesions typically occur in perimenopausal or late premenopausal age groups; differentiation from cancer may be difficult. Plaquelike calcifications that occur within the scar wall may be visible mammographically. Localized scarring of the ectatic lesion may cause lumps that are fixed within an inflamed scar of the breast, referred to as “comedo mastitis.” This refers to the grumous, pultaceous material within dilated ducts that may have many of the morphologic features of comedo carcinoma.

Fat Necrosis

Fat necrosis in the breast histologically is no different from its appearance in other organs. Although relatively uncommon, it may clinically and radiographically be confused with scirrhous or even inflammatory carcinoma. Fat necrosis may present after a history of chest wall or breast trauma. The mammographic appearance is quite characteristic and suggests benignity.

Collagenous scarring is the predominant feature in late stages of the disease; granular histiocytes surround “oil cysts” of varying size. These cysts contain free lipid material that results from necrosis of lipocytes. No associated risk for cancer has been established.

Fibroadenoma

This generic term refers to a benign focal tumor that has mixed glandular and mesenchymal elements. Fibrous tissue comprises most of the lesion; the stroma may surround rounded and easily definable ductlike epithelial structures, or epithelium may be skewed into a curvilinear arrangement (Fig. 14-20). The gross appearance is characteristic, with sharp circumscription and smooth boundaries; the cut surface is glistening white. If epithelial elements are excessive, they may appear as light brown areas.

Fibroadenomas typically stop growing when they reach 2 to 3 cm in diameter. Blacks have a greater propensity than whites to develop fibroadenomas and at a younger age. This lesion invariably has a relationship to estrogen sensitivity, and it occurs predominantly in the second and third decades of life. Pain and tenderness may be observed with pregnancy, and an inflammatory response may be accompanied by lymphadenopathy that mimics carcinoma.

Other variants of fibroadenoma are characterized by increased cellularity of the stroma or epithelium. “Adolescent cellular fibroadenoma” typically occurs in adolescence and bears some resemblance to *benign phyllodes tumors*, thus suggesting the term *juvenile adenofibroma*. Five to ten percent of adenofibromas occur around the time of menarche; they frequently have a ductal pattern of epithelial hyperplasia and stromal hypercellularity and are characterized by rapid growth.

Diagnosis of fibroadenomas can be established by fine-needle aspiration (FNA) cytologic techniques. While these tumors may evolve into phyllodes tumors, this is poorly documented. After cytologic documentation by FNA, it is the policy of many clinics simply to observe the characteristic 2- to 3- cm fibroadenoma. Data support this clinical approach for patients younger than 25 years and perhaps as old as 35 years; however, excision is appropriate in older women to exclude carcinoma, which has an increasing incidence above age thirty-five. In most patients, excision is readily accomplished under local anesthesia. Patients with the typical clinical and histologic fibroadenoma are *not* considered at greater risk than the general population for development of subsequent carcinoma. Approximately 100 cases of carcinoma arising in the lesion have been documented. The predominant carcinoma that presents concurrently with the fibroadenoma is *lobular carcinoma in situ*. *Tubular adenoma* is a variant of fibroadenoma that possesses tubular elements arranged in a circumscribed concentric mass with minimal supporting stroma. These lesions have fine nodularity, uniform tubular structures, and the absence of lobular anatomy. The *lactating adenoma* is analogous to tubular adenomas and represents the physiologic response to pregnancy. As a consequence of estriol excess to stimulate growth, these adenomas have a more pronounced anatomic alteration of the lobule than is evident in tubular adenomas.

Phyllodes Tumors

The nomenclature, presentation, and diagnosis of phyllodes tumors historically pose many problems for clinicians. This confusion was predicated on the classic terminology that placed the suffix “sarcoma” on benign and malignant examples of these lesions. Differential diagnostic problems arise with the separation of benign phyllodes tumors from closely related, but distinguishable, fibroadenomas and with recognition of the rare variant that is malignant.

No reliable histopathologic measures exist to differentiate the juvenile fibroadenoma from the benign phyllodes tumor. Typically, the latter harbors a sharp demarcation from the surrounding normal parenchyma, which is considerably compressed and distorted. Connective tissue composes the bulk of the mass, which is firm with mixed gelatinous, edematous, or dense areas. The cystic components owe their origin to sites of infarction, degeneration, and necrosis. These alterations give the breast surface its classic leaflike (phyllodes) appearance. The contour of the breast may assume a “teardrop” configuration with sarcomatous transformation (Fig. 14-21 A).

Histologically, phyllodes tumor may be indistinguishable from the large fibroadenoma. The stroma of the phyllodes tumor has greater cellular activity and cellular content than the fibroadenoma (<3 mitoses/high-power field). The epithelium is deformed into intracanalicular patterns seen in the more common fibroadenoma. Counting mitoses and evaluating margins with attention to identifying infiltrating foci may predict a more aggressive behavior.

Borderline lesions are less likely to assume true malignant potential but have greater potential to recur locally than the usual phyllodes tumor. The overwhelming majority of malignant phyllodes tumors that have metastasized harbor obvious sarcomatous elements (e.g., liposarcoma, rhabdomyosarcoma) rather than fibrosarcoma.

Mammographic foci of calcification and morphologic evidence of necrosis are evident for both malignant and benign phyllodes tumors and are of little value in differentiating the entities. With comprehensive sectioning for study of the lesion to grade the mitotic rate and margins, a more predictable assessment of clinical behavior will be achieved. Local recurrence has been documented in over one-half of the cases. These recurrences are not associated with unrecognized malignant features.

Treatment options are controversial. At the very least, the small phyllodes tumor should be locally excised with an obvious margin (≥1 cm) of normal breast tissue. Often these lesions have been initially enucleated after clinical confusion with a fibroadenoma. When the pathologic diagnosis of a phyllodes tumor with suspicious malignant elements is made (see Fig. 14- 21 B), reexcision of the biopsy scar is indicated to ensure complete local excision. With the larger phyllodes tumors, especially those with malignant elements, total (simple) mastectomy is often required for adequate treatment. Axillary dissection generally is not recommended; rarely, lymph node metastasis occurs.

Papilloma

The solitary papilloma is a mass of variable size that presents in epithelium of the large duct network of the subareolar breast. Peripherally, lesions may be multiple and continuous with hyperplastic alterations that occur within lobular units. These papillary lesions may harbor atypical hyperplasia and ductal carcinoma in situ that is contiguous with the papilloma.

The papilloma presents with a serosanguineous or bloody nipple discharge, typically unilateral. The natural history of solitary lesions suggests a higher propensity for subsequent development of carcinoma. The accompanying epithelial hyperplasia may be responsible for further elevating this risk. Multiple papillomas imply a higher risk for cancer.

The lesions are papillary epithelial outgrowths that occupy dilated ductal spaces. The papilloma may attain several centimeters in size and may appear encysted within the duct from which it arises. Microscopically, there is branching fibrovascular tissue surmounted by epithelium. Dense sclerotic foci may be present. Focal areas of hemorrhage and necrosis are observed. Sites of infarction may cause distortion and compression of epithelium that mimic the appearance of carcinoma on mammogram. When the double cell layer adjacent to the basement membrane is surmounted by excessive cellular columnar cells, the rules used for diagnosis of *atypia* and *in situ carcinoma* apply. If features of ductal carcinoma in situ are evident, the appropriate diagnosis is *noninvasive papillary carcinoma*. Intraductal papilloma is the most common cause of bloody nipple discharge but must be distinguished surgically from adenocarcinoma, which also may present with bloody nipple discharge. The offending duct is identified by radial compression of breast tissue and observation of which lactiferous duct the discharge emerges from. Ductography may identify the involved duct. Under local or general anesthesia, a probe can be placed in the offending duct and the nipple-areola complex partially elevated via a circumareolar incision. The

duct then can be identified by the contained probe and excised for pathologic confirmation of the diagnosis.

Nipple Adenoma

This lesion bears resemblance to the papilloma but presents in the nipple or tissues immediately adjacent to the ductal ampulla of the nipple. Hyperplasia and pseudoinvasion of dense stroma are basic features. *Paget's disease* of the nipple may mimic the lesion. Nipple adenomas have varying patterns of localized hyperplasia with supporting fibrous and cystic changes. Clinically the lesions are recognized when they enlarge to approximately 1 cm. They may demonstrate epithelial hyperplastic components and fibrosis, and confusion with a malignancy is possible.

Ductal Adenoma (Nodular Adenosis)

This lesion may clinically resemble the presentation of a papilloma; differentiation from carcinoma is essential. Histopathologically, adenomas are related to epithelial lesions that have unusual patterns of sclerosis and adenosis. Dense fibrous tissue within epithelial cells appears to be pseudoinvasive and may cause confusion with carcinoma.

CARCINOMA OF THE BREAST

Incidence

Carcinoma of the breast is the most common site-specific cancer in women and is the leading cause of death from cancer for females 40 to 44 years of age. Breast cancer accounts for 32 percent of all female cancers and is responsible for 19 percent of the cancer-related deaths in women. Approximately 180,200 invasive breast cancers were expected to be diagnosed in the United States in 1997; approximately 43,900 women will die because of the tumor. Cancer registries in Connecticut and upper New York State note that the age-adjusted incidence of new cases has been steadily increasing since the mid-1940s. In the 1970s the probability of a woman in the United States developing breast cancer was estimated at 1 in 13; in 1980 it was 1 in 11 and in 1996 the frequency was 1 in 8.

Until the past decade, breast cancer was the leading cause of cancer- related mortality in women. In 1985 the lung surpassed it as the leading site of cancer-related mortality in women (Fig. 14-22). Despite the steady increase in incidence, the overall breast cancer mortality has remained static. This relative decrease in mortality rate reflects the detection of an increasing percentage of early disease. From 1960 to 1963, 5-year survival rates were 63 percent and 46 percent for white and black women, respectively; the figures for the 1981 to 1987 interval were 78 percent and 63 percent, respectively.

Worldwide, breast carcinoma is an epidemiologic problem. England and Wales have the highest national age-adjusted mortality for breast cancer (27.7 per 100,000 population). The United States ranks thirteenth with 22.0 cases per 100,000. For 1986 to 1988, women of South Korea ranked lowest among all nations, with an incidence of 2.6 cancers per 100,000 population. Mormon, Seventh-Day Adventist, Alaskan, American Indian and Eskimo, Mexican-American, and Japanese and Filipino women living in Hawaii have a lower per capita incidence of breast cancer than other Americans; nuns and Jewish women have a higher than average incidence. There is at least a fivefold variation in the incidence of the disease reported among different countries, although this difference appears to be diminishing. Women living in less-industrialized nations tend to have lower rates of breast cancer than those living in industrialized countries, but Japan appears to be an exception.

Etiology

Genetic Factors

Henderson and associates and Lynch and associates (1991) have documented the importance of heredity and the genetic predisposition for breast cancer. Definitions suggested by Lynch and colleagues are as follows:

Sporadic Breast Cancer (SBC): A breast cancer patient with no family history of cancer of the breast through two generations involving siblings, offspring, parents, aunts and uncles, and both sets of grandparents.

Familial Breast Cancer (FBC): A breast cancer patient with a family history including one or more first- or second-degree relatives with breast cancer that does not fit the hereditary breast cancer definition.

Hereditary Breast Cancer (HBC): A breast cancer patient with a positive family history of breast cancer and, sometimes, related cancers (e.g., ovarian, colonic) and with high incidence and a distribution in the pedigree that is consistent with an autosomal dominant, highly penetrant, cancer susceptibility factor. Other factors supporting the HBC classification include frequent early age at premenopausal onset and an increased incidence of bilateral breast cancer and other multiple primary cancers.

Lynch and associates documented the frequency of *sporadic*, *familial*, and *hereditary breast cancer* variants. In an original cohort of 225 consecutive breast cancer patients, updated with 103 new patients, 68 percent of the 328 probands studied were sporadic, 23 percent were familial, and 9 percent were hereditary. The authors state that this may represent a considerable underestimation of the frequency of hereditary breast cancer. With documentation of pedigree, familial breast cancer may constitute as great as one-third of the total incidence of breast cancer cases. Approximately one-fourth of these cancers fall into the special subset of hereditary breast cancer and represent an autosomal dominant susceptibility pattern with early age of onset, with excess bilateral disease, and with other multiple primary cancers.

The risk for developing hereditary breast cancer is determined by pedigree, appears to be independent of age at first pregnancy, and is higher when a biopsy confirms atypical hyperplasia. Lynch considers HBC a site-specific heterogeneous entity in which other presentations may be evident. These variants include *SBLA syndrome* (sarcoma; *breast cancer and brain tumors*; *lung and laryngeal carcinoma and leukemia*; and *adrenal cortical carcinoma*) and *Cowden 's disease* (cancer-associated genodermatosis with multiple trichilemmomas, including cutaneous involvement of the face and multiple areas of the hands, feet, and forearms).

Currently biomarkers do not have the sensitivity or specificity to identify HBC individuals before cancer is expressed. BRCA1, a major breast cancer susceptibility gene, was mapped to a region of the chromosome 17q and identified. Linkage to BRCA1 is present in most breast-ovarian cancer families with hereditary site-specific breast cancer. Another major breast cancer susceptibility gene, BRCA2, was mapped to the q12-13 region of chromosome 13 and isolated. BRCA2 is linked to most families with male hereditary breast cancer. These two breast cancer susceptibility genes together account for most of the inherited breast cancer. However, it is likely that other undiscovered susceptibility genes may also play a role in the development of sporadic breast cancers.

Dietary Influences

The committee on Diet, Nutrition, and Cancer of the National Academy of Sciences concluded that a causal relationship exists between dietary mammalian fat and the incidence of breast cancer. Fried, high-fat foods can increase the risk of developing breast cancer approximately twofold. A study of five ethnic groups in Hawaii demonstrated a strong relationship between breast cancer incidence and consumption of total fat, saturated fat, animal fat, and unsaturated fat. A study conducted by the National Cancer Institute noted that dietary fat intake in breast cancer patients was contributory. Compared to controls, the highest quartile for beef or pork consumption was associated with a relative risk for cancer 2.7 times higher than that of the lowest quartile.

Both the quality and the quantity of dietary fat intake may influence the incidence of this disease. Epidemiologic studies also suggest that Japanese and Greenland Eskimo women have a low incidence of breast cancer despite consumption of large quantities of fat. Their diets include a large volume of omega-3 fatty acids, which are unique to the marine-derived lipids. The number of annual breast cancer deaths among Japanese women doubled over the 20-year interval from 1955 to 1975. During that period the Japanese diet became increasingly similar to that of Western societies. For second- and third-generation American offspring of Japanese immigrants, the incidence approaches that of Caucasian women born in the United States.

Hormone Usage

Kalache and associates demonstrated that combined oral contraceptives have no effect on breast cancer risk when used by women in the middle of their reproductive lives (ages 25 to 39 years). These data pertained even if oral contraceptives were taken for many years. By contrast, Lipnick and colleagues noted an adverse effect of these combined hormones on breast cancer risk when taken for a prolonged period at a very early age or when taken before the first full-term pregnancy. The World Health Organization study suggests that there is neither an increase nor a decrease in the risk of breast cancer with the use of the injectable contraceptive depot- medroxyprogesterone acetate (DMPA).

Vessey reviewed the epidemiologic literature and concluded that estrogen use by perimenopausal and postmenopausal women for hormonal replacement may slightly increase the risk of breast cancer. The risk is said to be accentuated in women with preexisting benign disease of the breast. The possibility exists that some of this excess risk is attributable to more thorough evaluation and diagnosis of breast cancer in women placed on estrogen supplementation therapy. This point of view is supported by a favorable stage distribution in estrogen users. Very little data have been generated on the effects for cancer risk with hormonal replacement therapy using combined estrogens and progestational agents.

Obesity

The majority of data suggest that breast cancer risk is directly correlated with relative weight; the risk for obese women is 1.5 to 2 times higher than for nonobese women. This relative risk is restricted to postmenopausal individuals. The increasing incidence of breast cancer mortality in Japanese women has been paralleled by a proportionate increase in both height and weight in this population.

Breast-Feeding and Menopause

Previously, breast-feeding of long duration (>36 months in a lifetime) was thought to reduce the risk of breast cancer. This observation is no longer considered valid. For women in whom menopause occurs after the age of fifty-five the risk of developing the disease is twice as high as for those whose menopause started before age forty-five. Artificially induced surgical menopause appears to be protective for breast cancer; protection is lifelong, and removal of endogenous estrogen dramatically reduces breast cancer risk. The earlier the surgical menopause, the lower the risk. The risk for breast cancer is one-third as high in women having oophorectomy at age thirty-five or younger as in women whose natural menopause was age fifty or later.

Child Bearing and Fertility

Infertility and nulliparity are associated with a higher probability (30 to 70 percent) for developing breast cancer in comparison with the probability for parous women. With decreasing age at the time of first pregnancy, the risk decreases proportionately. Women impregnated before 18 years of age who have a full-term pregnancy have a breast cancer risk approximately one- third that for women who become pregnant for the first time after 35 years of age. This increase in the relative risk in the latter group is related to persistent exposure to endogenous estrogens in the absence of appropriate concentrations of progesterone. Women who have their first full-term pregnancy after age thirty have an even greater risk for breast cancer than do nulliparas.

Multiple Primary Neoplasms

Harvey and Brinton concluded that women with a history of primary breast cancer have a risk three to four times higher for primary cancer in the contralateral breast. This risk for a second primary cancer in the breast is higher in women with a positive family history of the disease. Other factors that potentially affect the risk of a second primary cancer, including reproductive factors, body build, and prior radiation treatment for cancer, remain undetermined. Women with a history of previous ovarian or endometrial carcinoma have a relative risk of about 1.3 to 1.4 for development of a primary cancer of the breast.

Irradiation

Atomic bomb survivors from Nagasaki and Hiroshima, women treated with high-dose radiation for acute postpartum mastitis, and women who have received multiple chest fluoroscopic examinations for treatment of pulmonary tuberculosis have a higher incidence of breast cancer. Risk from multiple exposures to relatively low doses is similar to the risk of one large dose of similar radiation yield.

It was previously suggested that susceptibility to the carcinogenic potential of irradiation had its greatest magnitude between the ages of ten and twenty, with relative protection following exposure before age ten and after age forty. Data suggest that women exposed to ionizing radiation from infancy to age ten have an increased risk but that this risk is within the expectant ranges for development of breast cancer. Less than 1 percent of breast cancer cases result from diagnostic radiologic procedures. Radiotherapy for breast cancer may increase the risk for cancer of the contralateral breast. Risk of breast cancer is reduced after radiation treatment for cancer of the cervix as a result of reduction of estrogens.

Conclusions

Table 14-4 lists the established risk factors for cancer of the breast in women and the magnitude of these factors. With the exception of age, country of birth, and history of breast cancer in both mother and sister, all of the relative risks reported to date are of modest magnitude. Inconsistent data suggest protective effects from parity and lactation in various age groups and a higher risk associated with alcohol consumption and diethylstilbestrol (DES) exposure during pregnancy. Physical activity has emerged as a factor worthy of study.

Natural History

The natural history of breast cancer has been reported by Bloom and associates on the basis of the records of 250 patients with *untreated* lesions cared for on cancer charity wards in Middlesex Hospital, London, between 1805 and 1933. The median survival of this population was 2.7 years after initial diagnosis (Fig. 14-23). The 5-year and 10-year survival rates for these untreated patients were 18 and 3.6 percent, respectively; only 0.8 percent survived for 15 years or longer. Autopsy data confirmed that 95 percent of these patients died of breast cancer, and the remaining 5 percent died of intercurrent disease. Almost three-quarters of the patients had ulceration of the breast during the course of the disease.

The mean survival in the Middlesex series and for over 1,000 untreated cases obtained from the world literature was 38.7 months (range 30.2 to 39.8 months). It should be noted, however, that for all reports of untreated patients, survival is calculated from onset of first symptom. The longest survivor died in the nineteenth year after the onset of symptoms. Although this suggests that the disease may not always be a rapidly lethal one, 60 percent of patients who develop metastases do so within the first 24 months after mastectomy.

Evidence from the Middlesex data and from thousands of other reports suggests that 5-year survival does not always equate with cure. Metastatic foci may become evident 20 or 30 years after treatment of the index lesion; conclusive results cannot be derived from breast cancer data until at least 5 years have elapsed after institution of a therapeutic regimen. For the breast cancer patient, metastatic cancer is the most common cause of death for years 5 through 10 following mastectomy.

The typical carcinoma of the breast (80 to 85 percent) is a scirrhous adenocarcinoma with productive fibrosis that originates in the ductules and invades the parenchyma. For the majority of lesions, there is a long preclinical (occult) period when the tumor and/or host factors can modulate metastasis. If a tumor doubled in size every 100 days, it would take more than 8 years for a solitary neoplastic cell to grow to a 1-cm detectable clinical mass (10⁹) cells. Laboratory and clinical evidence suggests that growth rates are not consistent, especially within the first 30 doublings.

Metastases presumably may occur within any period of neoplastic growth after the first few doublings. Growth rates of tumors at distant sites have a wide range; this accounts for the observation that primary lesions may be diagnosed many years before the detection of metastases. Increasingly important roles of cytokines and growth factors on metastatic growth have been uncovered.

As the size of the small mass of breast cancer cells increases, numbers of these cells may be shed into cellular spaces and may be transported into the rich lymphatic network of the breast or into venous spaces. At approximately the twentieth doubling, these tiny tumor masses acquire their own blood supply as a network of neovascularization. Thereafter, these cells may be shed directly into the systemic venous blood. Fisher and colleagues suggest that these cells may enter the lymphatics with early crossover into the venous blood via lymphaticovenous communications. In systemic blood, tumor cells are rampantly scavenged by natural killer (NK) lymphocytes and macrophages. Successful implantation of metastatic foci from breast cancer rarely occurs until the index lesion exceeds 0.5 cm in transverse diameter; this size corresponds to the twenty-seventh doubling of the tumor mass.

In Table 14-5 observed survival rates for patients with breast cancer are related to the clinical and histologic stage. The number of lymph nodes that are involved with metastatic disease is inversely proportional to patient survival time. Tumor size is another important prognostic indicator and is directly correlated with the presence of nodal metastases. Occult (micrometastatic) axillary lymph node metastases of 5-year survivors closely parallel those for patients whose nodes are free of disease.

With tumor enlargement and invasion of the surrounding breast parenchyma, the accompanying fibrosis and desmoplastic response entrap and shorten the suspensory ligaments of Cooper to produce characteristic peau d'orange or retraction of the skin. Subdermal emboli of neoplastic cells fill the endolymphatic spaces and ultimately invade the corium. Skin invasion is preceded by localized edema; effective drainage of lymphatic fluid from the skin is disrupted. Should tumor cells in the corium continue to grow, ulceration of epithelium will occur. As new areas of skin are invaded, small satellite nodules are evident near the ulcer crater. Venous capillaries are invaded, and tumor cells seed the circulation passing via lateral, axillary, or medial central intercostal veins to enter the pulmonary circulation or via vertebral veins that course up and down the spinal column (Batson's plexus).

With expansion of the tumor mass after cellular doubling, tumor cells exfoliate and transgress along lymphatic spaces toward the upper-outer quadrants or enter the rich capillary plexus to communicate with parasternal nodes of the systemic circulation. Local, regional, or distant lymphatic implantation and growth are possible. With involvement of any regional lymphatic area, the nodes are at first ill-defined and soft, and then assume a firm, hard, or fixed configuration with increasing expansion of tumor growth. Eventually nodes adhere to each other and form a large conglomerate (fixed) mass. Tumor cells may break through the capsule with fixation of lymphatics to soft tissues or contiguous structures of the axilla or chest wall. Typically, axillary nodes are involved progressively from low (Level I) to central (Level II) to apical (Level III) regions.

Systemic dissemination is critical, because more than 95 percent of patients who die of uncontrolled breast cancer have distant metastases. The most important prognostic correlate for recurrent disease and survival is the nodal status (Fig. 14-24 A). Node-negative patients have a 20 to 25 percent incidence of relapse, compared to 50 to 75 percent for node- positive patients. Node-positive patients have recurrence preferentially in distant organs and tissues. The more common sites of disseminated disease are bone (49 to 60 percent), lung (15 to 20 percent), pleura (10 to 15 percent), soft tissues (7 to 15 percent), and liver (5 to 15 percent). In general, 10 to 30 percent of recurrences are local, 60 to 70 percent are distant, and 10 to 30 percent are both local and distant.

Staging of Breast Cancer

The staging of breast cancer is an attempt to predict potential survival rates from objective data.

Tumor Characteristics and Lymph Node Metastases

Koscielny and associates demonstrated that metastases are positively correlated with tumor size; this correlation does not occur in one-half of the cases until the primary tumor attains a size of 3.6 cm in diameter (Fig. 14-24 B). Nemoto and colleagues and Fisher and colleagues have shown a distinct relationship between the increase in tumor size, the probability of axillary nodal metastasis, and disease-free survival.

The single most significant predictive factor of 10- and 20-year survival is the absolute *number* of lymph nodes involved with metastatic neoplasm. Physical examination is notoriously inaccurate in determining the presence of lymphatic involvement and may have false-positive rates and false- negative rates for detection of axillary metastasis that range from 25 to 31 percent and from 27 to 33 percent, respectively. Henderson and Canellos report that patients with negative axillary lymphatics have 5-year and 10- year survival rates of 78 and 65 percent, respectively; for patients with four or more positive lymphatics, survival rates are 32 and 13 percent, respectively. Fisher and associates observed that the number of positive nodes is correlated with the percentage of 5-year and 10-year treatment failures. The absence of positive nodes was associated with a 20 percent failure rate at 10 years; the presence of more than four positive nodes was associated with a 71 percent treatment failure rate; the presence of more than 13 positive nodes increased the failure rate to 87 percent. Patients with occult micrometastases in lymph nodes initially reported as histologically negative may have survival rates that are not significantly different from those of patients with negative nodes.

The location of the nodes is important; apical axillary (Level III) node metastases carry an ominous prognosis that is distinctly worse than that of Level I involvement. Level I dissection can be predictive of axillary nodal involvement of the residual contents. Fisher and colleagues have determined that dissection of Levels I and II is more than adequate in most cases to predict systemic spread of disease. This report suggests that in order to achieve definition of axillary nodal involvement, sampling of more than ten nodes is necessary.

Clinical trials are investigating the issue of "sentinel node" biopsy as an alternative to formal axillary dissection for staging. A radioisotope or a dye is injected into the region of the tumor, and radioactivity or presence of dye is assessed in the axilla. When a node is identified by this technique, it is removed for biopsy. If this so-called sentinel node shows no tumor, this is regarded as equivalent to a negative axilla.

We do not recommend the sampling of internal mammary nodes in routine dissections. Positive internal mammary nodal metastasis may be expected in central and medial quadrant primary tumors; this frequency increases proportionately with size of the index tumor. Clinical or pathologic evidence of lymph node extension to supraclavicular sites is indicative of advanced (stage IV) disease, considered systemic. Routine scalene or supraclavicular nodal biopsies generally are not indicated.

Evolution of Staging

Three commonly used staging systems have evolved: the Manchester, the Columbia Clinical Classification, and the TNM (tumor, nodes, metastasis) systems. The American Joint Committee on Cancer (AJCC) modified the TNM system for breast cancer (Table 14-6).

A number of physical and radiologic parameters must be reevaluated: (1) comprehensive history and physical examination; (2) bilateral breast imaging (e.g., film screen mammography or xeromammography); (3) clinical pathology laboratory evaluations, including hemogram and hepatic function; (4) chest x-ray (posteroanterior and lateral); and (5) skeletal roentgenologic survey (indicated if symptomatic).

Select examinations include the following:

(1) Abdominal computed tomography (CT) when the following are evident:

(a) Abnormal liver function

(b) Hepatosplenomegaly

(2) Radionuclide bone scans for any of the following lesions:

(a) Advanced local disease (T3, T4)

(b) Lymph node metastasis (N1, N2, N3)

(c) Distant metastases (M1)

(d) Osseous symptoms in the absence of a, b, or c

Bone Scans

Bone scanning remains controversial. These radionuclide tests do not precisely identify metastatic disease. Inflammation associated with degenerative joint disease, osteoarthritis, or overlying soft tissues may provide false-positive results. The presence of a positive scan is indicative of advanced stages. Application of the technique should be applied in a cost-effective manner only for patients with T1, T2, or T1,N1 lesions. Scans are indicated in the presence of positive skeletal roentgenograms, bone pain, or palpable regional or metastatic disease.

CT/MRI

CT and MRI are equivalent methods for diagnosis of visceral metastases. Extracavitary ultrasonography and radionuclide scans for detection of hepatic or pulmonary metastases may complement CT or MRI. CT or MRI scanning is indicated for patients who have suspected distant metastatic disease as evidenced by symptoms, abnormal roentgenograms, abnormal liver function, bone metastases, or supraclavicular adenopathy. These modalities are state-of-the-art techniques for the detection of brain, chest, liver, abdominal, and pelvic metastases.

Future Staging Trends

New technology data allow the detection of hormonal, cytosol protein receptor, and functional characteristics in the cytoplasm of breast tumor cells. Flow cytometry, which evaluates cell surface and nuclear characteristics, has prognostic implications. Cytokine and growth factor analyses and the use of genetic information (e.g., on proto-oncogenes) will have increasing application in the staging process. Physical and cellular measurements of the primary tumor growth rate and percentage of malignant involvement of the breast have been developed. These refinements will potentially be included in future staging systems.

Histopathology

Noninfiltrating (In Situ) Carcinoma of Ductal and Lobular Origin General Considerations

The literature suggests that all cases of invasive breast cancer go through a period in which normal epithelial cells undergo malignant transformation but do not “invade” beyond the investing basement cell membrane. There is reason to question whether in situ carcinoma is truly a malignancy, whether it merits substantial efforts at detection, whether treatment affects the subsequent development of invasive cancer, and what form of therapy, if any, should be applied.

Foote and Stewart published the landmark description of *lobular carcinoma in situ (LCIS)* in 1941, and distinguished this pathologic entity, with its unique biologic behavior, from *ductal carcinoma in situ (DCIS)*. DCIS is the most common histologic variant of the noninvasive stage of carcinoma that originates in the major lactiferous ducts. In the late 1960s Gallagher and Martin published the results of their whole organ section studies and affirmed the transition that established a stepwise evolution of invasive breast cancer from benign epithelium through the in situ and subsequent invasive stages. This recognition allowed them to coin the term *minimal breast cancer* and to stress the importance of early detection of malignancy at a stage when proper therapy would translate into a 10-year cure probability of 90 percent or more. These authors further acknowledged that minimal breast carcinoma included LCIS, DCIS, and minimally invasive cancers smaller than 0.5 cm.

We now recognize that all of these entities have distinct clinical and biologic implications and that each entity deserves distinct therapeutic considerations. Figure 14-25 shows the histopathologic phases of proliferative changes that transform to atypical lobular or ductal hyperplasia and the potential pathway to initiate LCIS or DCIS. Although probably irreversible to normal epithelium, in situ carcinoma stages do not all inevitably transform to invasive disease.

Epidemiology

The increasing frequency of diagnosis of LCIS and DCIS is attributable primarily to mammography. When physical examination was the most common initiator for breast biopsy, in situ lesions constituted only 1.4 percent of all biopsies and only 3 to 6 percent of all breast malignancies. LCIS was more commonly diagnosed than DCIS by ratios of 2:1 or 3:1 in these series. Of 21 clinical series reported by Frykberg and Bland, a total of 9,472 mammographically detected nonpalpable breast lesions demonstrated a sevenfold increase in the incidence of in situ disease among all biopsies (9.6 percent) and a fourteenfold increase in incidence of in situ disease among all breast cancers (45 percent). Mammography has detected a predominance of DCIS over LCIS, averaging a 3:1 ratio in many series.

Lobular Carcinoma In Situ

Lobular elements of the breast from which LCIS originates are not noted in males; this form of noninvasive cancer is observed only in females. The average age at diagnosis is forty-four to forty-seven years, which is 15 years younger than the age at which invasive breast cancer is diagnosed. Over 90 percent of women with LCIS are premenopausal; this is distinctively different from the incidence (30 percent) for women with invasive cancer. This epidemiologic observation emphasizes the importance of estrogen influence on the biologic behavior of LCIS. Ninety percent of the invasive lobular cancers have estrogen receptor activity, compared to only 55 percent in duct carcinoma.

The frequency of LCIS in the base population cannot be reliably determined because it usually presents as an incidental finding. Published series suggest a wide frequency, ranging from 0.8 to 8 percent of all breast biopsies. LCIS has a distinct racial predilection, occurring twelve times more frequently in white women than in black women; black women, however, have a tenfold increase in the recurrence rate after therapy. A review of 1,455 nonpalpable breast malignancies in 18 recent series indicates that LCIS constitutes 2.3 percent of 6,287 biopsies and 9.8 percent of all malignancies.

Ductal Carcinoma In Situ

This histologic variant of in situ carcinoma of the breast is observed predominantly in the female but constitutes approximately 5 percent of all male breast cancers. Most patients present with DCIS in early menopausal years. A DCIS was found in nearly 7 percent of 6,287 breast biopsies and in nearly one-third of 1,455 nonpalpable breast malignancies. The predominance of this lesion among all in situ breast cancers is related to its mammographic definition and its presentation as a clinically palpable mass in over one-half of all cases.

Pathology of In Situ Disease

The original description of in situ carcinoma by Broders stressed the absence of invasion of cells into surrounding stroma and their confinement within the natural ductal or lobular boundaries of the cell membrane. The basement membrane is the crucial anatomic structure that defines the presence or absence of invasion. Diagnosis of an in situ lesion necessitates multiple sections to exclude invasion; frozen section is rarely relied on.

Evidence suggests that LCIS originates from the terminal duct–lobular apparatus. This explains its tendency to present as a nonpalpable mass and its diffuse distribution throughout the breast. The normal lobular anatomy undergoes a disorderly proliferation of epithelial cells to the point of filling and distending the terminal lobular lumina while the overall lobular architecture is maintained. Cells remain uniformly homogeneous, with a normal ratio of nuclear to cytoplasmic area and the absence of necrosis and mitoses. The cells are enlarged without loss of cohesion. Cytoplasmic mucoid globules are a distinctive cytologic feature of LCIS that distinguishes it from DCIS. The disease process frequently is observed in breast biopsy specimens that harbor microcalcifications; however, the process is not generally associated with the calcific sites. The process typically occurs in surrounding tissues that are clinically and radiologically normal. The “neighborhood calcification” is a feature unique to LCIS and contributes to its diagnosis.

The earliest phases of DCIS are characterized by proliferation of the inner cuboidal layer of the epithelial cells in major lactiferous ducts to form papillary ingrowths within the lumen. Cells in DCIS are well differentiated without evidence of significant pleomorphism, mitoses, or atypia. This may lead to difficulty in differentiating DCIS from benign hyperplasia (Fig. 14-26). With growth of the “papillary pattern” of DCIS, ingrowths coalesce to fill the ductal lumen until scattered rounded spaces remain interspersed among solid clumps of cells, which themselves tend to show atypia, hyperchromasia, and loss of polarity. This event was termed by Schultz- Brauns the “cribriform” growth pattern of DCIS. By contrast, a “solid” histologic pattern of DCIS is recognized when cellular growth obliterates these spaces and the ducts become distended with more anaplastic cells and mitotic figures.

With continued growth, these cells outstrip their blood supply, become necrotic, and lead to the classic “comedo” pattern that has been confused with benign inflammatory diseases. In the comedo variant, calcium deposition generally occurs in areas of necrosis, leading to typical DCIS radiographic manifestations. This latter pattern of DCIS has a significantly higher degree of nuclear grade, multicentricity, and microinvasion, suggesting that its biologic behavior is more aggressive than the papillary or solid types. Other indicators of aggressive biologic behavior and poor prognosis associated with comedo DCIS are high DNA proliferative activity, C-erbB-2 oncogene amplification, and overexpression of

HER-2-neu oncogene activity. Aneuploid DNA patterns and chromosome abnormalities that predominate in comedo DCIS also are correlated with a strong invasive potential and poor prognosis. Estrogen receptor activity is found as frequently in this variant as in invasive breast carcinoma. Consideration of these factors assists in determining management of the disease.

LCIS and DCIS may coexist, and their cytologic similarities may lead to diagnostic and therapeutic confusion. Histochemical and ultrastructural studies as well as monoclonal antibody technology have been used to distinguish the two pathologic entities. The term *intraductal carcinoma* is commonly applied to DCIS and denotes the basic pathologic features of ductal elements and their containment within the basement membrane.

Natural History of In Situ Carcinoma

Between 10 and 37 percent of women with LCIS will develop a breast malignancy, which represents a risk ratio six to twelve times higher than that of the base population. The majority of cancers that develop subsequent to LCIS occur more than 15 years after diagnosis, and over one-third occur more than 20 years later. Malignancies that develop later are recognized in either or both breasts, regardless of which breast harbors the focus of LCIS. Invasive cancers are discovered synchronously with LCIS in approximately 5 percent of cases.

A confounding variable is the evidence that 50 to 65 percent of future *invasive* malignancies are not lobular carcinoma but of ductal origin. This raises speculation about the validity of the transition theory. Invasive lobular carcinoma also occurs in this setting, however, at eighteen times the expected rate, and it is presumed that this entity does develop from LCIS. The distinctive pathologic and behavioral characteristics of LCIS are regarded as a “marker” of increased risk rather than an anatomic precursor that develops into an invasive lesion.

Since DCIS has been recognized as a distinct entity for a shorter period than LCIS, less evidence is available for the natural history of DCIS. The risk for invasive cancer from DCIS is considered to be in the range of 30 to 50 percent over 10 years. The risk for a subsequent malignancy is increased elevenfold after the diagnosis of DCIS. The future cancers are observed in the ipsilateral breast, usually in the same quadrant as the original biopsy, suggesting that DCIS is a true precursor of its invasive counterpart. Table 14-7 lists the salient clinical and pathologic characteristics of DCIS and LCIS. The report of the American College of Surgeons indicates a lower survival rate for DCIS than for LCIS, suggesting the need for distinct management of the two entities.

Studies of mastectomy specimens from patients with a diagnosis of DCIS confirm that residual disease in the biopsy site exists in as many as three- quarters of patients. These data suggest that a more complete excision of the primary lesion may reduce the risk of recurrent disease. By contrast, resection of smaller, nonpalpable specimens of DCIS for diagnosis has been observed to eliminate the risk completely.

Data suggest the need to distinguish grossly palpable and microscopically nonpalpable variants of DCIS to plan appropriate therapy more rationally. In comparison to nonpalpable or microscopic presentation, large palpable forms of DCIS have been noted to have occult invasive rates as high as 46 percent, higher rates of local recurrence, multicentricity, axillary nodal metastases, and evidence of poor survival rates. A more cautiously optimistic view of DCIS suggests that an occult microscopic presentation is important, but if it is left unresected, invasive carcinoma may develop.

Bilaterality and Multicentricity of In Situ Cancer

The frequency of synchronous bilaterality depends on the extent to which it is sought, suggesting that the published rates probably underestimate the true incidence. LCIS has a known statistically significant rate of bilaterality that has been reported to be as great as 90 percent. In contrast, DCIS is associated with only a 10 to 15 percent incidence of bilaterality, with an occasional series reporting an incidence as high as 30 percent.

Cancer of the breast originates and develops from multiple foci diffusely scattered throughout the breast. This explains both the heterogeneity of histologic forms found within individual tumors and the presence of other foci of malignancy in breasts that harbor noninvasive cancer. This frequency has been known to range from 10 to 90 percent. These foci of cancer consist largely of the in situ disease identified within the index lesion, and the frequency is not dependent on histology.

Multicentricity refers to occult malignancies found outside the quadrant of the primary (index) tumor, whereas *multifocality* and *residual disease* are appropriate terms for sites within the same quadrant as the index lesion. Studies that adhere to the true definition of multicentricity report that this phenomenon occurs in approximately one-third of patients with DCIS; some series note that LCIS has a much higher rate of true multicentricity that may approach 100 percent. A lower incidence of multicentricity is associated with invasive cancer as compared to in situ disease. This supports the theory that breast malignancies develop from a coalescence of multiple sites of origin. The small number of valid long-term studies of treatment of DCIS and the greater frequency of multicentricity for the lesion suggest caution in recommending conservation therapy that has been designed and implemented for invasive disease.

Infiltrating Malignancies General Considerations

Cancers of the mammary ducts may be classified according to histogenesis (duct, lobule, acini), histologic characteristics (adenocarcinoma, epidermoid carcinoma, sarcoma, etc.), gross characteristics (scirrhous, colloid, medullary), and invasive criteria (infiltrating, in situ). Approximately three- quarters of infiltrating carcinomas of the breast have been included in the imprecise characterization of “infiltrating ductal” or “adenocarcinoma, not otherwise specified” (NOS) category. The term “adenocarcinoma NOS” is preferred to “ductal carcinoma” because the site of origin of most breast adenocarcinomas remains indeterminate. The terminal ductal lobular unit is the most probable site for origin of the majority of breast adenocarcinomas. On the basis solely of morphologic features, no clinically significant differences have been observed between the less differentiated “unspecified” breast carcinomas and mixed variants, unless qualified by the objective assessment of nuclear and architectural degrees of differentiation, i.e., grade.

Current terminology describes histology that is based on the dominant architecture of the lesion, but many patterns may be observed in any one breast cancer. The following classification was originally proposed by Foote and Stewart:

Paget's disease of the nipple	
Carcinoma of duct origin	
Noninfiltrating (in situ, intraductal)	Infiltrating
1	Adenocarcinoma with productive fibrosis (scirrhous, simplex)
1	Medullary
1	Comedo
1	Colloid
1	Papillary
1	Tubular

Carcinoma of mammary lobulesNoninfiltrating (in situ) Infiltrating Relatively rare carcinomas Sarcoma of the breast The rarely encountered histologic patterns include melanoma, adenoid cystic carcinoma, squamous cell carcinoma, sweat gland carcinoma, and carcinoma with mesenchymal metaplasia of chondromatous or osseous types.

Approximately 40 to 50 percent of breast carcinomas are located in the upper outer quadrant, owing to the relatively larger volume of breast tissue in this sector. Almost one-quarter occur in the juxtaareolar area; the remainder are randomly distributed throughout medial and lower outer quadrants of the breast.

Paget's Disease of the Nipple

Described by Sir James Paget in 1874, this lesion presents as a chronic, eczematoid eruption of the nipple. Paget's disease constitutes approximately 2 percent of histologic types, and it is almost always associated with an underlying intraductal or invasive carcinoma. It presents as an encrusted, scaly, hyperemic, and enlarged tumor that occupies the surface of the nipple-areola complex.

Symptoms include tenderness, itching, burning, and intermittent hemorrhage. Intraductal adenocarcinomas often involve the epidermis of the nipple and areola by intraepithelial dissemination. Physical findings in the nipple-areola complex precede the identification of a palpable mass in the subareolar area. One-quarter to one-third of patients have axillary node metastasis at diagnosis. In general, this breast cancer has a better prognosis than the majority of lesions, because the nipple-areola changes promote early consultation, biopsy, and diagnosis.

Microscopically, Paget's disease presents as an intraepithelial tumor composed of single or small groups of clear cells with large vesicular and prominent nuclei. The intraductal lesion often is multifocal; ducts throughout the entire breast may be dilated as a result of obstruction of central collecting ducts at the ampulla of the nipple. Pathognomonic of this disease is the presence of very large, pale, vacuolated cells (Paget's cells) in the rete pegs of the epithelium. The lesion may be confused with superficial melanoma; differentiation between pagetoid intraepithelial malignant melanoma and Paget's disease of the nipple is difficult. The diagnosis is differentiated by demonstration of S-100 protein or melanoma-specific antigen immunoreactivity in malignant melanoma. The application of immunohistochemistry, specifically demonstrating carcinoembryonic antigen (CEA) within the Paget cells, has greatly facilitated diagnosis of the lesion. Melanoma does not contain CEA.

The origin of the Paget cell remains controversial, and two hypotheses are considered: (1) *epidermotrophism* of underlying tumor cells, and (2) *intraepithelial carcinomatous metaplasia*. The presence of typical Paget cells and the associated findings are diagnostic of Paget's disease of the nipple even in the absence of a subareolar mass. The ductal malignancy is most commonly invasive; but Paget's disease may be associated with carcinoma in situ of ductal origin.

Infiltrating Ductal Carcinoma with Productive Fibrosis

The 78 percent frequency of adenocarcinoma of the breast (ductal carcinoma) with productive fibrosis (*scirrhous, simplex, NOS*) is shown in Table 14-8. One-third of these tumors have recognizable elements of a specific histologic type, but the presence of specific tumor types in small volumes does not appear to affect prognosis. The prototypical common adenocarcinoma of the breast presents in a perimenopausal or postmenopausal woman in the sixth decade as a solitary, nontender, firm, ill-defined mass.

The tumor characteristically possesses a poorly defined border that is typically better defined by palpation than inspection. Cut surfaces suggest a central radiating stellate tumor

with a chalky-white or yellow streak extending into surrounding parenchyma. The histologic picture may reveal variable cellular and nuclear grade. A broad spectrum of variants is observed, from in situ to highly anaplastic, suggesting significant heterogeneity (see Fig. 14-26). Other lesions can possess bland homogeneity of cellular differentiation throughout the specimen. Neoplastic cells are arranged in small clusters or stacked in single rows (to produce “Indian filing”) that occupy irregular cleft spaces between collagen bundles (Fig. 14-27).

With profound desmoplastic response of tumor growth, the resultant fibrosis and tumor infiltration can shorten Cooper's ligaments as they course from the deep layer of clavicopectoral fascia to the superficial fascia of the corium. With hyalinization, these ligaments become entrapped within the expanding desmoplastic border of the tumor. With progressive growth, Cooper's ligaments are further shortened to initiate the classic physical finding of skin dimpling directly over the tumor and to initiate advanced local and regional presentations. This physical characteristic is exaggerated when the patient's arms are elevated above her head. This variant of skin dimpling and fixation does not represent a grave sign because it does not indicate direct involvement of the skin by the tumor. With progressive diffuse skin infiltration in the subdermal plexus and the characteristic involvement of Cooper's ligaments, there is extensive edema of the skin, referred to as peau d'orange.

Medullary Carcinoma

This cancer represents 2 to 15 percent of the histopathologic types and originates in large ducts. Grossly, the tumor is characterized by its soft, hemorrhagic bulky presentation. Commonly, the lesion is positioned deep within the breast and is mobile. The skin is often stretched over a bulky, spherical mass that exceeds 3 cm in diameter. There is usually delay in its initial progression, although rapid growth may occur secondary to tumor necrosis or hemorrhage. Bilaterality is reported in fewer than 20 percent of cases; fewer than 10 percent of these neoplasms contain detectable estrogen or progesterone receptors.

Microscopically, medullary carcinoma is characterized by: (1) a dense lymphoreticular infiltrate composed predominantly of lymphocytes and a variable number of plasma cells; (2) large pleomorphic nuclei that are poorly differentiated and accompanied by active cellular mitosis; and (3) a syncytial sheetlike growth pattern with minimal or absence of tubuloacinar differentiation (Fig. 14-28). Approximately one-half of these tumors are associated with intraductal cancer, with the intraductal component characteristically present at the periphery of the tumor mass. In rare circumstances, mesenchymal metaplasia or transformational anaplasia is noted.

Diagnosis of this lesion connotes a better 5-year survival than pure invasive ductal or lobular carcinoma. The most important prognostic determinant of medullary carcinoma is the presence or absence of axillary node metastasis. Because of the intense lymphohistiocytic response in and about the tumor, benign or hyperplastic enlargement of the nodes of the axilla may contribute to erroneous clinical staging. Metastases to axillary lymphatics are reported in more than 40 percent of patients.

Mucinous Carcinoma (Colloid Carcinoma)

This adenocarcinoma of ductal origin constitutes approximately 2 percent of all breast cancers and typically presents as a bulky, mucinous (colloid) tumor that is largely confined to the elderly population. The pathologic features of mucinous carcinoma are quite distinctive: the cut surface is glistening, glaring, and gelatinous. Fibrosis is variable and, when abundant, imparts a firm consistency to the tumor. Approximately one-third of patients have axillary metastases, and 5-year and 10-year survival rates are reported at 73 and 59 percent, respectively.

Characteristic microscopic features that are identifiable include large pools of mucin that surround variable groups of tumor cells. Tumor cells may not be evident in all sections. Signet-ring cells generally are not seen in mucin- producing adenocarcinomas of the breast. Approximately two-thirds of pure mucin-producing breast adenocarcinomas contain detectable estrogen receptors.

The lesion should be distinguished from benign *granular cell myoblastoma*. To confirm the malignant features of colloid carcinoma, multiple microscopic sections are essential. Frozen-section analyses are seldom diagnostic; relying on this technique is inadvisable.

Tubular Carcinoma

This lesion is a well-differentiated variant of breast carcinoma with an incidence of approximately 2 percent. Increasingly diagnosed mammographically, this tumor is reported in as many as one-fifth of women whose cancers are diagnosed by screening. Microscopically, tubular differentiation is distinctive. Under low magnification, a haphazard array of small, randomly arranged tubular elements are identifiable. The small glandular (tubular) pattern and single-cell lining of neoplastic tubules are important histologic characteristics of the tumor. Absence of myoepithelial cells and a well-defined basement membrane serve to distinguish common proliferative, microglandular, and sclerosing adenosis lesions from tubular carcinoma. Most commonly, the lesion is diagnosed in the perimenopausal or early menopausal population. These lesions are typically discovered mammographically when small (i.e., 1 cm maximum dimension).

Approximately 10 percent of patients with typical lesions develop axillary metastasis. Long-term survival approaches 100 percent if the carcinoma contains 90 percent or more of the tubular components. Metastases generally are confined to small numbers in low axillary nodes (Level I). Rosen and associates confirmed lower recurrences (3.5 percent) in patients treated for this disease.

Papillary Carcinoma

Papillary carcinoma accounts for less than 2 percent of all breast carcinomas and generally presents in the seventh decade. The lesion has been observed in a disproportionate number of non-Caucasian patients. Typically papillary cancer is small and rarely attains sizes greater than 2 to 3 cm in diameter. Morphologically, these cancers are well circumscribed; papillary differentiation in the form of papillae with well-defined fibrovascular stalks and multilayered epithelium may harbor moderately pleomorphic cells. McDivitt and colleagues noted that this tumor had the lowest frequency of axillary nodal involvement and the best 5-year and 10-year survival rates. Disease-free survival is similar to that for mucinous and tubular carcinoma. Despite the presence of axillary metastases, which may occur in up to one- third of patients, papillary carcinoma is a more indolent, slowly progressive disease than the common adenocarcinoma.

Adenoid Cystic Carcinoma

This lesion is very rare, accounting for less than 0.1 percent of breast cancers. It is typically indistinguishable from the more common adenoid cystic carcinoma that occurs in salivary glands. The age distribution is similar to that for typical adenocarcinoma. These cancers present as small lesions, 1 to 3 cm in diameter, characteristically well circumscribed with well- defined margins. On close inspection the tumors are found to contain dense mucoid material within glandular spaces that ultrastructurally mimics the lamina densa of the basement membrane. Axillary metastases are rare with adenoid cystic carcinoma. Only seven deaths from pulmonary metastases from this tumor have been confirmed.

Apocrine Carcinoma

These lesions present a ductal or acinar growth pattern with the unusual tendency to involve the lobular epithelium and are well-differentiated with rounded vesicular nuclei and prominent nucleoli. There is a very low mitotic rate and little variance in cytomorphologic features. These lesions can contain potentially aggressive biologic behavior; low to absent levels of estrogen and progesterone receptors are frequent.

Carcinoma of Lobular Origin

The histopathologic features include characteristic small cells with rounded nuclei, inconspicuous nucleoli, and scant, indistinct cytoplasm (Fig. 14-29). Special stains confirm the infrequent presence of intracytoplasmic mucin. Similar to colloid carcinoma, mucin may displace the nucleus, resembling signet-ring carcinoma of the gastrointestinal tract. These carcinomas originate in terminal ductules of the lobule and possess characteristic features that distinguish them from lesions of the larger, lactiferous ducts. The noninvasive variant is referred to as lobular carcinoma in situ (LCIS). In LCIS, lobules are packed with small hyperplastic cells of significant uniformity, arranged in rows or beads with few mitoses. Hyperchromatism, nuclear anaplasia, and other variants of invasive breast cancer are characteristic of the malignancy. The incidence of the in situ variant is approximately 3 percent of breast cancers; infiltrating lobular carcinoma constitutes approximately 10 percent of breast cancers.

Grossly, the infiltrating lobular variant deserves consideration, as these lesions vary from clinically inapparent microscopic tumors to those that replace the entire breast with a poorly defined, somewhat firm mass. Occasionally this cancer mimics inflammatory or benign lesions. Because the lesions have a high propensity for bilaterality, multicentricity, and multifocality, they sometimes present perplexing problems.

Squamous Cell (Epidermoid) Carcinoma

This infrequently observed cancer of epithelial origin arises from metaplasia within the lactiferous duct system. These cancers are typically devoid of distinctive clinical or radiographic characteristics. Similar to epidermoid carcinoma of the skin, metastases occur almost exclusively via the lymphatic route and are evident in approximately one-quarter of patients.

Sarcomas

Sarcomas of breast origin are a heterogeneous group of lesions. These tumors include *fibromatosis (low-grade fibrosarcoma or desmoid tumor)*, *fibrosarcoma*, *malignant fibrous histiocytoma*, *liposarcoma*, *leiomyosarcoma*, *osteogenic sarcoma*, and *chondrosarcoma*. *Stromal sarcoma* is a term that describes the diverse array of tumors of this type that are histologically identical with comparable soft-tissue tumors arising in extramammary sites.

The clinical presentation is typically that of a large, painless breast mass with rapid growth. Routine mammography is not a useful diagnostic aid; false-negative rates are high.

Morphologically, the tumors are predominantly of a solid type, although small cysts in degenerative areas may be observed. The typical sarcoma lacks the “cut cabbage” or leafy laminated surface configuration of the benign cystosarcoma phyllodes. Some sarcomas are well circumscribed, while others have infiltrative, ill-defined margins. Histologically, these lesions assume cellular features identical to those of malignancies of other body parts. These spindle-cell neoplasms grow as expansile, solid masses with microscopic margins that are sharp, pushing, or infiltrative. As a consequence, the lesions invade fat and tend to intervene between glandular aspects of the breast parenchyma to expand the lobules and intralobular spaces. Tumors are graded on the basis of cellularity, degree of cellular pleomorphism and nuclear atypia, evidence of differentiation, and mitotic activity.

Angiosarcoma

In 1948 Stewart and Treves described the syndrome of lymphangiosarcoma in patients with ipsilateral lymphedema following radical mastectomy. Angiosarcoma is the preferred term. It develops in a lymphedematous extremity that locally initiates an impaired immune mechanism. The average interval between the mastectomy and onset of the angiosarcoma is 10.5 years; 60 percent of patients have a history of postoperative radiotherapy to the operative site. Irradiation is considered a cofactor in the development of angiosarcoma only in the respect that it contributes to the development of lymphedema. The overall incidence of lymphedema following radical mastectomy is 15 to 25 percent, compared to 5.5 percent after modified radical mastectomy.

Poorly differentiated variants grow as solid nests and masses of spindle- shaped or epithelioid cells, the former mimicking Kaposi's sarcoma and the latter carcinoma. Exuberant mitotic activity, necrosis, and hemorrhage are evident in high-grade tumors. Less well-differentiated tumors grow as complex papillary proliferations of malignant cells forming anastomotic vascular channels. The typical presentation is a spectrum of differentiation with small capillary-sized vessels formed by atypical endothelial cells.

Ultrastructural features of postmastectomy angiosarcoma are identical with those of other types of angiosarcoma. Factor VIII–related antigen, a protein produced by endothelial cells, has been identified in this tumor and constitutes a reliable marker. This marker is not useful for distinguishing benign and malignant vascular proliferations; however, demonstration of factor VIII in an anaplastic cutaneous neoplasm excludes the diagnosis of carcinoma and melanoma.

The prognosis for patients with angiosarcoma is dismal; median survival is 19 months. No correlation has been observed between histologic features and survival. Radical forequarter amputation of the involved extremity has been proposed to manage the ulcerative complications of the arm and axilla and to palliate the massive progressive lymphedema. Five-year survivals are extremely rare.

Lymphomas

Primary lymphomas of the breast are rare. Presentation is that of a large lesion (mean size 4 cm) in the postmenopausal patient. DeCosse and associates noted a high incidence of tumor-positive axillary nodes. An occult breast lesion may be diagnosed after detection of palpable axillary lymphadenopathy.

Mammary lymphomas are identical to other malignant lymphomas, with tumor cells that are densely infiltrative throughout the breast parenchyma. There is a predominance of diffuse histiocytic lymphomas. Total mastectomy and axillary node sampling are advocated for large lymphomas of the breast. Recurrent or progressive local disease and/or regional disease is best managed by radiotherapy and multimodal systemic chemotherapy using protocols standard for non-Hodgkin's lymphoma. Prognosis is favorable, with 5-year and 10-year survival rates of 74 and 51 percent, respectively.

Inflammatory Carcinoma

The presence of lymphatic or vascular invasion portends a reduction in survival rate and a shorter disease-free interval. In *inflammatory carcinoma*, characteristic clinical features of erythema, peau d'orange, and skin ridging with or without the presence of a palpable mass are evident. This relatively rare entity constitutes approximately 1.5 to 3 percent of breast cancers. No specific histologic type predominates.

Typically the skin over the lesion is warm, diffusely scaly, and indurated with ridging. It may present with the characteristics of a cellulitis. An interval of treatment with antibiotics by the physician who mistakes this for a breast abscess is common. The tumor mass may be diffuse or nondefinable. The breast is diffusely “brawny” and the nipple is often retracted when the index lesion is subareolar. Diagnosis is established by generous biopsy of skin, subcutaneous tissue, and parenchyma.

Pathologically, subdermal lymphatics and vascular channels are permeated with microscopic foci of highly differentiated tumor. As many as 15 percent of patients free of axillary metastases have microscopic tumor emboli in tissues that surround the primary neoplasm. Inflammatory carcinoma refers to a clinicopathologic entity with characteristic absence of polymorphonuclear leukocytes and lymphocytes near the tumor.

This disease progresses rapidly, and more than three-quarters of patients have palpable axillary metastases at the time of presentation. It is important to distinguish inflammatory carcinoma from the contiguous extension of a scirrhou carcinoma that invades subdermal lymphatic spaces and skin to produce characteristic peau d'orange and lymphangitis of locally advanced disease (Fig. 14-30). Patients with inflammatory carcinoma have distant metastatic disease evident at a much higher frequency than that for more common breast cancers. Taylor and Meltzer found bone and visceral metastases in 36 percent of their patients. The extensive report of the Surveillance, Epidemiology, and End Results (SEER) Program by Levine and associates revealed metastatic disease at diagnosis in one-quarter of 3,171 white patients with inflammatory carcinoma.

Treatment

Historical Perspectives

The Edwin Smith Surgical Papyrus (3000–2500 b.c.) was the first document that referred to carcinoma of the breast. The lesion was in a man, but the description encompassed most of the clinical features of breast carcinoma. The author of the papyrus concluded, “there is no treatment [for cancer of the breast].” Few writings referred to tumors of the breast until the first century a.d. Direct reference to treatment of breast cancer is conspicuously absent in the Corpus Hippocraticum.

Celsus recognized the value of operations for early breast cancer in his early Roman writings of the first century a.d. A translation notes: “None of these can be removed but the cacoethes [early lesion], the rest are irritated by every method of cure. The more violent the operations are, the more angry they grow.” In the second century a.d., Galen inscribed one of the classic clinical observations:

We have often seen in the breast a tumor exactly resembling the animal the crab. Just as the crab has legs on both sides of his body, so in this disease the veins extending out from the unnatural growth take the shape of a crab's legs. We have often cured this disease in its early stages, but after it has reached a large size no one has cured it without operation. In all operations we attempt to excise a pathological tumor in a circle in the region where it borders on the healthy tissue.

The Galenic system of medicine ascribed neoplasms to an excess of “black bile” and concluded that excision of a local outbreak could not cure the systemic imbalance. Theories espoused by Galen dominated medicine until the Renaissance. Operative intervention was considered a misdirected, futile, ill-advised approach by the majority of established and respected physicians.

Beginning with Morgagni's systematic approaches, modalities that varied from those espoused by Galen became acceptable. More radical approaches for treatment of the breast became acceptable, including some early and primitive attempts at total mastectomy and axillary dissection. The procedure evolved slowly from simple amputation of the breast. LeDran repudiated Galen's humoral theory in the eighteenth century and stated that cancer of the breast was a local disease that spread by way of the lymphatics to regional nodes. He removed large axillary nodes in his operations on patients with cancer.

In the nineteenth century, Moore, of the Middlesex Hospital of London, emphasized wide removal of the breast and felt that the axillary contents should be removed en bloc, together with the breast, when neoplasm was evident in the axilla. In a presentation before the British Medical Association in 1877, Banks supported Moore's concepts and also advocated en bloc resection of axillary contents with the breast even when palpable nodes were not evident. Banks recognized that occult involvement of axillary nodes could be present.

In 1894 Halsted and Meyer simultaneously reported their operations for treatment of cancer of the breast. By demonstrating superior local and regional control rates after en bloc radical resection, these eminent surgeons established radical mastectomy as “state of the art” for that era. Both Halsted and Meyer advocated complete axillary dissection of all nodal levels from the latissimus dorsi muscle laterally to the thoracic outlet medially. Both routinely resected the long thoracic nerve and the thoracodorsal neurovascular bundle en bloc with the axillary contents.

D. H. Patey of Middlesex Hospital, London, is credited with the demonstration of the worth of the “modified radical mastectomy” technique. In the 1930s Patey was able to refute the unproved postulates on which the original radical operations were based for cases in which advanced local disease was not evident. Thereafter, Patey and colleagues developed the technique for in-continuity removal of the breast and axillary contents with preservation of the pectoralis major muscle. Removal of the pectoralis minor with retraction of the pectoralis major allowed access and clearance of the axillary contents. Madden and Auchincloss advocated the modified radical approach with preservation of both the pectoralis major and minor muscles. This restricts the dissection of the apical (Level III) nodes, and nodal recovery is less than with the Patey modified technique.

Patient Selection

As late as the 1960s, radical mastectomy was the only procedure used for the treatment of breast cancer. Surgeons attempted to exclude from operation patients who would almost certainly develop distant metastasis later. These concepts led to the adoption of the generally accepted principles espoused by Haagensen, which he referred to as “criteria of inoperability.” These general criteria included fixation of the local breast cancer to the chest wall, fixation of the involved lymph nodes of the axilla, and inflammatory carcinoma. This detailed list of criteria included:

- (1) Extensive edema of the skin over the breast (Fig. 14-31)
- (2) Satellite nodules in the skin over the breast
- (3) Carcinoma of the inflammatory type (see Fig. 14-30)
- (4) Parasternal tumor nodules
- (5) Proved supraclavicular metastases
- (6) Edema of the arm
- (7) Distant metastases
- (8) Any two or more of the following grave signs of locally advanced carcinoma:
 - (a) Ulceration of the skin
 - (b) Edema of the skin of limited extent (less than one-third of breast skin involved)

- (c) Solid fixation of tumor to the chest wall
- (d) Axillary lymph nodes measuring 2.5 cm or more in transverse diameter
- (e) Fixation of the axillary nodes to the skin or deep structures of the axilla

At the time these criteria were developed, more than one-quarter of patients were excluded from surgical therapy. Today, more than 10 percent of patients would be found to have such advanced tumors, and the utility of combined modality therapy would perhaps revert the majority to a simple extended procedure. The success of modern adjuvant chemotherapy and radiotherapy has greatly altered the approach to therapy; the majority of patients (80 percent) can at least expect local-regional control despite systemic metastases.

Current Therapy

Both patient and physician should have a clear perspective on the planned course of therapy. The physician should discuss with the patient the possibility that the suspicious lesion may be a cancer that necessitates a therapeutic regimen, including surgical options and the proposed multimodal therapy. The duration of therapy and the issue of irradiation also must be addressed.

Biopsy

The morbidity and mortality related to breast biopsy procedures are acceptably low. Local anesthesia can be used. A 1980 study on the cost- effectiveness of breast cancer management demonstrated that the most effective measure for containing economic morbidity of the illness is via targeted selection of high-risk patients for biopsy under local anesthesia. Although most (60 to 80 percent) biopsies of “suspicious” breast lesions prove to be benign, specific clinical and mammographic characteristics are associated with a high probability of malignancy.

Medical history, physical examination, and results of clinical staging each influence the timing and method of breast biopsy. Properly done, the biopsy is of immense value to determine subsequent work-up and definitive therapy. By contrast, inconclusive data derived from inadequately sampled tissue or a skin biopsy incision placed in the inappropriate skin contour or quadrant may limit therapeutic options and alter subsequent management.

Nonpalpable Lesions

The mammographically detected nonpalpable lesion may present in the “normal” breast without physical signs of an underlying cancer. In recent decades the wide application of screening mammography has resulted in the detection of increasing numbers of nonpalpable lesions. Specific criteria that commonly lead to the diagnosis of the nonpalpable breast mass include: (1) localized soft tissue mass within the breast parenchyma; (2) architectural distortion, including contracture of trabeculae, which produces stellate alterations, and asymmetry with thickening of the lobular or periductal architecture; and (3) clustered microcalcifications.

Despite the simplicity of mammographic techniques using craniocaudal and mediolateral views, intraoperative localization with adequate excision presents challenges that have led to the development of several methods, including noninvasive and invasive localization. *Noninvasive techniques for localization* include visual estimation, external breast markers, stereomammograms, plotted coordinates on a breast diagram, and grid compression devices. *Invasive localization methods* have improved remarkably over the past decade with the use of small radiopaque needles that may be radiographically guided into the suspicious lesion. After insertion of the localization needle, subsequent mammograms will demonstrate orientation of the needle tip to the suspect mass. This technique requires cooperation and communication between the radiologist and the surgeon, and specimen radiography is necessary. Figure [14-32](#) demonstrates the operative technique for needle localization biopsy.

Stereotactic core needle biopsy has become increasingly accepted as an alternative to needle localization biopsy for mammographically detected nonpalpable breast lesions. The advantages of this technique include a lower complication rate, decreased trauma to the patient, avoidance of scarring and disfigurement of the breast, and lower cost. The accuracy of stereotactic core needle biopsy is higher than that of stereotactic fine- needle aspiration and has improved significantly with the use of larger needles and a larger number of cores per sample. The incidence of insufficient sample is negligible. The immediate establishment of a diagnosis of breast cancer enables definitive surgical management of the disease with only one visit to the operating room.

Palpable Lesions

The biopsy technique for a palpable lesion often is influenced by the physical characteristics, the size, and the location of the suspicious lesion, the type of anesthesia desired by the patient, and the therapy planned if a malignancy is confirmed. An *incisional biopsy* of a large breast mass can be performed under local anesthesia if the patient presents with bony metastasis. This technique provides histologic confirmation of the malignancy and adequate tissue for hormonal receptor analysis before initiation of radiation or chemotherapy. *Fine-needle aspiration* of a small, suspicious palpable lesion is appropriate in an outpatient setting for the patient with clinical stage I breast cancer. Regardless of the method, it is essential that the biopsy specimen be handled expeditiously and appropriately to render a valid specimen for histologic and hormone receptor analyses.

Fine-Needle Aspiration (FNA)

This technique usually is performed when a palpable mass is evident. The combination of physical examination, mammography, and FNA provides a diagnostic accuracy that approaches 100 percent. A negative FNA cytology in the presence of a palpable mass, however, does not conclusively exclude carcinoma. When the mass is clinically and mammographically suspicious, the sensitivity (true-positive) of FNA is 80 to 98 percent. The false- negative rate of FNA is 2 to 10 percent. Sensitivity and efficacy are influenced primarily by false-negative results. The specificity and predictive value of FNA approach 100 percent because false-positive results are rare. Fine-needle aspiration of nonpalpable, mammographically detected breast lesions can be performed with stereotactic technique. However, the stereotactic fine-needle aspiration is less accurate and reliable than the stereotactic core needle biopsy, and therefore the latter is preferable for nonpalpable breast lesions.

Cutting Needle Biopsy

The standard Tru-Cut needle (Travenol, Deerfield, IL) is the most commonly used cutting needle for biopsy of breast masses. False-positive diagnostic rates are lower with tissue procured by cutting needles than with FNA specimens because more tissue is submitted for analysis. But a core biopsy specimen without malignant tissue cannot conclusively be considered a “negative” biopsy, as it might reflect a sampling error.

Incisional and Excisional Biopsies

Both techniques procure suspicious breast tissue to be submitted for microscopic examination. The *incisional technique* is indicated for patients with large (4 cm or larger) primary lesions for whom preoperative chemotherapy and/or radiation therapy is desirable. The surgeon should carefully incise tissues that are not necrotic to permit histologic and hormonal receptor analyses. Tissue procured from both incisional and excisional techniques should be obtained with the cold scalpel, because electrocautery may distort histologic features of the tumor and invalidate tissue levels of hormonal receptors.

Excisional biopsy implies removal of the entire lesion and generally a margin of normal breast parenchyma surrounding the suspicious lesion. The surgeon should avoid transection or disruption of the lesion for fear of tumor implantation. When the volume of tissue excised is small (<1 cm³) permanent histologic sections should be planned, as it may be difficult pathologically to differentiate an invasive carcinoma from severe atypia or in situ disease on frozen-section specimens. Both incisional and excisional biopsies can be closed in layers with absorbable sutures.

Planning Incisions

Figure [14-33](#) illustrates nonradial approaches for cosmetically acceptable scars of the breast. Radial incisions in the upper half of the breast are ill- advised because of scar contracture and asymmetric displacement of the nipple-areola and parenchyma. Incisions should be cosmetically designed, since approximately 70 percent of the biopsies confirm benign (proliferative and nonproliferative) disease. Lines of tension in the skin of the breast (Langer’s lines) are generally concentric with the nipple. Incisions that parallel these lines generally result in thin, cosmetically acceptable scars. It is important to keep incisions within the boundaries of potential incisions for future mastectomy or wide-local excision that may be required for definitive treatment (Fig. [14-34](#)). The most cosmetically acceptable scars result from circumareolar (curvilinear) incisions. Centrally located subareolar lesions are best approached in this manner.

Biopsy is best performed with a scalpel rather than electrocautery. Hormonal receptor profiles should be obtained from index tumor tissue. Use of residual tumor from the mastectomy specimen for estrogen and progesterone receptor studies may provide invalid hormone receptors, especially when warm ischemia time is prolonged in the course of the procedure.

After scalpel removal of the lesion, electrocautery of bleeding sites or absorbable suture ligatures are used to achieve hemostasis. Wound drainage with soft rubber Penrose drains (¼-inch) is optional. Although closure of the breast tissue defect is not mandatory, we recommend closure with interrupted 2-0 or 3-0 absorbable chromic gut or polyglycolic sutures. Subcutaneous tissues can be closed with 3-0 or 4-0 absorbable sutures. A running subcuticular suture of the skin using 5-0 synthetic suture is performed, followed by approximation of the defect with adhesive tape strips (Steri-Strips). A light occlusive dressing is applied.

For most palpable and nonpalpable breast masses, local anesthesia and sedation incur minimal morbidity and no mortality. When general anesthesia is required for biopsy of a suspicious breast mass, a full work- up with clinical staging and plans to implement comprehensive therapy should be completed before the procedure to avoid a secondary operation. If needle localization is required, the operating room personnel must arrange, preoperatively, for *specimen radiography* to be performed as soon as the tissue is procured. In the case of a suspicious lesion, an informed consent for the previously agreed treatment should be obtained; if frozen section confirms an invasive neoplasm, this allows the surgeon to proceed with definitive therapy. After biopsy excision of breast tissue, three dimensions of the tumor should be carefully marked with suture or clips and recorded; the surgeon should orient the pathologist to these findings if subsequent margins are questioned on final pathologic review.

Therapeutic Options

The past three decades have witnessed significant progress in multimodality therapy for the treatment of breast cancer and the integration of these modalities to enhance survival and, in appropriate patients, to employ conservation surgical principles. In addition, large breast cancers (>3 cm) with matted axillary metastases usually are treated with preoperative chemotherapy to initiate a cytoreductive effect for the index tumor and the metastases. This approach allows the surgeon to complete a planned mastectomy without the use of skin grafts. After mastectomy, irradiation of the chest wall and the internal mammary and supraclavicular nodes eradicates residual microscopic disease in the skin flap or regional nodes outside the operative field. This multimodal approach enhances local-regional control of the advanced primary tumor and has the additional benefit of treating

potential systemic metastases with the preoperative chemotherapy.

The therapeutic objective of the surgeon and the radiation therapist is local-regional control; the objective of the medical oncologist is control of systemic disease. While cytotoxic agents, especially in combination, are effective for breast cancer metastases, *complete eradication* of documented metastasis is theoretically impossible because of the log-kill cytokinetics of chemotherapeutics. A small fraction of cells that enter the G₀ phase outside the cell cycle and are nonreplicating will potentially develop resistance or become refractory when exposed to these cytotoxic agents. Prospective trials in North America and Europe have allowed the clinician to predict control rates, survival rates, and prevailing toxicities. It is the integration of surgery, radiation therapy, and chemotherapy that has achieved the unprecedented response rates afforded to patients with this disease.

Therapy of Early Breast Cancer and In Situ Disease

Adenocarcinomas of the breast in which the diameter is less than 5 cm, limited to the lateral aspect of the breast, and with involvement of the pectoral fascial or skin fixation (T1a, T2a) often are treated by surgery alone, provided that lymph node metastases are absent in the pathology specimen. Ductal and lobular carcinoma in situ are not invasive lesions and therefore do not have nodal metastases. These too are treatable by surgery alone, and in most clinics axillary node dissections are not performed. By contrast, when lymphatic metastases are evident pathologically and when the adenocarcinoma is medially or centrally positioned in the breast, the combination of surgery and postoperative radiation therapy is often used to ensure local and regional chest wall control.

Lobular Carcinoma In Situ (LCIS)

LCIS is considered a *marker* for increased risk rather than an inevitable *precursor* of invasive disease. The goal of observation is to detect subsequent invasive cancers that develop in a minority of these patients. If subsequent invasive cancers do develop, bilateral mastectomy will have a high probability of cure. There is no demonstrable benefit to widely excising LCIS and obtaining clear margins if nonoperative observation is chosen, because the disease is assumed to diffusely involve all breast tissue as well as the contralateral breast.

A 5 percent rate of associated invasive carcinoma and a high rate of multicentricity and bilaterality are the major arguments that support operative therapy of the disease. Routine bilateral mastectomy for LCIS appears to be an aggressive radical approach for a lesion that has a low risk potential. Routine contralateral biopsy in the absence of standard indications is not justified, because the likelihood of identifying a lesion requiring therapy (i.e., invasive carcinoma or DCIS) is minimal and the clinical significance of other foci of LCIS is negligible. If operation is chosen, anything less than total mastectomy is inappropriate, because the disease process is diffuse and often bilateral. In some instances bilateral mastectomy and reconstruction are applicable. The incidence of axillary node metastasis is generally less than 1 percent; routine dissection is not advisable, but sampling of Level I nodes in conjunction with the mastectomy adds virtually no morbidity to the procedure.

Ductal Carcinoma In Situ (DCIS)

For many decades unilateral mastectomy has been used for this lesion. Mastectomy is still considered the “gold standard” against which lesser procedures must be evaluated. In combined data from 14 series, a total of 1,061 women with DCIS were treated with mastectomy; a local recurrence rate of only 0.75 percent and a mortality of 1.7 percent were demonstrated. Ipsilateral therapy is adequate because the overall incidence of occult disease (10 to 15 percent) is essentially equivalent to that for invasive carcinoma. Breast reconstruction can be performed at the time of the ipsilateral mastectomy (Fig. [14-35](#)).

The excellent results achieved with breast-sparing therapy of invasive disease have led to its use for DCIS. Rosner and colleagues showed that wedge resection of DCIS was equivalent to mastectomy with regard to overall survival. Follow-up intervals of up to 14 years for DCIS treated by local excision alone suggest local recurrence rates that range from 10 to 63 percent. Lagios and associates and Gump and associates note that recurrence is greatest for palpable tumors of large size (>25 mm), when the criteria for confirmation of pathologically clear margins are less rigorously applied, and when the histology is the comedo type. Recurrences are noted most often within the original biopsy site, implicating inadequate marginal clearance rather than intrinsic biologic behavior of DCIS. Combined data from 14 studies of conservative surgery couple with postoperative radiation therapy in 1,098 women with DCIS showed an overall recurrence of 9.1 percent with 45 percent of the recurrent lesions being invasive carcinomas.

Results of prospective randomized trials assessing the value of radiation therapy in the treatment of DCIS are relatively limited. The first published trial containing such results was the report on Protocol 6 of the National Surgical Adjuvant Breast and Bowel Project (NSABP). Although the trial was designed to study the value of radiation therapy in addition to lumpectomy for invasive breast cancer, 51 patients included in the study were later identified as having only DCIS. Of these 51 patients, 22 had lumpectomy alone and 29 had lumpectomy followed by breast irradiation. The recurrence rate in the lumpectomy-alone group was 23 percent, compared to 7 percent in the lumpectomy-plus-radiation group. The short follow-up period of only 39 months combined with the relatively large size of the tumors and some uncertainty regarding the assurance of negative tumor margins, create some doubt about the validity of these results. The results of Protocol 17 of the NSABP, which was designed to compare lumpectomy alone and lumpectomy plus radiation therapy for the treatment of DCIS, suggested an enhanced 5-year event-free survival in the women who had lumpectomy followed by breast irradiation. The 5-year cumulative incidence of ipsilateral breast cancer was reduced by breast irradiation from 10.4 percent to 7.5 percent for noninvasive cancers ($p < .05$) and from 10.5 percent to 2.9 percent for invasive cancers ($p < .001$). The mean duration of follow-up was 43 months. This is a relatively short follow-up period for patients with DCIS, and a longer period of assessment will be necessary before definitive conclusions can be drawn. An additional limitation of this study is the fact that 42 to 45 percent of the tumors were less than 0.1cm, 30 to 31 percent were 0.1–1.0 cm, and only 8 percent were more than 2 cm. Since the majority of the tumors in this study were smaller than 1 cm or microscopic, extrapolating the results of the treatment to larger tumors should be done with caution. Protocol 17 also did not address the histopathology or grade of the DCIS, which are considered predictors of recurrence in other studies. Whether or not the results of breast-sparing treatments are comparable to the excellent results achieved by mastectomy has not been addressed.

An additionally important issue with the use of breast-sparing treatment for DCIS is that over one-half of all recurrences are invasive. Studies suggest that local recurrence after conservation treatment may be successfully treated with “salvage mastectomy,” a technique that does not carry the ominous prognosis documented after recurrence following mastectomy. Other studies suggest that local/regional recurrence in this setting may result in diminished survival. In the absence of conclusive evidence, caution is warranted, as the safety and efficacy of breast-sparing therapy for DCIS are less certain. Until data are available from ongoing prospective trials, it would be prudent to offer mastectomy for lesions that suggest a substantial risk of local/regional recurrence and reduced survival (palpable mass > 25 mm, comedo histology, extensive multicentricity, multifocality, high nuclear grade, negative estrogen and progesterone receptor studies, aneuploid DNA pattern, high proliferative index, etc.). Breast-sparing therapy may be offered to women with DCIS when uncertainties and risks are fully discussed. Mastectomy is recommended when attempts at wide local excision reveal extensive foci of residual DCIS and when margins cannot be cleared pathologically. We continue to recommend postoperative radiotherapy after conservation treatment. No role currently exists for use of cytotoxic chemotherapy in this disease. Axillary nodal dissection beyond Level I generally is not recommended in view of the low yield of positive nodes (<2 percent). Comprehensive lifelong surveillance is indicated for the contralateral breast in women previously treated for DCIS.

Stage I and Stage II Breast Cancer

In the United States in 1972 45.3 percent of patients were reported to have had a Halsted-type radical mastectomy. In 1981 3.4 percent of patients underwent the procedure, and in 1990, only 0.4 percent. Conversely, partial mastectomy was performed in only 3.4 percent of patients in 1972, 7.2 percent in 1981, and 25.4 percent in 1990. These changes in the distribution of patients by the type of operation performed are shown in Table [14-9](#). Although breast conservation surgery has been used with increasing frequency, modified radical mastectomy remains the most commonly used surgical treatment for carcinoma of the breast. The proportion of patients receiving adjuvant radiation therapy has remained around one-fifth or one-sixth over the past three decades, while the proportion of those receiving cytotoxic chemotherapy, tamoxifen, or both climbed steadily, from 17.6 percent in 1976 to 46.6 percent in 1990 (Table [14-10](#)).

Evaluation of Modified Radical Mastectomy

In the 1982 American College of Surgeons Survey, 5-year survival rates by stage, type of treatment, and type of adjuvant therapy were evaluated. Similar survival rates for patients with localized disease were observed for treatment by partial mastectomy alone and by partial mastectomy plus irradiation of the breast, the axilla, or both. The five-year survival rates for patients treated by modified radical mastectomy were equivalent to those of patients treated by the Halsted radical mastectomy. Wilson and associates noted that survival rates were similar for those who received additional radiotherapy or chemotherapy with either of the two procedures.

There has been a transition in curative surgical procedures used by American surgeons from the Halsted radical to the modified radical mastectomy procedure. The transition was apparent at the time of the 1977 survey reported by Nemoto and colleagues, and a 1981 analysis confirmed that the vast majority of patients treated had a modified rather than a radical mastectomy (77 percent versus 3 percent). This change was based on the conclusion that extirpation of the pectoralis major muscle is not essential to provide local/regional control for stage I and stage II disease. Also, either the modified radical mastectomy or the Halsted procedure *alone* is inadequate procedure for achieving local/regional control of TNM stage III and Columbia Clinical Classification C and D tumors.

Subsequent prospective trials for modified radical mastectomy as compared to the radical procedure have included the Manchester Trial, reported by Turner and associates, and the University of Alabama Trial, reported by Maddox and associates. In both studies, recurrence rates for stage I and stage II patients were comparable at 5 years. Disease-free survival was different at 10 years in the analysis by Maddox and colleagues, and trends favored the more radical procedure.

Fisher, Safer, and Fisher confirm that biologic rather than anatomic factors are responsible for metastatic dissemination. They demonstrated that hematogenously carried tumor cells enter the lymph nodes, concluding that hematopoietic and lymphatic systems are unified as routes of tumor cell dissemination. Because total (simple) mastectomy is designed to treat local or regional disease, some authors have proposed that the addition of the regional node dissection should *not* influence survival. These series have made comparisons of total mastectomy with and without radiation.

Five-year survival rates for clinical TNM stage I tumors range from 51 to 78 percent; for clinical TNM stage II cancers the range was 33.7 to 71 percent. Radical radiotherapy administered to the peripheral lymphatics in these nine series appeared to have had little overall benefit at 5 years. At 10 years the absolute survival appeared to be improved for patients who underwent radical irradiation of peripheral lymphatics. Survival rates achieved with total mastectomy (with or without radiotherapy) are comparable to those obtained with radical mastectomy.

In the original report of the Cancer Research Campaign of the United Kingdom, Kyle and associates compared results of a “radical” therapeutic regimen (total mastectomy plus radiotherapy) with those of a conservative policy (total mastectomy alone) in a prospectively controlled study of 2,268 patients. At 5-year follow-up, although there was no evidence that routine postoperative radiation therapy was detrimental to wound repair, this modality was found to confer no additional benefit for survival or distant recurrence. Radiotherapy did, however, significantly reduce the incidence of local/regional recurrence. Further, a fourfold reduction in chest wall recurrence and a threefold reduction in supraclavicular recurrence was evident at 5 years in the treatment group. Similar trends were observed for recurrence in the operative site.

An identical trial was conducted by Berstock and associates with follow-up periods ranging from 9 to 14 years (median 11.4 years). Analysis confirmed no significant difference in survival and distant recurrence between the two treatment groups. Prophylactic postoperative radiotherapy reduced the risk for development of local recurrence.

In the randomized trial conducted by Langlands, Prescott, and Hamilton, simple mastectomy and radiotherapy had an overall survival equivalent to radical mastectomy. At 12-year follow-up they confirmed that survival in the radical mastectomy treatment group was significantly better, but only for those with clinical stage I disease. There was a significant prolongation of survival after detection of recurrence in the radical mastectomy group; this was greatest when local recurrence and distant metastases coincided.

The National Surgical Adjuvant Breast and Bowel Project B-04 trial conducted by Fisher and collaborators compared local and regional treatments of breast cancer. Life table estimates were obtained for 1,665 women enrolled for a mean of 120 months. This NSABP trial randomized patients with *clinically negative* axillae into three groups: (1) Halsted radical mastectomy (RM); (2) total (simple) mastectomy plus local/regional radiotherapy (TM + RT); and (3) total mastectomy alone (TM). Removal of axillary regional nodes was to be completed only when nodes became clinically positive. *Clinically node-positive* patients were treated with RM or TM + RT. Adjunctive chemotherapy was not given to patients in any of the three randomization groups.

For patients treated by total mastectomy and regional radiotherapy versus those treated by total mastectomy alone, no differences were observed between treatment groups for patients with clinically positive nodes or clinically negative nodes with respect to disease-free, distant-disease-free, or overall survival at 10-year follow-up. Ten-year survival was approximately 57 percent for node-negative and 38 percent for node-positive patients. These investigators conclude that *variations of local and regional therapy were not important in determining survival of patients with breast cancer*. Further, results obtained at 5 years accurately predicted outcome at 10 years. Despite similarity in survival, chest wall recurrence was significantly greater at 10-year follow-up for stage I patients treated with mastectomy alone (5.2 percent) than those treated with total mastectomy and radical irradiation (0.9 percent).

Figure 14-36 confirms that there was no significant difference in disease-free survival during the entire period of follow-up among groups of patients with *clinically negative nodes* treated by RM, TM + RT, or TM (panel A of Fig. 14-36). When disease-free survival was evaluated in the first and second 5-year periods of follow-up, Fisher and colleagues observed no differences among groups within the first 5 years after surgery (panel B). Panel C of Fig. 14-36 confirms no statistical differences in the probability of failure among the three groups during the *second* 5 years of follow-up. For each group, approximately 75 percent of patients who were free of disease at the end of 5 years remained so at the end of the tenth year. Patients undergoing TM + RT also had a lower incidence of local and regional recurrence than did those in the other two treatment groups.

Fig. 14-36 further confirms no significant differences in disease-free survival among patients treated with RM or TM + RT for individuals who presented with *clinically positive nodes*. These data were consistent for the first and second 5-year follow-up intervals. In addition, little difference between the two groups was observed in occurrence of distant, local, or regional disease (Fig. 14-37).

Figure 14-38 illustrates the probability of survival for node-negative and node-positive patients. For node-negative patients among the three treatment groups, no significant differences were evident for distant-disease-free or overall survival. These data also were confirmed for the two treatment groups with positive nodes. Approximately 75 percent of patients with negative nodes who were alive at 5-year follow-up remained alive at 10 years, and approximately 65 percent of those with positive nodes who lived 5 years survived another five years.

Figure 14-39 further confirms the relationship of treatment to survival according to tumor *location* in the NSABP B-04 study. For patients with clinically negative nodes in medial-central or laterally located tumors, no differences in outcome were observed among the three treatment groups. Similarly, for patients with clinically positive nodes, treatment was not observed to affect survival in those with lateral or medial-central tumors. Thus, location of breast tumor does not influence prognosis, and irradiation of the internal mammary chain in patients with inner quadrant lesions does not improve survival. This study served to reiterate the important principle that variations of local and regional treatment parameters have less importance in determining survival of the patient with breast cancer than originally presumed.

Mastectomy

Both the Halsted radical mastectomy and the Patey modified radical mastectomy necessitate en bloc resection of the breast, the axillary lymphatics, and overlying skin near the tumor with a 3-cm to 5-cm margin that ensures histologic clearance of the tumor. The Patey mastectomy acknowledges the importance of the complete axillary dissection and the anatomic necessity for preservation of the medial and lateral pectoral (anterior thoracic) nerves, which may provide dual innervation to the pectoralis major muscle. The Halsted mastectomy necessitates resection of the pectoralis major by virtue of size of the lesions (T2, T3, T4) that present with gross infiltration (fixation) of the skin or pectoralis major and for peripheral (high-lying) lesions near the clavicle in patients who are otherwise not candidates for radiotherapy. Major considerations for operations less extensive than the classic Halsted mastectomy are based on tissue preservation to enhance the cosmetic result. The modified radical mastectomy, with removal of the pectoralis minor muscle (Patey dissection), allows access to Level III nodes. The Patey modified technique is intended for lesions that cannot be removed with clear margins by segmental mastectomy and for lesions of large size (>T2, >5 cm) in which cosmetic reconstruction and regional control cannot be accomplished. The modified radical technique is not intended for large tumors with evidence of skin or pectoral muscle fixation, for which resection of the muscle is necessary to achieve adequate margins.

The Madden and Auchincloss mastectomies advocate preservation of both the pectoralis major and minor muscles, thus allowing adequate access to Level II lymphatics with incomplete dissection (or preservation) of apical (Level III) nodes. These approaches (Figs. 14-40, 14-41, 14-42, 14-43) require total mastectomy with at least partial axillary lymph node dissection. With limitation for dissection of the apical (subclavicular) nodal group, the Auchincloss and Madden procedures allow higher probability for preservation of the medial (anterior thoracic) pectoral nerve, which courses in the lateral neurovascular bundle of the axilla and commonly penetrates the pectoralis minor to supply the lateral border of the pectoralis major.

Regardless of the skin incisions chosen, the limits of the modified radical mastectomy are delineated *laterally* by the anterior margin of the latissimus dorsi muscle, *medially* by the midline of the sternum, *superiorly* by the subclavius muscle, and *inferiorly* by the caudal extension of the breast some 2 to 3 cm inferior to the inframammary fold (see Fig. 14-40, *inset*). The surgeon should be cognizant of the thoracodorsal nerve, whose origin is medial to the thoracodorsal artery and vein en route to innervation of the latissimus dorsi. With medial dissection, the surgeon encounters the chest wall deep in the medial axillary space and is able to identify the long thoracic nerve (Bell's respiratory nerve) in the deep investing fascia of the serratus anterior muscle. This nerve is constant in its location anterior to the subscapularis muscle and is closely applied to this fascial compartment. Every effort should be made to preserve the long thoracic nerve; otherwise, permanent disability with a "winged" scapula and shoulder apraxia will follow denervation of the serratus anterior.

After completion of the extirpation of breast tissue from the chest wall, the entire mastectomy specimen and axillary contents are submitted en bloc for pathologic analysis. *Estrogen receptor (ER)* and *progesterone receptor (PR)* activity studies should be obtained on all pathologic breast cancer specimens to aid therapeutic planning of endocrine replacement treatment should metastatic disease occur. The results of these studies will guide future management for neoadjuvant therapy as well. Despite the importance of ER and PR activity to guide future therapies, processing of neoplastic tissue for pathologic examination, in all cases, must take precedence over determination of steroid receptor activity. Procurement of tissues for pathologic diagnosis and determination of qualitative and quantitative steroid receptor activity is best accomplished with the cold scalpel. This technique avoids the possibility with electrocautery of heat-induction artifact, tissue necrosis, cellular death, and temperature-dependent inactivation in steroid receptor activity and in procured tissues.

The responsibility of the surgeon and the radiation therapist is to provide the patient with the greatest probability of control of chest wall disease and to ensure minimal morbidity and mortality related to the therapy. After complete dissection of the axilla for Levels I, II, and III nodes, axillary irradiation should *not* be used, as the incidence of lymphedema to the ipsilateral extremity is six to eight times higher with this combination of modalities. Major lymphatics are removed surgically, and residual lymphatic collaterals may be destroyed by irradiation. In addition, the axilla dissected for operable disease should not require irradiation after radical or modified radical mastectomy unless extracapsular involvement of the lymphatics or tumor implantation in soft tissues of the axilla is evident.

Conservation Surgery

Conservation surgery of the breast implies the resection of minimal volumes of diseased breast tissue to achieve control rates equivalent to those accomplished by mastectomy. It has the goal of preservation of cosmesis and function. These procedures are variously termed *segmental resection*, *lumpectomy*, or *tylectomy*.

The breast may be preserved when adequate removal of all primary breast cancer can be accomplished without incision into cancerous tissue in appropriately selected patients. In all circumstances, frozen-section evaluation and permanent-section analyses of resected margins should be performed to ensure that all breast cancer has been removed en bloc with the specimen. Margins that harbor residual breast cancer warrant further excision. If clearance of tumor margins is not possible, or if multicentric disease is evident, total mastectomy is appropriate.

After reconstruction of the peripheral breast tissues in the operative site, sampling of ipsilateral axillary lymphatics is completed. To determine the necessity for adjuvant chemotherapy, the status of the axillary lymphatics must be determined. Adequate sampling is accomplished via a curvilinear incision between the lateral border of the pectoralis major and latissimus dorsi muscles 4 to 6 cm below the apex of the axilla. Lateral axillary contents that would be removed with an extended simple mastectomy (Level I) are taken with approximately 10 to 15 lymph nodes in the sampling. This volume of lymph nodes assures adequate sampling that is indicative of the regional nodal status. Some clinics demand complete axillary dissection of Level I through Level III. Subsequent irradiation of the axilla should not be performed, in order to avoid lymphedema of the ipsilateral extremity.

Indications for lumpectomy, axillary sampling, and comprehensive irradiation of the ipsilateral breast include: (1) a small breast cancer (<4 cm in transverse diameter); (2) clinically negative axillary lymphatics; (3) breast volume of adequate size to allow a uniform dosage of radiation; and (4) a radiation therapist experienced with the technique. Excessive radiation doses or nonhomogeneous distribution of the radiotherapeutic field may initiate painful, edematous, ulcerative, and/or fibrotic residual breast tissue. As experience with this treatment modality has increased, the indications have been extended successfully to larger-breasted women and to women with small but clinically positive axillary lymph nodes.

It is the surgeon's responsibility to ensure complete removal of the cancer within the breast. If viable cancer cells remain within the breast parenchyma at the periphery of the resection, they will be incorporated into the desmoplastic response of scarring and become poorly oxygenated. Marginally or poorly oxygenated anoxic cells entrapped within scar tissue may not be eradicated by irradiation; recurrence of breast cancer in the scar would be anticipated. Data confirm that clearance of the surgical margins of the index lesion is required to minimize the chances of local recurrence and to enhance cure and control rates. It is the practice of most North American and European clinics to reexcise scar in such patients and complete the axillary sampling procedure. If reexcision with histologically negative margins is not obtainable, these patients are best treated by total mastectomy. Approximately one-half of patients who have had scars reexcised were found to have viable cancer cells present in the wound margins after what was originally deemed an adequate segmental mastectomy. For patients in whom reexcision of the scar has not been advocated, external beam radiation boosted by implantation of iridium (¹⁹²Ir) needles in the area of the scar has been used.

Conservation surgery for patients who meet the above criteria results in long-term disease control and survival data that are equivalent to those achieved in patients treated by

modified radical mastectomy. Protocol B-06 of the National Surgical Adjuvant Breast and Bowel Project compared total mastectomy to lumpectomy with or without radiotherapy in the treatment of stage I or stage II breast cancer. The results of this study after 5 and 8 years of follow-up showed that lumpectomy with or without irradiation of the breast resulted in rates of disease-free, distant-disease-free, and overall survival that were not significantly different from the rates observed after total mastectomy. The results also showed that there was no statistically significant difference in the rates of survival with or without radiotherapy among the women who underwent lumpectomy, although the incidence of ipsilateral breast tumor recurrence was higher in the nonirradiated group. These findings support the use of lumpectomy in the treatment of stage I and stage II breast cancer and also indicate that irradiation of the breast reduces the probability of local recurrence after treatment of breast cancer by lumpectomy.

Reanalysis of this study after 12 years of follow-up was performed after the discovery of falsified information on patients enrolled in the study by one of the participating institutions. Separate audits were performed, and the results were recently reported. Again, no statistically significant differences were found in overall disease-free and distant-disease-free survival after mastectomy and lumpectomy alone or with radiotherapy. However, the cumulative incidence of ipsilateral tumor recurrence was significantly higher in the lumpectomy-alone group (35 percent) as compared to the lumpectomy-plus-radiotherapy group (10 percent, $p < .001$). These findings suggest that lumpectomy followed by breast irradiation is an appropriate treatment for patients with breast cancer 4 cm or less in diameter with negative or positive axillary lymph nodes. These findings are detailed in Table [14-11](#) and Fig. [14-44](#).

The long-term effects of irradiating the breast are not yet known, and patients wishing conservation surgery should be apprised of the uncertainty of the prolonged effects of radiobiologic injury. Properly done, lumpectomy, axillary sampling, and comprehensive irradiation of the remaining breast can provide a very satisfactory functional and cosmetic result. Breast cancer cannot be considered cured, however, by lumpectomy alone. Several studies have shown the disease to be potentially multicentric within the ipsilateral breast, requiring irradiation for sterilization of residual foci of invasive microscopic disease.

If lymphatics removed at the time of axillary sampling confirm histologic metastases, *adjuvant chemotherapy* should be considered in the postoperative period. For approximately 20 to 30 percent of patients, clinically negative axillary lymph nodes will be proved pathologically positive. In many clinics, the concomitant administration of adjuvant therapy (CMF—cyclophosphamide, methotrexate, and 5-fluorouracil) with comprehensive irradiation is being used with minimal toxic systemic effects.

QU.A.R.T. Procedure (Quadrantectomy, Axillary Dissection, and Radiation Therapy)

Quadrantectomy implies that a quadrant of the breast that harbors carcinoma is resected. Resection of an entire quadrant in a small- or medium-sized breast can produce an unacceptable cosmetic result. This procedure has yielded excellent local control and survival results when followed by radiation therapy and axillary dissection. The technique aims to remove an entire quadrant of the breast, including the skin and superficial pectoralis fascia. The objective is radical removal of the primary tumor and potential foci of infiltration via en bloc excision of one-fourth of the entire breast. Veronesi notes that it is difficult to do an appropriate operation for tumors whose diameter exceeds 2 to 3 cm unless the breast is large. The segmental mastectomy, however, has been applied in moderate-size breasts for lesions as great as 4 cm, so long as 1-cm margins are obtained. The QU.A.R.T. procedure provides a more radical resection (≥ 2 cm from the biopsy incision) of the tumor.

Veronesi suggests that the axillary dissection be performed in continuity; excision of a previously made biopsy incision allows adequate exposure of the axilla. The en bloc operation is generally performed when the primary tumor is situated in the upper outer quadrant close to the axilla. For index lesions in other quadrants, quadrantectomy is performed separately from the axillary dissection, which incorporates nodes of Levels I through III (Figs. [14-45](#) and [14-46](#)). Comprehensive irradiation of the intact breast is an important component of therapy. A dose of 50 Gy is delivered through two opposing tangential fields with high-energy photons (cobalt unit or a 6 MeV linear accelerator), and another dose of 10 Gy is given with orthovoltage radiotherapy as a booster to the skin surrounding the scar.

Figure [14-47](#) depicts equivalent disease-free survival curves for the QU.A.R.T. procedure and the classic Halsted radical mastectomy. These data apply to node-negative and node-positive disease confirmed at axillary dissection (Fig. [14-48](#)). An algorithm depicting treatment options for early stage breast cancer is presented in Fig. [14-49](#).

Future Developments

Prospective trials conducted by Veronesi for the QU.A.R.T. procedure and Fisher and colleagues of the NSABP for segmental mastectomy indicate that conservation surgery results in equivalent disease-free and overall survival compared to more radical procedures. Veronesi makes two important conclusions about the value of these procedures: (1) with specific reference to QU.A.R.T., the operation appears to be safe and allows preservation of the majority of breast substance and form; and (2) the patient treated with *inadequate* local-regional surgery or *inadequate* radiotherapy are at greater risk for local-regional recurrences and thus of overall and disease-free survival.

International trends favor a reduction in radical procedures, with increasing selection of breast conservation. An increasing realization has emerged that lumpectomy *without* radiation may be suitable for an undefined proportion of patients with early breast cancer. At 12-year follow-up, the B-06 protocol of the NSABP demonstrated no statistically significant reduction in survival (distant-disease-free or overall) for patients who had local excision *alone* compared to those who underwent local excision with radiation therapy for cancers 4 cm or less in transverse diameter. The trend for management of this disease over the past decade has been toward selective breast-sparing procedures with restriction of radiation therapy (i.e., elimination of irradiation of axillary, supraclavicular, and internal mammary node sites), and more liberal application of systemic adjuvant therapy.

We can expect the subset of patients treated by local excision *without* radiotherapy to increase. These patients will be the beneficiaries of early detection, definition of risk for local recurrence, pathologic analysis, and the emerging application of the adjuvant antiestrogen agent tamoxifen. Expanded therapeutic regimens will invoke the application of sophisticated genetic, biochemical, and pathologic prognostic indicators to verify and predict recurrence following conservative management parameters.

The role of axillary lymphadenectomy in the management of early breast cancer has been questioned in recent years, though it remains a required component of the surgical management of invasive breast cancer. Lymphatic mapping and sentinel lymphadenectomy has been shown in a single-institution trial to provide accurate staging of breast cancer without the morbidity associated with axillary lymphadenectomy. A multi-institution trial for the assessment of this technique is under way.

Limitations

Kurtz and associates and Recht and associates identified patients at risk for failure after breast-preserving techniques. There is ample reason to expect that recurrence after conservation surgery might be reduced to approximately 3 percent for patients who do not have an *extensive intraductal component (EIC)* and who are older than 35 years (Fig. [14-50](#)). Certain morphologic and histochemical features allow the selection of patients in whom breast conservation can be performed with increasing confidence: (1) mammographically detected lesions; (2) decreasing size of the primary invasive cancer; and (3) low S-phase component of DNA flow cytometry.

Despite the importance of breast preservation to enhance cosmesis and allow control and survival rates equivalent to radical procedures, certain women will desire total mastectomy for a variety of economic and psychosocial reasons. Patients less concerned about the cosmetic or functional importance of the breast may view total mastectomy as the most expeditious and desirable therapeutic option, especially with the economic morbidity and inconvenience of radiotherapy. Women who have EIC within and surrounding the index lesion should undergo total mastectomy because of the very high local failure rates in the remaining irradiated breast. This high local recurrence rate for EIC was observed despite extensive lumpectomy and evidence of pathologically negative margins with frozen-section control of the inked specimen.

For large lesions that occupy the central, subareolar portion of the breast, margins sufficient to achieve local control require sacrifice of large volumes of breast tissue, which may yield an unsatisfactory cosmetic deformity. These patients are best treated by total mastectomy and autogenous tissue reconstruction. Patients with two or more primary lesions in the same breast have a higher probability of local recurrence after local excision and irradiation. Patients with synchronous, multiple primary tumors have a higher likelihood of EIC. Very young patients (under 35 years of age) also have a higher incidence of recurrence in the presence of the EIC component after local excision and radiotherapy (see Fig. [14-50](#)).

Advanced Local Disease (Stage IIIA, IIIB, and Inflammatory Carcinoma)

Inflammatory breast carcinoma (see Fig. [14-30](#)) is an ominous clinicopathologic variant, with 5-year survival rates of 3 to 5 percent. Combination chemotherapeutic regimens (e.g., CAF—cyclophosphamide, doxorubicin, and 5-fluorouracil) have effected dramatic regressions of the breast lesion in approximately 60 to 75 percent of these patients. The index lesion and the associated breast changes related to subdermal lymphatic infiltration, as well as any axillary metastases, may disappear. With induction responses following two to six drug cycles of CAF, an extended simple mastectomy (inclusive of Level I nodes) is performed to remove residual malignant disease from the chest wall. Thereafter, peripheral lymphatics, skin flaps, and the central-apical axilla (Levels II and III) are treated with comprehensive radiation therapy. This multimodal approach may result in 5-year survival rates with inflammatory breast cancer that approximate 30 percent.

Polychemotherapy has been valuable in the therapy of stage III disease. In patients with large, matted axillary metastases and grave signs (edema, ulceration, peau d'orange, and skin or pectoralis major muscle fixation), extended simple mastectomy may not be technically possible. Preoperative chemotherapy allows a regression in size of the index lesion and of axillary metastasis. Sites of tumor infiltration and fixation to skin with ulceration also may undergo regression. Subsequently, an extended simple mastectomy might be feasible. Comprehensive irradiation of the chest wall defect and peripheral lymphatics is important to enhance local and regional control. These multimodal approaches may reduce chest wall recurrence to 4 to 10 percent; 5-year and 10-year survival rates are reported as 45 and 28 percent, respectively.

Breast Reconstruction

The use of immediate chest wall reconstruction at the time of mastectomy for invasive carcinoma has increased. In most instances, when radiation therapy or chemotherapy is planned, breast reconstruction can be delayed until these treatments have been completed. Increasingly, chest wall irradiation is being used in patients with implants, although planning and implementation of radiation physics are less precise. Optimally, reconstruction should be deferred until radiation therapy has been completed, because capsular scarring around the breast prosthesis may be stimulated by irradiation and various types of chemotherapy, particularly with doxorubicin. In 1992 the Food and Drug Administration allowed the placement of gel prostheses for reconstruction purposes only in the patient with neoplastic disease. The increasing use of autogenous tissue transfers for reconstruction does not appear deleterious to the radiation therapy.

Immediate breast reconstruction also can be considered in patients having total mastectomies for ductal carcinoma in situ or lobular carcinoma in situ, as well as individuals who, because of high risk, undergo prophylactic mastectomy. The use of autogenous tissues for reconstruction is highly desirable in these clinical settings.

The mastectomy is ideally performed via a transverse or oblique incision (see Fig. [14-34](#)). If the oblique incision is used, off-shoulder gowns may be worn without the medial portion of the scar visible. Laterally, however, the oblique scar should not extend into the apex of the axilla, as scarring may limit mobility of the arm and shoulder.

The major criticism of breast reconstruction has been the potential for delay in diagnosis of recurrent chest wall disease. For stage I breast cancer, chest wall recurrence as the first sign of failure is unusual. When proper local therapy of breast cancer has been completed, local chest wall recurrence rates of stage I and early stage II disease are 0 to 2 percent.

As the breast prosthesis is placed in the subpectoral position, superficial recurrence of the skin and subdermal connective tissues usually is palpable and is not obscured clinically or radiographically by the prosthetic implant.

Adjuvant Therapy

No solid tumor has been as extensively studied to determine the effects of systemic therapy as has carcinoma of the breast. Clinical trials indicate that *adjuvant cytotoxic therapy* and possibly *hormonal therapy*, when used in patients with axillary metastasis but without established distant metastasis, prolong the disease-free interval and may enhance survival rates. For patients with established distant metastasis (stage IV), therapy with several drugs that are less effective as single agents has resulted in response rates upwards of 50 percent when used in combination (polychemotherapy). The most prevalent and most studied polychemotherapy combinations are (1) cyclophosphamide, methotrexate, and 5-fluorouracil (CMF), and (2) cyclophosphamide, doxorubicin, and 5- fluorouracil (CAF). Prednisone and vincristine have sometimes been added to these regimens. Liberal use of vincristine, however, is not justified because of its high neurotoxicity.

Response rates for cytotoxic combinations vary from 20 to 70 percent. Complete responses (patients in whom all evidence of disease resolves) are rare with rates consistently below 20 percent. Complete responses with combination cytotoxic therapy are rare because heterogeneity of the breast cancer cell population often denies response by cells that are in a resting (G₀) phase of the cell cycle. With reentry into the cell cycle, neoplastic cell growth and mitosis are evident. Heterogeneous cell populations have variable response rates to the administered agents. This heterogeneity of cell population may explain why multiple drug combinations, with different sites of cytotoxic action within the cell, have a better overall response rate than is associated with single-agent therapy.

The toxicity noted with cytotoxic agents for breast cancer is similar to that observed with chemotherapy of other malignancies, and includes nausea, vomiting, myelosuppression, alopecia, thrombocytopenia, and exercise intolerance. These toxic events may be reversible with discontinuance of the cytotoxic agents. Doxorubicin initiates a profound and predictable cardiomyopathy with cumulative, dose-limiting side effects. Only 550 mg/m² may be given to a patient because cardiomyopathy, if it occurs, is irreversible.

Hormonal Receptors

Within the cytosol of breast cancer cells are specific proteins that bind and transfer steroid moieties into the cell nucleus to exert specific hormonal effects. The most widely studied and available receptor proteins are the estrogen receptor (ER) and progesterone (PR) receptor proteins. To obtain a quantitative hormonal assay of either hormone receptor, 1 g of fresh tissue obtained from the tumor is essential; the receptors are thermal and ischemia labile. Use of the electrocautery current near the fresh tumor at the time of excision should be avoided, as ER and PR activity will be invalidated. The cold scalpel should be used for obtaining tissue useful for valid analyses.

Specimens must be rapidly frozen (-70°C) for ER and PR assay, because decay in activity is evident within 20 to 30 min after extirpation of the neoplasm. Tissue cytosol is obtained by homogenization and centrifugation of the prepared specimen, which is then incubated with ³H-tritium-labeled estradiol-17 β . Labeled unbound hormone is removed from the incubation mixture, and the bound estrogen sediment is measured by multipoint titration with Scatchard plot analysis. Binding capacity is expressed in femtomoles (fmol) of ³H-estradiol-bound tissue per milligram cytosol protein.

Values of 10 fmol/mg or higher are considered *receptor-positive*; values below 3 to 4 fmol/mg are *receptor-negative*. Intermediate values are considered *borderline*. In all cases the laboratory analysis values should be reviewed to ensure appropriate interpretation.

The degree of positivity is proportional to the differentiation and histologic subtype of the lesion. Ninety percent or more of well-differentiated ductal and lobular carcinomas are ER-positive. Sequential studies of ER activity in the same patient usually reveal no significant difference between lesions of the primary and metastatic sites. There also appears to be no evolution or change of activity in metastatic sites from that of the index lesion.

Clinical response to various forms of endocrine manipulation is evident in patients who have ER activity. Less than 10 percent of ER-negative patients are responders; more than 60 percent of ER-positive patients respond to exogenous estrogens or endocrine ablative measures.

In the past, oophorectomy, adrenalectomy, and/or hypophysectomy were the primary *endocrine ablative procedures* used to treat metastatic foci. *Oophorectomy* was primarily used for premenopausal patients who presented with skin or bony metastasis with a prolonged disease-free interval that exceeded 18 months between treatment of the primary tumor and the discovery of metastases. Visceral metastases (e.g., lung, liver) were infrequently observed to respond to any form of hormonal manipulation. Pharmacologic doses of exogenous estrogens were provided for postmenopausal women who were observed to have an 18-month disease- free interval and metastasis primarily to bone or skin. Response rates for each of these subgroups approximated 30 percent. *Adrenalectomy* and *hypophysectomy* were effective in individuals who had previously responded to either oophorectomy or exogenous estrogen therapy. The response rates of these additional ablative techniques were also seen in one-third of the treated population.

Receptor activity is the most commonly used measure for determining the applicability and selection of additive hormonal or ablative endocrine procedures. There are correlations between tumor differentiation characteristics and reactivity of ER. In one review, only 8 percent of patients with ER-positive tumors were observed to have relapse. Ninety- one percent of patients with ER-positive tumors were free of disease at 24 months, compared to only 62 percent of ER-negative patients.

Other prognostic variables potentially account for differences in recurrence rates according to ER activity. Younger patients were observed to have trends toward positive nodes and greater need for adjuvant chemotherapy and more commonly had ER-negative tumors.

PR activity in the cytosol is also a measure of hormonal responsiveness of the index tumor or metastatic foci of disease. This receptor is measured concomitantly with ER from the primary tumor. Premenopausal patients have a lower incidence of ER-positive activity (30 percent) than postmenopausal patients (60 percent). These data suggest that premenopausal patients have a lower responsiveness to hormonal manipulation. However, response rates of premenopausal and postmenopausal patients are similar, and PR activity may be more indicative of an opportunity for hormonal manipulation in the premenopausal patient. Commonly, the premenopausal patient may have a tumor that is strongly PR-positive, yet may be ER-negative. This profile may indicate a high clinical correlation for response of the malignancy to hormonal manipulative therapy.

Premenopausal patients with undetectable ER have a threefold increase in PR as compared to postmenopausal groups. Since high endogenous estrogens in premenopausal patients may mask ER in tumor biopsies, it appears advantageous to perform PR determinations to identify an additional 15 percent of women with metastatic breast cancer who may benefit from endocrine therapy (Table 14-12). McGuire observed a correlation of the level of ER (fmol/mg cytosol protein) in breast neoplasms with the response rate to endocrine therapy. He observed that synthesis of PR is strictly estrogen-dependent and is the end product of estradiol- stimulated pathways in breast cancer tissues. An 80 percent objective response was observed in patients whose ER was 100 fmol/mg cytosol protein or more. A response rate of 46 percent was observed in women with lower values. This objective response rate to endocrine therapy as a function of content of ER has been confirmed by others. Studies also indicate a trend toward higher quantitative values of ER and PR in tumors that are histologically well differentiated. These correlations have been confirmed as high mean ER and PR values in tissues harboring low-grade (grade 1) neoplasms when both receptors are positive.

Antiestrogen Therapy

The antifertility drug tamoxifen was originally observed to have antiestrogen activity and initiate regression of breast cancer. Approximately one-third of patients initially treated with the drug showed objective regression of metastatic disease. Antitumor activity was closely correlated with the reactivity of ER and/or PR. Antiestrogens block the uptake of estrogen by the target tissue after cytosol binding to the ER. Diminished responsiveness at one dose level may be reversed by escalation of the dose.

The most striking advantage of tamoxifen over chemotherapy is the absence of toxicity and severe side effects. There may be, however, a “flare” of bone pain with induction of hypercalcemia when therapy is initiated; this effect usually is short-lived. In addition, nausea, vomiting, and fluid retention may be induced. An important long-term risk of tamoxifen use is the development of endometrial cancer. Thrombotic events occur rather infrequently, in 1 to 3 percent of treated patients.

The overview analysis of adjuvant systemic therapy trials for early breast cancer by the Early Breast Cancer Trialists' Collaborative Group showed that adjuvant therapy with tamoxifen produced highly significant reductions in the annual rates of recurrence and death (25 percent reduction in annual risk of recurrence and 17 percent reduction in annual mortality). The analysis also noted a reduction in the risk of development of contralateral breast cancer by 39 percent. Tamoxifen therapy usually is discontinued after 5 years.

Adrenalectomy

The initial application of adrenalectomy was in postmenopausal females after ablation of all other sources of estrogen, particularly in individuals who initially responded to exogenous estrogens. After menopause the adrenal glands are the major site for production of endogenous estrogens. Aminoglutethimide blocks enzymatic conversion of cholesterol to g-5- pregnenolone and inhibits the conversion of androstenedione to estrogen in peripheral tissues. After treatment with this agent, adrenal suppression is evident, with a reduction of cortisol secretion and feedback increase in ACTH that may override the aminoglutethimide blockade. Thereafter, glucocorticoid therapy is required for suppression of ACTH secretion by the adrenal cortex. This therapy amounts to a “medical adrenalectomy” and has been compared prospectively with surgical adrenalectomy and hypophysectomy. Medical therapy is equivalent to surgical ablation and represents an alternative. Neither permanent adrenal insufficiency nor acute crises were observed. Side effects included ataxia, dizziness, and lethargy; these effects were dose-dependent and transient.

Applications

Fisher and associates and Bonadonna and associates suggested that disease-free and overall survival may be improved when additive therapy is initiated before clinically detectable distant disease. The goal of therapy is eradication of well-established but as yet unidentified micrometastases. Original recommendations were for adjuvant chemotherapy using multiple combinations in premenopausal women with three or more positive axillary lymph nodes. Data suggest that this approach also may be applicable for women with one to three positive nodes regardless of menopausal status. The addition of the antiestrogen agent tamoxifen to the chemotherapeutic regimen for ER-positive patients appears to provide even greater protection against the development of distant disease.

An improvement in survival of at least 5 to 20 percent over that anticipated for untreated patients with either stage II or stage III disease is being realized in most clinical trials.

The overview analysis of adjuvant systemic therapy trials for early breast cancer by the Early Breast Cancer Trialists' Collaborative Group showed that combination chemotherapy produced a 28 percent reduction in the annual rate of recurrence and a 16 percent reduction in mortality. These reductions in the risk of recurrence and mortality were observed in both node-positive and node-negative patients. The risk reductions produced by chemotherapy were highly significant in patients of all ages, but in younger patients there was a significantly higher efficacy (36 percent annual reduction in risk of recurrent and 24 percent reduction in annual mortality in women younger than 50 years of age compared to 24 percent and 13 percent, respectively, for women older than 50 years of age). Although long-term (12 months) polychemotherapy was not shown to be better than short-term (6 month) treatment by both direct and indirect randomized comparisons, polychemotherapy was found to be significantly better than single-agent chemotherapy. No significant

differences were found between different polychemotherapy regimens or between different tamoxifen doses. However, long-term tamoxifen treatment was shown to be significantly more effective than treatment of shorter duration.

Node-Negative Breast Cancer

Data from the NIH Consensus Conference in 1991 indicate that (1) the majority of patients with node-negative cancers are cured by breast conservation treatment or total mastectomy and axillary dissection; and (2) the prevailing evidence suggests that the rate of local and distant recurrence is decreased by *both* adjuvant combination cytotoxic chemotherapy and adjuvant tamoxifen. Ten randomized trials confirm that adjuvant systemic therapy reduces the observed rate of recurrence by approximately one-third.

At present, prospective randomized trials are too immature and often lack sufficient numbers to estimate with acceptable precision the correlation between menopausal status or steroid reactivity and the effects of adjuvant therapy in these node-negative patients. Few patients with ER- negative tumors have been included in tamoxifen studies. Reduced mortality, however, is seen in nearly all trials, although this has not reached statistical significance in most trials. The rate of death in patients with node- negative disease is low; thus, a clinically important reduction in mortality will require prolonged follow-up with large numbers of patients to achieve statistical significance.

The major benefits of chemotherapy are seen when the antimetabolites (5- fluorouracil and methotrexate) are administered intravenously, rather than given orally. Trials with tamoxifen suggest that the use of this drug for more than 2 years (usually 5 years) results in greater reduction in the risk of recurrence.

In all prospective-study patients who are node negative, tamoxifen reduces the clinical incidence of contralateral primary breast cancer. Overall benefits from tamoxifen in postmenopausal patients clearly outweigh the toxic side effects encountered. For premenopausal patients, the administration of tamoxifen may initiate endocrine abnormalities that have not been defined by long-term analysis (e.g., endometrial carcinoma, effect on the developing fetus). There are no data on the effects of combination cytotoxic agents with tamoxifen for treatment of node-negative breast cancer patients.

The NIH Consensus Conference recommended that patients who are not candidates for prospective trials or who are noncompliant for participation in these trials should be made aware of the benefits and risks of adjuvant systemic therapy. Adjuvant therapy should consist of either combination chemotherapy in these patients or tamoxifen (20 mg/day for at least 2 years). Comprehensive prospective studies have not directly compared tamoxifen and chemotherapy (with or without tamoxifen) in the node- negative subset.

Prognostic Factors

The majority of node-negative patients should be cured after local-regional therapy. Factors considered important to the risk for recurrence include tumor size, ER and PR status, tumor grade, histologic type, proliferative rate (thymidine labeling indices, S-phase fraction), and other incompletely defined risk parameters (protease, cathepsin, HER-2/neu, EGF receptors, stress-response proteins). Despite the proliferation of definable risk factors and biologic markers, tumor size and the axillary nodal status remain the most important of these risk variables (Table [14-13](#)).

The incorporation of such prognostic parameters into equations that mathematically estimate risk proportion are gaining attention and may have theoretical and clinical investigational value. With the exception of clinical trials, it is not reasonable to treat patients with tumors 1 cm or less in diameter, because the chance of recurrence in 10 years is less than 10 percent. With increasing tumor diameter with or without positive nodes, other prognostic variables must be considered in deciding whether to use adjuvant therapy.

Systemic Therapy Trials for Early Breast Cancer

Surgery for early breast cancer allows all clinically apparent (macroscopic) disease to be removed. Various forms of systemic “adjuvant” therapy should be considered postoperatively, namely, antiestrogen therapy (tamoxifen) and those that involve a cytotoxic agent, or a combination of the two drugs. Analysis of collected data by the Early Breast Cancer Trialists' Collaborative Group confirmed significant reductions in annual rates of recurrence and of death with tamoxifen-treated patients, in patients who had ovarian ablation before the age of fifty, and in patients treated with chemotherapy, but no improvement in survival or reduction in recurrence rates for patients treated by immunotherapy.

Tamoxifen also was shown to reduce the risk for developing contralateral breast cancer by 39 percent. The salient benefits of reduction in *recurrence* by both polychemotherapy and tamoxifen were seen chiefly during years 0 to 4 postoperatively; the reduction of *mortality* was highly significant both during and after years 0 to 4 in these relatively brief treatments. The Cancer Trial Collaborative Group provided little information beyond year 10, except for ovarian ablation, which produced significant mortality reductions during and after years 0 to 9 postoperatively. These collaborators verified that long-term (12 months) polychemotherapy was no better than shorter (6 months) regimens. Polychemotherapy provided reduction of recurrences and overall survival that was superior to single-agent chemotherapy. There was no significant difference between various forms of polychemotherapy or different tamoxifen doses, but long-term tamoxifen (e.g., 2 years, or even 5 years) was significantly more effective than short-term regimens.

Tamoxifen was effective in older patients (over 70 years of age); chemotherapy has not been properly evaluated in this geriatric group. In patients aged 50 to 69 years direct comparisons confirm that chemotherapy *plus* tamoxifen is superior to chemotherapy alone or tamoxifen alone, for both recurrence and survival. In younger women (under 50 years), chemotherapy and ovarian ablation have comparable effects. Trends suggest that the combination of chemotherapy and ovarian ablation might be superior to either alone.

A 30 to 40 percent proportional reduction in risk was achieved with combination chemoendocrine therapy for middle-aged patients. This risk reduction was evident for node-positive and for node-negative patients. The improvement in absolute 10-year survival was *twice* as great for the node-positive patients as for the node-negative groups (see Table [14-13](#)). An algorithm of treatment options for adjuvant systemic therapy is shown in Fig. [14-51](#).

Metastatic Disease

Polychemotherapy was designed for therapy of metastasis to bone, liver, soft tissue, lung, and occasionally brain. The combinations of cytotoxic agents most commonly used were 5-fluorouracil, cyclophosphamide, methotrexate, vincristine, and doxorubicin. Randomized trials comparing cytotoxic agents indicated an increase in response rates from approximately 25 percent with single agents to 50 to 60 percent with combination therapy. Further, median survival for combination therapy in those individuals who do not respond is longer than that obtained with single-agent therapy. While relationships between responsiveness to chemotherapy and the ER-PR status have been controversial, more recent studies confirm that hormonal receptor data properly guide anticipated response rates. The heterogeneity of the cellular population within the breast cancer primary and metastatic sites probably determines the therapeutic response cytokinetics and cell kill relative to the cell cycle. No reproducible parameters to evaluate heterogeneity of the cell population are as yet available. The majority of trials identify significant prolongation of chemotherapy-induced responses for patients who previously were responders to hormone therapy and who have ER-positive tumors.

Age, menstrual status, family history, number of positive nodes, tumor size, and type of surgery have been poor predictors of response to chemotherapy. A prolonged disease-free interval usually does not indicate a response advantage. The median duration of response to combination chemotherapy ranges from 12 to 18 months. Once failure has occurred, the use of an additional combination of agents may initiate a remission (partial or complete). Stage IV patients who ultimately achieve complete remissions have a median survival of 32 months.

Patients who experience relapses from combination therapy remain eligible for hormonal manipulation, particularly if their tumors are hormone- receptor positive and they have not previously been treated by hormonal manipulation. Patients who have rapidly progressive disease and a short disease-free interval should be treated initially with combination cytotoxic agents before therapy with hormonal manipulation. As tamoxifen has shown therapeutic advantage when combined with chemotherapy in the age group 50 to 69 years, it may prolong survival in advanced disease as well.

A relationship exists between *response* to chemotherapy and *dose* of the cytotoxic agent used, but there is dose-limiting toxicity. For most chemotherapeutic agents this toxicity factor is myelosuppression. Recent trials using transplantation of normal autologous bone marrow cells have allowed dose-intensive chemotherapy regimens to overcome this limitation.

In the case of advanced local disease (stage III), the cytotoxic dose may be escalated as great as tenfold, depending on the toxicities encountered. Anteman and Gale reviewed 27 trials of high-dose therapy in which 172 patients received single-agent or multiagent therapy, radiation, or both. Many had been treated extensively with multiple agents before receiving high-dose intensive therapy after marrow transplantation. The multiple alkylating agents offer the best response rate (76 percent). Although response rates are high, duration of response may be short and is often less than 6 months.

In a trial with high-dose cyclophosphamide, cisplatin, carmustine, and melphalan along with bone marrow support as initial therapy, a *complete response* was evident in 55 percent with an *overall response* rate of 73 percent. The median duration of response for patients who had complete remission was 9 months. These studies suggest that a large cell kill can be obtained with single intensive therapy using autologous marrow rescue. However, this therapy appears to be inadequate to eliminate enough cell populations in order to provide durable remission responses. The use of hematopoietic growth factors (GM-CSF—granulocyte-macrophage-colony- stimulating factor) may allow applications of multiple-dose intensive regimens to obtain an enhanced cell kill with autologous marrow rescue. The future of successful therapeutic interventions for stage IV disease depends on innovative approaches that provide high cell kill with reduced myelosuppression and systemic toxicities.

New agents that produce high response rates have been evaluated in the treatment of metastatic breast cancer. They include spindle poisons such as paclitaxel, docetaxel, and vinorelbine, topoisomerase inhibitors, and new antifolates. New hormonal treatments include the use of new aromatase inhibitors, GnRH agonists, antiprogestins, and new antiestrogens. Finally, biologic therapies targeting growth factors and growth-factor receptors such as HER-2/neu and epidermal growth factor receptors (EGFR) have been developed and trials using these approaches are under way.

Carcinoma of the Male Breast

Less than 1 percent of all breast cancers occur in men. The incidence appears to be highest among North Americans and the British, in whom it constitutes 0.4 to 1.5 percent of all male cancers. Jewish and black males have a higher incidence. Gynecomastia precedes approximately one-fifth of these malignancies. Male breast cancer has been associated with Klinefelter's syndrome (XXY), estrogen therapy, high endogenous estrogen levels related to testicular feminizing syndromes, irradiation, and trauma.

This tumor is rarely seen in young males; the incidence peaks between 60 and 69 years of age. Hormonal dependence of the neoplasm is typical, and the tumor is commonly estrogen-receptor positive. Clinical presentations of the disease are similar to those for women except that the diagnosis is delayed owing to infrequent recognition in the male patient.

Clinical characteristics include breast mass, nipple retraction, discharge, skin fixation, ulceration, and pain. Stage for stage, men with this neoplasm appear to have the same survival rate as women, but the overall prognosis is poor because of the advanced stage of disease (stages III, IV) at the time of diagnosis. Overall survival rates for node-negative

patients correspond favorably with those for women; survival in node-positive males is poor, suggesting the need for adjuvant therapy in this group.

The preferred treatment is modified radical mastectomy and use of postoperative radiotherapy for ulcerative and/or high-grade anaplastic tumors to reduce local recurrence. Orchiectomy and administration of estrogenic steroids may induce remissions of metastatic disease. Hormonal manipulation, by medication or by ablation, often provides objective responses in the management of metastatic disease. Cytotoxic chemotherapeutic agents have been used infrequently in the therapy of male cancer. Early trials suggest response rates similar to those for female breast cancer patients.

Breast Cancer During Pregnancy and Lactation

From the time of Billroth until Halsted's pioneering work, carcinoma of the breast diagnosed during pregnancy was considered incurable. White and White in 1956 reported an incidence of approximately 2.8 percent for cases occurring during pregnancy. Recent reviews suggest an incidence of 3 cancers per 10,000 pregnancies, or a range from 0.4 to 3.8 percent of reported breast cancers. The average age of the pregnant patient with concomitant breast cancer is 34 years.

In the past, the association of cancer with pregnancy was considered ominous. The profound estrogen and progesterone stimulation of breast cancer cells from the placenta and corpus luteum reportedly increases the risk of distant disease with provision of an excellent hormonal milieu to support cellular growth of the neoplasm. More recent data indicate that, stage for stage, carcinoma of the breast in pregnancy is associated with a prognosis similar to that of the nonpregnant female. There are, however, proportionally more patients with stage II and stage III breast cancers diagnosed in pregnancy than in the general population. Whether this late stage at diagnosis is secondary to the hormonal stimulation of pregnancy or to delay because of the coincident physiologic changes expected with pregnancy has not been determined.

Diagnosis

A careful breast examination should be performed at the initial obstetric visit and, at least, in each trimester of the pregnancy. Diagnostic work-up in pregnancy and lactation is similar to that in the nonpregnant patient, although breast mammography tends to be less reliable because of the extensive parenchymal changes associated with gestation. Radiation exposure of the fetus should be negligible with modern techniques for proper shielding of the abdomen. Regardless of the mammographic findings, any dominant mass should be evaluated promptly in the pregnant or lactating patient. Evaluation may begin with fine-needle aspiration to distinguish cysts from solid lesions. Any solid, discrete mass requires biopsy, which usually can be completed under local anesthesia. Risk of spontaneous abortion during mastectomy is approximately 1 percent and is correlated with stage of gestation.

Treatment

Therapy is identical to that for the nonpregnant patient. Patients who have advanced cancer require appropriate chemotherapy and radiotherapy. Modified radical mastectomy is optimal therapy for patients with stage I or stage II disease. When segmental mastectomy is selected, it must be followed by whole-breast irradiation, after delivery if the diagnosis is established late in the third trimester. In patients with positive axillary lymph nodes, chemotherapy should be delayed until the second trimester of pregnancy to diminish fetal toxicity and spontaneous abortion.

Termination of pregnancy has no role in the management of stage I or stage II disease. There is no evidence that oophorectomy influences the course of breast cancer during pregnancy and lactation. With mastectomy, normal pregnancy is allowed to continue, with minimal risk to the mother or the fetus. Modern series do not provide evidence that abortion benefits control or survival rates for patients with the disease. For stage I and stage II disease some patients with early disease will elect breast conservation. Segmental mastectomy should be strongly discouraged except in the third trimester, when radiotherapy could be reasonably delayed (4 to 6 weeks) until delivery.

Risks associated with chemotherapy during pregnancy have not been established. Schapira and Chudley found the incidence for teratogenicity of chemotherapeutics given to humans in the first trimester to be 12.7 percent, whereas Sweet and Kinzie noted an 11.5 percent incidence. There was no evidence of teratogenicity from administration of chemotherapeutic agents in the second and third trimesters. Little data exist with regard to combination regimens, as most of the reports considered single-agent usage. Long-term consequences of fetal exposure to chemotherapeutic agents are unknown.

In general, the use of cytotoxic agents during the first trimester should be discouraged; use during the second and third trimesters probably induces very few fetal abnormalities. For patients with positive lymph nodes at mastectomy, chemotherapy may be delayed until the second trimester. In the third trimester, after determination of fetal age and maturity, consideration should be given to early cesarean section to minimize delay in initiating cytotoxic therapy for stage II and stage III disease.

Lactation should be suppressed promptly in the postpartum state, even if the biopsy identifies a benign lesion, as milk from transected lactiferous ducts will drain via the biopsy site. If the infant is breast feeding, rapid weaning is desirable.

During postoperative radiotherapy or chemotherapy to control an aggressive primary lesion (stage III, inflammatory cancer), the patient who becomes pregnant should consider therapeutic abortion. In the first trimester, both radiotherapy and chemotherapy are potentially teratogenic, particularly in the first 9 weeks of gestation. Radiation in doses greater than 500 cGy, by itself, administered in early fetal development may be teratogenic.

Recurrence During Pregnancy

Data do not convincingly establish that survival is diminished nor that recurrence is enhanced for women who become pregnant after treatment of a breast cancer. In contrast, studies suggest an improved survival among breast cancer patients who later become pregnant. These data document no detrimental effect of subsequent pregnancy even among patients with positive nodes or of pregnancies that occur less than 2 years after mastectomy. Abortion does not appear to provide an improvement in the survival rate.

While no therapeutic grounds exist for recommending avoidance or termination of pregnancy among patients without documented recurrence, theoretically the disease-free interval may be shortened, particularly in ER- positive patients in whom the enhanced estrogen milieu will support growth of tumor cells. Thus, child bearing or estrogen-containing compounds must be considered cautiously before either is recommended to the patient at high risk for recurrent disease (e.g., stage II, III, inflammatory cancer). For patients at high risk for recurrent disease and in whom estrogen antagonists may be useful as an antineoplastic agent, the use of oral contraceptives and estrogen-containing compounds should be avoided.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 15 - Tumors of the Head and Neck

John J. Coleman III
Mark R. Sultan

CONGENITAL LESIONS

Thyroglossal Duct Cysts

The thyroid gland originates from the pharyngeal floor at the foramen cecum during the fourth week of gestation. It enlarges, becomes bilobed, and descends ventrally in the midline of the neck in close approximation to the developing hyoid bone. During this descent the patent diverticulum is called the thyroglossal duct. The duct normally resorbs by the tenth week of gestation. When all or a portion of this duct persists, thyroglossal duct cysts or sinuses are formed.

Classically these cysts present as midline masses in childhood, although they have been reported to be as much as 2 cm from the midline and may present for the first time in adults. Eighty percent occur at or just below the hyoid bone. A maneuver to differentiate them from Delphian lymph nodes or other central masses is to have the patient protrude the tongue. The level of the cyst is elevated by protrusion, demonstrating its embryologic origin from the base of the tongue. Unlike branchial cleft remnants, thyroglossal duct cysts generally do not have external sinuses. A significant proportion do become infected, however, usually during the course of an upper respiratory tract infection. Approximately 5 percent of the cysts contain functional thyroid tissue, and rare cases of thyroglossal duct carcinoma have been reported.

The differential diagnosis for midline neck masses about the hyoid bone includes lingual thyroid tissue. In rare cases this is the patient's only active thyroid gland. Therefore, the presence of thyroid tissue in its normal anatomic location must be confirmed either clinically or by radioactive scan before any midline neck mass is excised, and careful postoperative observation for hypothyroidism is imperative.

Sistrunk is credited with the development of a technique for excising thyroglossal duct cysts and sinuses that minimizes the risk of recurrence. He described resection of the cyst with the central portion of the hyoid bone, following the sinus superiorly to its presumed site of origin, the foramen cecum, and excising it in its entirety.

Branchial Cleft Anomalies

Branchial cleft cysts, sinuses, and cartilaginous remnants result from the incomplete fusion of the branchial clefts. The branchial clefts, appearing in week 4 of embryonic life and normally involuting fully by week 7, contribute to the formation of various head and neck structures in the developing embryo. When a portion of a cleft persists, epithelium-lined cysts or sinuses, with or without cutaneous openings and cartilaginous rests, may result. These anomalies usually present in the first decade of life but may go undetected until adulthood. The majority of branchial cleft cysts and sinuses are lined by squamous epithelium, although ciliated columnar epithelium has been reported as well. Branchial cleft carcinoma occurs rarely when there is a history of a branchial cleft cyst, the subsequent development of epidermoid cancer at that site, and no other primary lesion. Branchial cleft cysts also contain lymphoid tissue and may enlarge in response to upper respiratory tract infections.

The most common types of branchial cleft anomalies are those of the second cleft and are found at the middle and lower thirds of the sternocleidomastoid muscle. Cartilaginous rests from the first branchial cleft typically are subcutaneous, usually appear medial to the tragus, and may be managed by simple excision. Cysts and sinuses often extend more deeply into the neck. The fistulous tract courses superiorly along the carotid sheath and then medially over the hypoglossal nerve between the internal and external carotid arteries to end at the pharynx adjacent to the tonsillar fossa. A stairstep incision is sometimes needed to follow this circuitous route. First branchial cleft cysts and sinuses are located above the level of the hyoid bone just below the body of the mandible and extend superolaterally through the parotid gland to end within the membranous external auditory canal. Excision of these cysts and sinuses is recommended to avoid the complications associated with recurrent infection. Dissection must be meticulous to avoid injury to the facial, hypoglossal, vagus, and lingual nerves and to the carotid vessels. Anomalies of the third branchial cleft are rare. Like second cleft anomalies, they arise anterior to the middle and lower thirds of the sternocleidomastoid muscle. However, they course behind the carotid artery to end at the pyriform sinus.

Hemangiomas and Vascular Malformations

Congenital vascular lesions must be clearly classified as hemangiomas or vascular malformations in order to assess their prognosis and establish appropriate management plans. The distinctions between the two have been clarified on the basis of cellular and clinical characteristics. Hemangiomas have an increased mitotic activity and as such may be considered true neoplasms. They are typically absent at birth or may be present as a faint vascular blush. During the first several months of life they may undergo a rapid proliferative phase, during which they sometimes grow to large size. Although the majority of hemangiomas undergo spontaneous involution by the age of seven, complications such as ulceration and bleeding, obstruction of the eye with subsequent amblyopia, nasal airway obstruction, and rarely thrombocytopenia (Kasabach-Merritt syndrome) may mandate early surgical resection. Systemic dexamethasone therapy, or the intravenous administration of interferon- α , for a short course has been found to arrest the growth of large lesions during their proliferative phase. The use of intralesional steroid injections or sclerosing agents may be successful in achieving temporary control of smaller hemangiomas in certain locations such as the lip or eyelid. Photodynamic laser therapy may be helpful in preventing the onset of the proliferative phase of hemangiomas.

Vascular malformations, unlike hemangiomas, have a normal rate of endothelial cell turnover. They result from congenital errors in vascular morphogenesis and are classified by their vessel of involvement—capillary, venous, arterial, lymphatic, or combined. High-flow lesions result from gross abnormalities connecting the arterial and venous systems and may cause catastrophic problems of massive hemorrhage, high-output congestive heart failure, and hemolytic anemia. Chronic increased blood flow may be associated with skeletal abnormalities such as bone hypertrophy and distortion. Capillary, venous, and arterial malformations may occur anywhere in the head and neck. Lymphatic malformations (cystic hygromas) classically occur in the neck or the floor of the mouth. All malformations are present at birth, although they may not be clinically evident until the ectatic vessels suddenly dilate under hormonal or other physiologic influences. They normally grow proportionally with the child and do not regress spontaneously, unlike hemangiomas. Therefore, the management of malformations often is surgical. Indications for early surgical resection include recurrent infections, obstructive symptoms (e.g., respiratory distress), hemorrhage, and significant aesthetic deformities.

Because vascular malformations of the head and neck often are highly infiltrative, a complete preoperative evaluation of the extent of the lesion and its vascularity must be obtained. Angiography usually is necessary, sometimes for diagnosis, but usually to determine the contributory vessels. When technically possible, complete extirpation of the malformation should be performed while preserving normal anatomic structures. Preoperative embolization by arteriographic technique may decrease blood flow temporarily, making surgical resection safer and more practical.

Vascular birthmarks, in the past descriptively named “strawberry” or “capillary” hemangioma, “port-wine stains,” “cavernous hemangiomas,” and “cystic hygromas,” are now classified by their major vessel of involvement—capillary, venous, or lymphatic malformation. Accurate description is crucial, since the prognosis and management of vascular birthmarks depend directly on their correct classification as hemangiomas or malformations.

In deep-seated subcutaneous vascular malformations such as those sometimes occurring in the cheek or pharynx, physical examination may be inadequate for accurate delineation. Further evaluation often is necessary and may include computed tomography (CT) or magnetic resonance imaging (MRI), angiography, or technetium-labeled red blood cell scintigraphy. Definitive diagnosis in any patient with a vascular mass may require open biopsy to rule out a malignancy, which may occasionally have similar radiographic characteristics.

BENIGN LESIONS

Lip

The lower lip may be subject to chronic irritants such as pipe smoking or lip biting and, more important, to the damaging effects of chronic actinic exposure. In affected individuals the epidermis of the vermilion becomes atrophic and the dermis reveals elastosis. The basal layer of the epidermis then develops dysplasia, creating thickening of the superficial mucosal layer, the stratum corneum. This thickening or hyperkeratosis is clinically visible and palpable. With progression, the dysplasia extends upward into the epithelium, producing parakeratosis, the accumulation of nucleated cells, near the epithelial surface. This is manifested clinically by scaling of the lip. A proliferation and abnormal orientation of epithelial cells, *dyskeratosis*, may then follow, ultimately leading to *carcinoma in situ*. With penetration of the basement membrane, an invasive squamous cell carcinoma develops. When dyskeratosis or carcinoma in situ is present over a large extent of the surface of the lower lip, excision of the entire vermilion (“lip shave”) should be considered. The lip is then resurfaced by advancing the labial and buccal mucosa to the mucocutaneous junction.

Oral Cavity

A variety of benign lesions arise in the oral cavity. They may be grouped by location into those affecting the buccal mucosa, gingiva, or tongue or by their causation, such as inflammatory or ulcerative. For many, their significance lies in their possible premalignant potential or in their mimicking of true malignancies.

The oral lining contains countless mucous glands, the minor salivary glands. A submucosal accumulation of mucus results in a mucous retention cyst. Since the majority of these cysts have no epithelial lining, it is believed that they are most often caused by rupture of the duct system, with extravasation of mucus, and not by ductal obstruction, as was previously thought. The most common location for a mucous cyst or mucocele is the labial mucosa of the lower lip (Fig. 15-1). These lesions are typically less than 1 cm in diameter, smooth, rounded, and have a bluish hue. The treatment of choice is excision. Marsupialization alone is reserved for extensive lesions. A *ranula* is a type of mucous retention cyst that arises from major salivary glands, most commonly the sublingual glands. This too is managed by excision, but given its location in the floor of the mouth and, in some cases, its large size, it may require a meticulous dissection to avoid nerve injury and postoperative hemorrhage. It is frequently necessary to resect the affected sublingual gland in-continuity to prevent recurrence.

An *epulis* is a granulomatous lesion of the gingiva; as elsewhere in the body, it represents an exaggerated inflammatory response to minor injury. Two subtypes exist: *congenital epulis* is typically found in the anterior maxilla of newborns; *epulis gravidarum* occurs in approximately 1 percent of pregnant women and normally resolves spontaneously when the pregnancy is concluded. Only symptomatic epulides need be excised.

Peripheral giant cell reparative granulomas also occur most commonly on the gingiva. The “giant” cell of origin resembles an osteoclast. These granulomas are polypoid, submucosal, and fibrous. They can create ulceration and hemorrhage of the overlying mucosa. Radiographs may reveal erosion of the underlying bone. Excision must be complete to prevent recurrence. The term “peripheral” refers to the soft-tissue origin of these tumors as opposed to the *central giant cell reparative granulomas*, which, although similar histologically, arise within bone. The peripheral giant cell reparative granulomas are four times more common than the central type. The latter are expansile endosteal lesions that typically present in the mandible of young adults. They also have been reported in the paranasal sinuses, orbit, cranial vault, and temporal bone. They must be distinguished from true giant cell tumors of bone (which have malignant potential), brown tumors of hyperparathyroidism, traumatic bone cysts, and fibrous dysplasia. Thorough curettage is generally

curative.

The tongue and the larynx are common locations for the development of *papillomas*. They are caused by the human papillomavirus, which induces squamous epithelial proliferation into soft, irregular, pedunculated lesions that may recur and cause obstruction of the airway. Eradication may be accomplished by excision or cauterization using a carbon dioxide laser. Other common benign masses of the tongue include fibromas, neurofibromas, and lingual thyroid nodules. The latter may lie dormant through childhood and then rapidly enlarge during puberty.

In 1926 Abrikossoff described rare benign tumors involving the tongue, which he named *granular cell myoblastoma* because of their presumed embryonal muscle cell of origin. They are now believed to derive from Schwann cells and have been found to arise throughout the aerodigestive tract, particularly in the larynx. In the tongue they typically occur as firm, submucosal swellings in the middle one-third and may therefore mimic squamous cell carcinoma. Wedge excision is curative.

The oral lining is subject to a number of ulcerative conditions. The *idiopathic aphthous ulcer* is the most common type. The cycle of painful ulceration and spontaneous healing may occur several times a year and persist for many years. Other similar ulcers have identifiable causative factors, including viral infection with herpes simplex, nutritional deficiencies including vitamin B, folate, or iron, and emotional stress. These ulcers often respond to topical steroids. Multiple painful oral ulcerations may be a manifestation of *pemphigus vulgaris*. This may be accompanied by severe, generalized toxicity. The disease typically occurs in the fifth to seventh decade of life in patients of Mediterranean descent. The ulcers begin as intraepithelial bullae that subsequently rupture and ulcerate. The overlying epithelium may be rubbed off easily (Nikolsky's sign). In *erythema multiforme*, persistent or recurrent painful oral ulcerations arise within a background of diffuse oral erythema. Biopsy examination reveals a perivascular lymphocytic cellular infiltrate. Separate skin involvement may or may not be present. A variety of causes have been proposed. Management with systemic steroids and antimetabolites is often necessary.

Discrete, painful ulcers are also present in *necrotizing sialometaplasia*. This is a benign inflammatory disease of minor salivary glands that usually occurs on the hard palate. It is believed that local trauma leads to progressive local ischemia and ulceration. The histologic differentiation from squamous cell cancer or mucoepidermoid carcinoma can be extremely difficult by incisional biopsy. Therefore, although these ulcers will ordinarily heal spontaneously in 6 to 10 weeks without specific treatment, in cases where the correct diagnosis remains unclear after incisional biopsy, complete excision of the ulcer is prudent.

The description and significance of leukoplakia and erythroplasia will be reviewed later in this chapter. Other white plaquelike lesions of the oral cavity include *white sponge nevus*, *lichen planus*, and *oral hairy leukoplakia*. Histologically, all reveal parakeratosis. White sponge nevus is a rare familial ectodermal disease that diffusely involves the oral cavity in a benign, self-limited manner. Lichen planus is a degenerative mucocutaneous disease with a probable autoimmune basis. Oral lesions may appear with or without cutaneous manifestations and may at times become erosive. Oral squamous cell carcinoma has been found in association with lichen planus with an incidence varying between 0.09 and 10 percent in different series. Therefore, as in oral leukoplakia, systemic and topical retinoids are being evaluated in the treatment of oral lichen planus to reverse the condition itself and, more important, to suppress the presumed heightened potential of the oral mucosa toward degeneration into invasive carcinomas in these conditions. Oral hairy leukoplakia is a form of parakeratosis recently described in patients with AIDS or other forms of immunosuppression such as in patients with renal transplants or leukemia. Thick, shaggy plaques typically appear on the lateral surface of the tongue and become symptomatic when superinfected with *Candida*. Management includes antifungal medication, antiviral agents, or surgical excision.

Nose

Polyps arise commonly within the nasal cavity and paranasal sinuses. They occur with equal frequency in males and females and in all age groups after adolescence. Ten percent of children with cystic fibrosis may also develop nasal polyps. The polyps are often multiple, involving both sides of the nasal cavity and the paranasal sinuses. They may present with nasal obstruction, mucoid nasal discharge, or anosmia. Those that arise in the region of the turbinates and ethmoids are mainly allergic in origin, while those of the posterior nasal cavity are most often infectious. Medical management should include an evaluation for allergies. Aspirin should be stopped or avoided, since an association exists between aspirin use and the formation of nasal polyps. If the allergic evaluation is negative, an empiric trial of antimicrobials is begun. Steroid nasal sprays may be helpful. When medical management fails, surgical intervention may be necessary. Simple polypectomy carries a high rate of recurrence, and more extensive endoscopic or open excision may be necessary.

Papillomatous growths may occur on the nasal skin or within the nasal cavity. The *squamous papilloma* is an exophytic verrucous lesion caused by the papillomavirus and is present on the skin of the nasal sills, columella, or alae. It is managed as are other cutaneous warts. The *inverted papilloma* (also called schneiderian papilloma, squamous papilloma, or papillomatosis) is a polypoid mass occurring on the lateral nasal wall, typically in middle-aged men. The name is derived from its histologic appearance of an “inverted” proliferative growth pattern. The significance of this lesion lies in its presentation with symptoms of nasal obstruction, its recurrence rate of 50 percent with polypectomy alone, and its association with concurrent (8 percent) and subsequent (4 percent) invasive squamous cell carcinoma. More extensive resection and close follow-up are therefore indicated and can lower the recurrence rate to 6 percent.

Juvenile nasopharyngeal angiofibromas are benign but highly expansible and destructive fibrovascular neoplasms that typically arise in adolescent males between 10 and 20 years of age. Originating in the superior nasal cavity, they can erode widely into the paranasal sinuses, orbit, pterygomaxillary fossa, and middle cranial fossa. Early symptoms include nasal obstruction and epistaxis, and more advanced lesions can produce anosmia, proptosis, or cranial nerve dysfunction. Management commonly requires preliminary angiographic embolization followed by surgical extirpation. Approximately 10 percent require a combined intracranial/extracranial approach. Radiation therapy is generally reserved for residual or recurrent disease, although its successful use as a primary modality has also been reported.

Paranasal Sinuses

Although the terms are often used interchangeably, mucous retention cysts and mucocèles of the paranasal sinuses are different entities with distinct pathogeneses, natural histories, and connotations. *Mucous retention cysts* arise as a result of blockage of secretion from *microscopic* secretory ducts of mucous glands within the lining of the paranasal sinus cavity, possibly as a sequela of sinusitis. The fluid mass that results remains separate from the bony wall of the sinus and so continues to be surrounded by air within the sinus, except at its base. Rarely, they can enlarge to occupy the entire sinus, at which point it would be difficult to distinguish them from the more virulent mucocèles. Radiographically they appear as discrete masses that are profiled by sinus air. The most common location for mucous retention cysts of the paranasal sinuses is in the maxillary sinus, where they usually present as an asymptomatic incidental finding on x-ray. Ten percent of routine sinus radiographs reveal evidence of these cysts in the maxillary sinus. They are considered the most common benign lesions of the maxilla. Treatment is rarely necessary.

As opposed to the indolent mucous cyst, *mucocèles* of the paranasal sinuses, although benign, can be expansile, highly destructive lesions. These result from *macroscopic* blockage of a sinus ostium by epithelial or osseous neoplasms or inflammatory processes, or as a result of trauma (e.g., facial bone fractures). Mucinous secretions then accumulate within the entire sinus. With persistent secretion a pressure effect on the entire sinus wall is produced, displacing both the epithelial lining and the bony wall. This ultimately can result in thinning and destruction of the wall such that the mucocèle can “invade” adjacent vital structures. CT or MRI scanning is needed to delineate the complete extent of the process, but clinical distinction from a carcinoma can be difficult. Mucocèles most commonly arise within the frontal sinuses, followed, in order of frequency, by the ethmoidal, maxillary, and sphenoidal sinuses. In the frontal sinus location they most often present with frontal headaches. Sixty percent of frontal sinus mucocèles erode through the floor of the sinus (the orbital roof) causing proptosis and frontal swelling. Left untreated, diplopia and even blindness can result. Symptoms resulting from mucocèles of the other sinuses depend on their direction of spread. Fortunately, intracranial extension is rare. Mucocèles that become infected are called *pyocèles* and present with the additional signs and symptoms of sinusitis. The most definitive treatment of a mucocèle includes the evacuation of the contents of the sinus through an open approach. The entire mucosal lining of the sinus must then be removed and the sinus duct occluded with muscle or bone. The question remains controversial as to whether or not the remaining sinus itself should then be obliterated, and if so, with what material (e.g., muscle, fat, cancellous bone).

Larynx

The most common benign neoplasm of the larynx is the *papilloma*, accounting for more than 90 percent of laryngeal tumors. Papillomas, probably caused by the human papillomavirus, usually arise on the vocal cords and usually present with hoarseness. They may be found at any site in the larynx as well as on the hypopharyngeal or pharyngeal walls and present as pedunculated exophytic masses. Laryngeal papillomas may be grouped into juvenile or adult types, depending on the age at onset. In the juvenile group there is a 2:1 female predominance, and ratio is reversed in the adult group. The typical course of each type is also distinctly different. In adults the masses are most often solitary and rarely recur after excision. In the juvenile group the lesions tend to be multiple and may recur and spread rapidly after excision. The reason for these differences is not known. Laryngeal papillomas are usually treated with laser obliteration.

Other, less common benign tumors of the larynx include oncocytic tumors and granular cell myoblastomas. The former present as a smooth submucosal mass, and the latter as a sessile mucosal mass. The most common location for both is the vocal cords, and so they too typically present with hoarseness. Chondromas of the larynx are rare benign cartilaginous neoplasms most commonly occurring on the cricoid cartilage. They can cause hoarseness, respiratory obstruction, or dysphagia. All of these benign neoplasms are managed by conservative excision.

A *laryngocele* is a herniation of the laryngeal ventricle. Three forms exist, categorized by their site of presentation. An *internal laryngocele* remains confined to the larynx and presents as an enlargement of the false cord. An *external laryngocele* protrudes through the thyrohyoid membrane, causing swelling in the anterior neck. Combinations of these two types, called *mixed laryngoceles*, occur as well. The pathogenesis is believed to be associated with chronic increases in intralaryngeal pressure. Corroborative evidence for this theory includes the fact that singers and musicians have a propensity for their development. Symptoms depend on the site of presentation, with the internal variety often causing hoarseness and the external most commonly remaining asymptomatic. Treatment includes ligation of the stalk of the laryngocele and repair of the ventricular weakness.

Odontogenic and Bone Tumors

Odontogenic Tumors

A variety of cysts and tumors of the mandible and maxilla may arise from the progenitor cells of tooth development. The majority of these odontogenic lesions are benign and may be treated conservatively. Ameloblastomas (adamantinomas) arise from dental lamina and often are associated with impacted teeth in young patients. Their usual presentation is that of a painless mass of the jaw with a multilocular, radiolucent appearance on x-ray. They occur four times more frequently in the mandible than in the maxilla. Although slow-growing, they may grow to large size and erode adjacent bone (Fig. 15-2). Treatment consists of resection of the entire lesion with a margin of bone to prevent local recurrence. Myxomas and Pindborg tumors (calcifying epithelial odontogenic tumors) are similar in their presentation and management, and they too require an en bloc resection for cure.

A second group of odontogenic tumors, including calcifying odontogenic cysts (Gorlin's cysts), ameloblastic fibromas, cementomas, and keratocysts, are generally less aggressive

than those discussed above and are treated effectively by enucleation and excision of the entire lining of the lesion.

Nonodontogenic Tumors

The tumors in this group arise from bone that is not involved in tooth development. Torus is a benign, slow-growing projection from the surface of bone. The torus palatinus occurs in the midline of the hard palate, and the torus mandibularis usually develops on the lingual surface of the mandible opposite the premolars, often bilaterally. They are both common lesions. In the United States 20 percent of the population have a torus palatinus and 8 percent a torus mandibularis. There is evidence that they are genetically inherited by autosomal dominant genes with incomplete penetrance. Tori often begin around puberty and are slow-growing. They can induce ulceration of the overlying mucosa, thus mimicking a mucosal neoplasm. No therapy is needed unless they interfere with speech, mastication, or the use of dentures, at which time simple excision is performed.

Exostoses are similar to tori and also commonly occur in the jaws. There are localized overgrowths of bone that may be nodular, pedunculated, or flat and are often multiple. They most often present in the maxilla at the canine fossa as a hard, discrete, submucosal mass. Only symptomatic masses require excision.

Osteomas are slow-growing tumors of mature bone that arise within (intraosseous) or at the periphery of the involved bone. The peripheral lesions often are attached to the cortical bone by a dense pedicle. Involvement of the bones outside of the face and skull is rare. They most commonly arise on the mandible on the lingual aspect of the ramus or on the lower border of the angle of the mandible. Osteomas also may occur in the paranasal sinuses, where they may grow to a large size. Excision is advised when continued growth encroaches upon vital structures or becomes cosmetically unacceptable. Multiple osteomas are one of the manifestations of Gardner's syndrome, the others being multiple inclusion cysts of the skin, supernumerary teeth, and familial polyposis.

Fibrous dysplasia is a benign bone disorder of unknown cause in which cortical bone is replaced by immature fibrous tissue. The fibrous tissue often proliferates and extends beyond normal boundaries, distorting and compressing vital structures. Seventy percent of patients have involvement of a single bone—monostotic—and 30 percent have the polyostotic form. In the head and neck, the mandible or maxilla is most often involved. Albright's syndrome includes polyostotic fibrous dysplasia, precocious puberty, café- au-lait spots, and several endocrine abnormalities. The management of the patient with fibrous dysplasia is dictated by the aggressiveness of the disease and ranges from observation only to extensive local resection and reconstruction. Malignant degeneration occurs in 1 percent of the patients and may be related to prior radiation therapy. Malignant degeneration should be suspected in any lesion that undergoes rapid growth or produces significant pain.

CARCINOMAS

General Considerations

Malignant neoplasms that arise in the head and neck area or upper aerodigestive tract share the general behavior of most solid tumors: local growth, locoregional spread, and distant metastasis. Their effect, however, before therapy and as a consequence of therapy, depends much more than most other neoplasms' on local disruption of function. The two main vegetative functions of the human being—alimentation and respiration—are effected by the intricate synergy of the bone, nerve, muscle, and mucosa-lined cavities that make up the head and neck. Invasive carcinoma disrupts this fine balance, resulting not only in the proliferation of abnormal cells but also in the derangement of feeding, breathing, and speaking, which may present as malnutrition, upper airway obstruction, and recurrent aspiration pneumonia. Thus unlike the case of adenocarcinoma of the breast, colon, and stomach, malignant melanoma, squamous cell carcinoma of the lung, and most other solid tumors, in which the fatal event is almost always disseminated malignancy, as many as 60 percent of patients with a fatal head and neck malignancy die without clinical evidence of metastasis beyond the local/regional disease. Central nervous system invasion, rupture of the great vessels, airway obstruction, and invasive local infection are common causes of death in these patients (Fig. 15-3). Because of the predominant local and locoregional natural history of this disease, significant attention must be paid to local diagnosis and therapy.

Most malignant tumors that develop above the clavicles are squamous cell carcinomas (epidermoid carcinomas) originating from the respiratory and stratified squamous epithelium of the upper aerodigestive tract. Although there are some differences in the natural history of tumors arising at the various sites of the upper aerodigestive tract, probably dependent on characteristics of blood supply, lymphatic drainage, or histologic variation specific to that area, most squamous cell carcinomas of the head and neck behave similarly. Their unique clinical expression then depends on the interruption of normal activity inherent in their epicenter and those areas to which they spread (Fig. 15-4). So nasopharyngeal carcinoma may present with nasal stuffiness and progress to cranial nerve dysfunction and central nervous system invasion as well as neck metastasis; and carcinoma of the floor of the mouth may present with pain, tethering of the tongue, and dysphagia, resulting in malnutrition and, sometimes, aspiration.

Moreover, the therapeutic approach to the malignancy may interfere with normal function. Laryngectomy and glossectomy obviously impair speech and swallowing tremendously. Radiotherapy, used alone or in combination with surgery, may result in deficiencies of olfaction, salivation, and infection control in the upper aerodigestive tract. Hence the survival of the patient with head and neck cancer requires consideration of both tumor growth and residual local function in formulating the therapeutic plan. The appropriate selection of radiotherapy, extirpative surgery, chemotherapy, and reconstructive surgery is crucial in the physician's attempt to prolong life and restore reasonable function and appearance. A multidisciplinary approach to this group of tumors is essential.

In both children and adults neoplasms other than squamous cell carcinoma, benign and malignant, and other clinically important masses arise in the head and neck area (Fig. 15-5). Although some of these are more likely to progress to early metastatic disease, most have significant local effect in both their clinical expression and therapy. In adults salivary neoplasms are the next most important group of lesions in the head and neck.

Epidemiology

Squamous cell carcinoma of the head and neck is not a major public health problem in the United States. Combined upper aerodigestive tract sites account for approximately 23 new cases per 100,000 males and 8 per 100,000 females, or roughly 6 percent of new cancers in males and 2 percent in females. There are 30,000 to 40,000 new cancers of the head and neck and 10,000 to 15,000 deaths per year. Approximately one-third of patients who develop a squamous cell carcinoma of the upper aerodigestive tract will die from it. Incidence and mortality rates in the United States have remained relatively stable over the past 40 years in white males but have increased dramatically for most sites in nonwhite males and in females of both groups (Fig. 15-6). Like many solid tumors, the incidence of disease increases with age in both sexes, being very rare below age thirty except in immunocompromised patients. In the United States there seems to be a clear-cut relationship between squamous cell carcinoma of the upper aerodigestive tract and the chronic use of tobacco and alcohol together. Although the use of either tobacco or alcohol alone increases the likelihood of squamous cell carcinoma of the upper aerodigestive tract, the cumulative abuse greatly increases the risk (Fig. 15-7). Previously noted increases in the incidence of squamous cell carcinoma in females were probably a result of the more widespread acceptance of cigarette smoking among women over the past 30 years. In the Western world the combination of tobacco and alcohol of various forms seems to be the predominant cause of this disease; and though there is some slight variation, the mortality rate is similar from country to country. Cigarette, cigar, and pipe smoking, chewing tobacco, and snuff have all been implicated, as have distilled beverages, beer, wine, and even mouthwash, which in many forms has a high content of alcohol. The previously demonstrated relationship of asbestos and tobacco in the carcinogenesis of lung cancer is probably operant in squamous cell carcinoma of the upper aerodigestive tract as well.

Squamous cell carcinoma of the head and neck is endemic in other parts of the world, with incidences reaching as high as 50 percent of new cancers in Bombay, India, where chewing of *pan*, a combination of tobacco, betel nut, and lime (CaOH 2), causes buccal carcinoma. Throughout southeast Asia the use of betel nut and tobacco combinations similarly results in a high frequency of oral cancers. In cultures such as those of South Asia and Southeast Asia, where alcohol is not commonly used, the prevalence of vitamin deficiency and local submucosal fibrosis of the oral mucosa may play an important adjuvant role. Reverse smoking (smoking of cigarettes and cigars with the lighted end inside the mouth) similarly contributes to the high incidence of palate and oral cancers seen in India and some parts of Central and South America. In South Africa unusually high rates of nasal and paranasal cancers have been discovered, where nickel contaminates the tobacco inhaled as snuff, and where it and other substances are industrial contaminants among workers in the nickel, furniture, and shoe industries.

In southern China and the Pacific Rim countries, carcinoma of the nasopharynx is extremely common in both men and women, up to twenty times more common than in the United States. The causative factors are unclear, although inhaled fumes, cooking smoke, chemical pollutants, or others have been proposed, as has the ingestion of salted fish. The role of the Epstein-Barr virus (EBV) in the causation and natural history of nasopharyngeal carcinoma is unclear. Infection with EBV is prevalent in this group, and there seems to be a close relationship between elevated serum EBV titers and risk. Despite the differences in proposed causative agents throughout the world, it appears clear that an important factor in the origin of squamous cell carcinoma is a chemical carcinogen and that a linear dose- time risk ratio is likely. Prevention should be possible.

Carcinogenesis

The simplest explanation of the causation of squamous cell carcinoma falls within the accepted principles of chemical carcinogenesis, more specifically cocarcinogenesis, where an *initiating* agent (perhaps some contaminant or metabolite of tobacco) induces an irreversible change in the DNA of the affected cell. A *promoting* agent (perhaps alcohol, vitamin deficiency, local inflammation secondary to trauma, poor hygiene, or submucosal fibrosis) allows that irreversible change to manifest itself as a histologically and clinically identifiable malignancy. Initiation involves alteration of the DNA by addition through covalent bonding, replacement, deletion, or other mechanism. Promotion theoretically involves facilitation of expression of the abnormal DNA via activation of cell surface and cytoplasmic protein receptors or other reactions with the epithelial cell. This theory is consistent with many of the epidemiologic and clinical characteristics of squamous cell carcinoma of the upper aerodigestive tract. The indictment of chronic use implies that initiation alone is not sufficient for malignancy. In addition to the sites in the nasopharynx and larynx where inhaled and exhaled tobacco smoke contact, high-dose exposure is possible to the mouth, tongue, and buccal mucosa as metabolites of tobacco are secreted into saliva and sit pooled in the lingual and buccal sulci. The chronic inflammation attendant to frequent alcohol use may stimulate local proliferation of epithelial cells and react with reproducing cells to allow expression of the previously altered DNA.

The role of viral carcinogenesis via initiation or promotion is unclear but certainly suspicious. Patients with papillomatosis caused by human papillomavirus of the nasal cavity and larynx have been noted to be at higher risk for development of invasive squamous cell carcinoma. Recent studies of DNA from squamous cell carcinoma of the nasal cavity and paranasal sinuses using the polymerase chain reaction have identified human papillomavirus (HPV) types 16 and 18, whereas they were not identified in tumors of other histologic type (adenocarcinoma, adenoid cystic, etc.). HPV 16 genomes also have been found in squamous cell carcinoma of the larynx, buccal mucosa, tongue, and cervical lymph node metastases. Elevated antibody to Epstein-Barr virus is associated with the presence of nasopharyngeal carcinoma but is not specific enough to be of clinical use, since viral infection is common. That the virus plays either an initiating or a promoting role in carcinogenesis is suggested by the presence of the Epstein-Barr viral genome in cervical metastases as well as in the primary tumor arising in the nasopharynx but its absence in lymph nodes of Epstein-Barr-antibody-seropositive patients that did not contain tumor and its absence in metastatic tumors of other histology.

Although malignancy of the head and neck is not commonly associated with familial syndromes, benign tumors such as exostoses, sebaceous cysts, and sebaceous adenomas may be seen in the genetic alteration that causes Gardner's syndrome and Muir-Torre's syndrome. Of the identified oncogenes, several have been discovered in squamous cell carcinoma. The *int-2* oncogene related to fibroblast growth factor and the *C-myc* gene that codes for a DNA-binding protein concerned with regulation of cell growth have been demonstrated in 11 of 21 and 2 of 21 squamous cell carcinoma tumors studied. Improved methods of DNA characterization will clarify the role of viral transformation, heredity, and increased susceptibility in the pathogenesis of squamous cell carcinoma.

Current molecular biologic and genetic investigations of squamous cell carcinoma of the head and neck have not yet produced useful mechanisms of therapy. Overexpression of oncogenes and increased levels of oncoproteins, such as the a6b4 integrins (cell surface adhesion molecules and receptors for extracellular matrix proteins), and the HER2/neu oncoproteins have been associated with rapid locoregional spread and metastasis. Inactivation of tumor suppressor genes by deletion or other mutations seems to be a common mechanism of tumor initiation and subsequent progression. It has been theorized and partially demonstrated that the formation of a head and neck tumor requires many genetic events. In vitro analysis of several lines of squamous cell carcinoma have demonstrated common allelotypes with deletions at the 9p21, ep, and 17p chromosome sites. The level of p53 mutations may predict the responsiveness of tumor to radiotherapy and chemotherapy. Another potential therapeutic mechanism being investigated is the elaboration of humorally and cellularly mediated antibodies to these overexpressed tumor cell proteins. The further delineation of the genetic identity of squamous cell carcinoma and the manipulation of protein synthesis will have a major role in future treatment.

The protean manifestations of AIDS include a number of abnormalities in the head and neck. Proliferative lymphocyte deposits in the nasopharynx have been noted to cause eustachian tube dysfunction and may be precursors of the commonly seen extranodal lymphomas. Squamous cell carcinomas of the upper aerodigestive tract appear to be disproportionately common and unusually aggressive in patients with AIDS, as in other immunodeficient states such as chronic lymphocytic leukemia. Oral and pharyngeal presentations of Kaposi's sarcoma are common in HIV-positive patients and occasionally require surgical or radiotherapeutic intervention (Fig. 15-8). Cystic degeneration of lymph nodes in the parotid and submandibular glands also occurs, and in one series this was the presenting symptom of newly diagnosed HIV infection.

Natural History

Malignancy in the head and neck follows a relatively predictable course along the continuum of histologic changes from early evidence of hyperplasia with atypia to poorly differentiated invasive malignancy. The sequential presentation of these abnormalities fits well with a theory of chronic exposure to one or several chemical carcinogens. The normal oral and oropharyngeal cavities are lined with stratified squamous epithelium similar to the skin but without the characteristic keratinization. Another difference between the skin and oral mucosa is the lack of distinct rete ridges in the mouth in normal states. Immediately subjacent is the basement membrane, beneath which lies the submucosa, containing lymphoid aggregates and lymphatic channels, blood vessels, and mucous and serous glands. In most areas of the oral cavity and oropharynx, mucosa and submucosa cover muscle. Within the epithelial layer are terminations of the nerves mediating the special senses of olfaction and taste as well as other nerve fibers. The larynx, nasal cavity, and paranasal sinuses are lined by pseudostratified columnar ciliated epithelium, which overlies a similar arrangement of minor salivary glands, nerves, and blood vessels.

Viral infection, chronic irritation by ill-fitting dentures, trauma, or infection from poor dental hygiene may elicit a response from the epithelium known as hyperplasia or papillomatosis, in which cells with normal DNA configuration and organelle structure proliferate, resulting in more prominent intraluminal projection as well as extension of the mucosa deeper into the submucosa. Despite these changes there are none of the cellular manifestations of malignancy (mitoses, pyknotic nuclei, prominent nucleoli, etc.), and the basement membrane beneath the mucosa remains intact.

Most of the clinical changes that reflect these histologic alterations, including hyperplasia, hyperkeratosis, and pseudoepitheliomatous hyperplasia (the aggressive end of this spectrum, which may be confused clinically with malignancy), have been grouped under the term *leukoplakia* (white patch). Histologic analysis shows hyperkeratosis and parakeratosis or orthokeratosis (appearance of nuclei in the most superficial layers of the mucosa with or without inflammatory cell infiltrate and acanthosis). Earlier approaches required the removal of all leukoplakia, which was believed to be a precursor of invasive malignancy. More recent thinking, however, suggests that this change is not in itself premalignant but simply evidence of chronic irritation.

Cellular manifestations of malignancy result in the diagnosis of epithelial dysplasia. Lack of a normal cellular progression to maturation characterizes dysplastic epithelium. Nuclei are larger, are hyperchromatic, and show mitotic activity. Cells may be pleomorphic with basophilic cytoplasm. Cell layers become disorganized, with loss of the gradual ascent to the epithelial surface and the presence of immature cells at the basement membrane as well as the epithelial surface. The extent of the individual cellular changes and loss of normal polarity determines whether the entity is termed dysplasia, severe dysplasia, or carcinoma in situ. All of these changes, however, overlie an intact basement membrane. The change from hyperplasia to dysplasia is thought to be irreversible and the initial step in carcinogenesis. The clinical manifestation of these histologic changes has been termed *erythroplasia* or *erythroplakia*, or red patch. These lesions appear reddish, are frequently exudative, and may have associated leukoplakia. Biopsy or excision is mandatory, because they are premalignant and may also indicate the presence of another adjacent malignancy.

The cellular changes of carcinoma in situ with loss of the integrity of the basement membrane become invasive squamous cell carcinoma. Deranged cellular function results in intracellular keratinization or keratin pearl formation (Fig. 15-9). Growth into the oral cavity may be manifested grossly as an exophytic carcinoma, whereas invasion into the adjacent muscle or bone produces ulcerative lesions.

In addition to the continuum of changes from hyperkeratosis to invasive carcinoma, patients with squamous cell carcinoma frequently demonstrate another phenomenon consistent with the theory of chemical carcinogenesis. Field *cancerization* or the *condemned mucosa* phenomenon is the finding of epithelial abnormalities throughout the entire upper aerodigestive tract in a patient with squamous cell carcinoma at one site (Fig. 15-10). Erythroplakia in the oral cavity increases the risk of invasive carcinoma in the pharynx, larynx, esophagus, and lung. The incidence of synchronous malignancies in the upper aerodigestive tract has been reported to range from 4.4 to more than 10 percent, with lung, esophagus, and other head and neck sites being most common.

Other anatomic abnormalities and functional disorders are also more frequent in patients with upper aerodigestive tract malignancy. Barrett's esophagus and esophagitis have been reported to occur in 33 percent of patients undergoing laryngoesophagectomy for advanced laryngeal carcinoma. An overall incidence of esophageal disease of 54 percent was found, with synchronous esophageal cancer in 25 percent of these patients. Metachronous or subsequent malignancies also appear at a higher frequencies in patients with squamous cell carcinoma of the upper aerodigestive tract than in the general public. Lung, esophagus, and other head and neck sites are more common, with a risk of up to 27 percent. Cessation of smoking and alcohol intake may or may not be protective.

The phenomenon of field cancerization and the high frequency of synchronous and metachronous malignancies in patients with squamous cell carcinoma of the head and neck fit nicely with the current theories of oncogenesis. Within the condemned mucosa there may be multiple subpopulations of neoplastic cells, each the clonal expression of a common initial abnormality. Different environmental conditions, subsequent mutations caused by the chronic insult of chemical carcinogens, and amplification of genetic abnormalities at one site more than another may explain the simultaneous appearance of erythroplasia, well-differentiated local carcinoma, and poorly differentiated lymph node metastasis as the disease progresses clinically and histologically. The application of the theory of *tumor heterogeneity* to the entire mucosal surface explains many of the characteristics of individual tumors such as resistance to chemotherapy and radiotherapy as well as the patient's systemic response to disease.

Although theories of malignancy that espouse a stepwise progression from local to locoregional to metastatic disease have been shown to be inaccurate for breast and gastrointestinal solid tumors, squamous cell carcinoma of the head and neck does seem to spend a considerable portion of its natural history in the local or locoregional stages. After passage through the basement membrane, malignancy invades the surrounding mesenchymal tissue, be it muscle, cartilage, or bone. Proteases and collagenolytic and osteoclastic enzymes facilitate the destruction of the adjacent tissues. Invasion of subjacent nerves opens the perineural spaces, and tumor emboli may pass in the perineural lymphatics. The usual spread of tumor is from the primary site via the subjacent lymphatics (Fig. 15-11) to the neck or by continued local invasion through nerve and bone to the base of the skull. The virulence of the primary tumor and its likelihood of metastasis or ability to cause the death of the host have been estimated by a number of methods. The size of the primary tumor is the main parameter determining its clinical stage in the TNM (tumor-node- metastasis) staging system adopted by the American Joint Committee on Cancer (AJCC) and other worldwide organizations (Fig. 15-12). Although it is a crude method, it is fairly reliable in suggesting the prognosis of the patient and the appropriate form of therapy.

Histologic characteristics such as degree of differentiation (primarily a function of nuclear morphology), pattern of invasion (blunt pushing borders vs. noncohesive jagged borders), microvascular invasion, perineural invasion, and tumor thickness have been used to predict the likelihood of lymph node metastasis and overall prognosis. The amount of tumor angiogenesis within the cancer as measured by the intensity of staining of the endothelium is a cogent predictor of aggressiveness. Advances in biochemical and molecular investigation have allowed more careful consideration of the primary tumors. Flow cytometric analysis of tumor suspensions shows that aneuploidy is common in primary tumors (68 percent) and in cervical metastases (82 percent). Aneuploid tumors are more likely to metastasize than are diploid tumors. Correlation also has been found between thymidine labeling index and the T stage of squamous cell carcinomas, with T3 lesions having significantly more activity than T1. Further subcellular analysis of squamous cell carcinoma has attempted to correlate abnormalities and thus predict behavior.

Expression of histocompatibility antigens HLA-1 and HLA-2 on squamous cell carcinoma demonstrates marked tumor heterogeneity, particularly in poorly differentiated tumors. The relationship of growth factors to primary tumor development and progression is unclear, but, as in other tumors, such as breast and brain, growth factor receptors have been identified on squamous cell carcinoma of the head and neck. Higher absolute levels of epithelial growth factor receptor and demonstration of amplification of the gene for epithelial growth factor receptor in squamous cell carcinoma of the upper aerodigestive tract may indicate that epithelial growth factor plays an important role in local tumor behavior. The histologic, biochemical, and genetic characteristics of head and neck tumors also have been used in an attempt to predict response to radiotherapy and chemotherapy.

As with most solid tumors, the most cogent prognosticator of head and neck cancer is the presence or absence of lymph node metastases. In the unoperated patient, passage of tumor emboli through the lymphatics follows an orderly pattern, with the anatomic triangles of the neck reflecting the site of primary disease. In the patient who has undergone previous surgical therapy or radiotherapy or who has extensive tumor that blocks normal lymphatic flow, metastasis from a primary tumor may not follow the usual pathways, and a neck mass may not predict accurately the site of recurrence or new primary disease (see Fig. 15-3 B). Clinical and ultimately pathologic staging of the neck according to the AJCC system depends on number of lymph nodes, size of lymph nodes, fixation to the skin or subjacent neck muscles, and laterality with respect to the primary tumor (ipsilateral, contralateral, and bilateral). Although there are many determinants of primary tumor behavior, size and differentiation are useful predictors of the risk of metastasis. Increasing T stage is generally reflected by increasing N stage. Distant metastasis, however, is more closely related to N stage than to T stage. In a study of 160 patients with extensive disease in the neck, N3a, only 63 completed combined therapy with surgery and radiation and, of these, only 13 showed no evidence of disease at 2 years. Disseminated disease was the most common cause of death, followed by failure at the primary site and in the neck.

Lymph node metastases appear to behave in some ways similarly to primary disease in that micrometastases are not as poor a prognostic indicator as clinically palpable metastatic disease. Spread of the tumor outside the capsule of the lymph node may occur in clinically negative and positive lymph nodes and is a poor prognostic sign, indicating a more aggressive tumor. In a study of stage III carcinoma of the oral cavity, the 2-year survival rate of patients with no lymph node metastases was 87 percent; with intracapsular lymph node metastases, 75 percent; and with cervical metastases with extracapsular spread, only 39 percent. The incidence of extracapsular spread of cervical disease increases with the size of the lymph node. Proportional to increase in extracapsular spread is the likelihood of recurrent disease, and thus of death.

Uncontrolled growth of squamous cell carcinoma in the neck results in carotid artery hemorrhage, invasion of the sympathetic ganglia resulting in Horner's syndrome, erosion of the

cervical vertebrae, invasion of cranial nerves IX, X, and XI (jugular foramen syndrome) and of cranial nerve XII at the base of the skull, airway obstruction, and brachial plexus palsy. The site of the primary tumor dictates not only the site of cervical metastasis but the frequency as well. Although the size of the primary lesion at presentation is important, it seems that independent of size, squamous cell carcinomas of the tonsil and the base of the tongue have high rates of metastasis to the neck, and lesions of the buccal mucosa and palate have low rates.

Distant metastases from squamous cell carcinoma have been reported to range from 31 percent of patients with no evidence of cervical metastases to 59 percent of those with extensive neck disease. Despite this relatively high incidence, they are not uniformly the cause of death. Although any site is possible, lung, bone, skin, and liver are the most common metastases. Systemic effects of both local and systemic disseminated tumors include hypercalcemia from bone metastasis and elaboration of parathormonelike peptides as well as syndrome of inappropriate antidiuretic hormone secretion (SIADH) from vasopressinlike substances. There is some evidence that the widespread use of chemotherapy has changed the natural history of metastatic squamous cell carcinoma, increasing the frequency of patients dying with disseminated disease.

Diagnosis and Evaluation

History and physical examination are the most important considerations in the diagnosis of carcinoma of the upper aerodigestive tract. A history of chronic tobacco and alcohol abuse places the patient in the high-risk category. Males over forty with such a history comprise 70 to 80 percent of most series of patients with head and neck cancer. A previous occurrence of lung cancer, esophageal cancer, or other head and neck malignancy, diseases causing immunodeficiency such as renal failure with its therapy by transplantation, malnutrition, and AIDS are also significant.

Symptoms referable to the tumor itself usually are mild and not commensurate with the size of the tumor, often because of the patient's generally stoic personality and denial. Late-stage presentation is common in these patients. Barkley noted that 36 percent of patients presenting with oropharyngeal lesions had local disease alone and that the remaining 64 percent already had cervical or disseminated metastases. Pain at the site of the tumor is not a frequent complaint. Because otitis media and externa are relatively rare problems in adults, pain in the ear in a patient over forty may be a manifestation of tumor in the oral cavity, oropharynx, or larynx via referred pain pathways, including the lingual to auriculotemporal nerves, glossopharyngeal to tympanic nerve, or vagus to auricular nerve. Prograde neural symptoms such as formication (the feeling of ants crawling along the lip or cheek) may represent mental or infraorbital nerve invasion by buccal, labial, or alveolar carcinoma. Family members or the patient may note changes in speech caused by tethering of the tongue (Fig. 15-13). Constant or variable hoarseness is the sign of vocal cord impairment by local growth of laryngeal or hypopharyngeal cancer or by paralysis in the neck of the recurrent laryngeal nerve. The sensation of scratchiness or tickling in the throat, gagging on food, or nocturnal choking secondary to aspiration may all represent interference with the synergistic mechanisms of swallowing and breathing. Airway compromise is usually a late symptom but occasionally may be the first presentation in an emergency setting. The typical American patient, then, is male, over forty, with a long history of tobacco and alcohol abuse, frequently with underlying psychologic or behavioral disorder such as depression or antisocial personality with withdrawal.

Visualization of the entire upper aerodigestive tract, using a systematic approach that includes inspection of the facial and cervical surface anatomy and contour, intraoral examination, and indirect (mirror) laryngoscopy with nasopharyngoscopy, is essential for diagnosis and staging (Fig. 15-14). In addition to size, shape, and projection into the cavity (exophytic or degree of ulceration), it is important to note the mobility of the tongue and its position relative to the midline (hypoglossal nerve function), any fixation of the tongue to the adjacent mandible, and any direct invasion of the mandible. Mirror examination of the oropharynx and hypopharynx and larynx should note the patency of the vallecula, the distensibility of the pyriform sinuses, and the normal adduction to the midline of the vocal cords. Interference with vocal cord function may result as hypopharyngeal lesions invade the medial wall of the pyriform sinus or by paralysis of the recurrent laryngeal nerve. By retracting the soft palate forward with a transnasal catheter, visualization of the nasopharynx is possible with the mirror. The vault of the nasopharynx, the choanae, and the eustachian tube should be examined. The condition of the teeth and gums should be evaluated and the presence of torus palatinus or torus mandibularis noted. Limited motion of the mandible may come from direct tumor invasion or, more frequently, from invasion of tumor through the retromolar trigone area or tonsillar fossa to the pterygoid muscles. Infiltration or inflammation of the internal and external pterygoids limits jaw motion and causes the uncomfortable symptom of trismus, an ominous clinical sign. Compilation of the appropriate information about the primary tumor allows clinical staging. The distinct anatomic site, size, pattern of growth, and invasion of adjacent structures are all necessary to assign the appropriate T stage.

Examination of the neck will reveal the presence or absence of lymph node metastases. The site of the primary tumor should predict the most likely site of metastatic disease (see Fig. 15-11). Enlargement or inflammation of the submandibular gland may interfere with evaluation of the neck. Careful palpation, preferably performed with the examiner standing behind the seated patient, allows systematic and sequential evaluation of the submental, submandibular, jugulodigastric, midjugular, juguloomohyoid, posterior triangle, and supraclavicular lymph node stations. The number of lymph nodes, lymph node size, and any fixation to the skin or subjacent muscle allows the examiner to assign an N stage. Tumors with no evident primary site are manifested as lymph node metastases in 3 to 4 percent of malignancies. Cervical metastases of squamous cell carcinoma comprise two-thirds of these cases. Furthermore, the initial manifestation of the subsequently discovered disease is a neck mass in 25 percent of the patients with carcinoma of the oral cavity, oropharynx, and thyroid, and in 50 percent of patients with carcinoma of the nasopharynx. Because of variable extension into the neck, parotid gland masses may present as a cervical lymph node. The characterization of masses in the neck can be guided largely by age. In children congenital and inflammatory masses predominate; in young adults, inflammation or proliferative disorders of the lymphatic system; and in adults, metastatic disease from the upper aerodigestive tract and skin.

Careful physical examination, including neurologic examination of the remainder of the head and neck area, may reveal evidence of more extensive disease such as cavernous sinus invasion, as documented by extraocular movement disorders, or invasion of the cervical sympathetics, as indicated by Horner's syndrome.

Distant metastases are evaluated by history, physical examination, laboratory procedures, and radiology. Pleuritic pain or shortness of breath may indicate lung involvement, and distinct pain at a specific site may indicate spread to distant bone.

Before treatment planning, definitive histologic confirmation of disease is necessary. If a primary site is visible, a wedge biopsy specimen should be taken at the edge of the tumor to include some adjacent normal tissue, fixed in formalin, and examined histologically. Touch preparations and other cytologic methods are occasionally helpful but suffer from lack of specificity and sensitivity. Although the extent of the tumor usually can be assessed by direct visualization, the use of vital dyes such as toluidine blue may help to delineate the extent of the disease, staining only the malignant and dysplastic epithelium and washing off the surrounding inflamed tissue. Hematoporphyrin dyes share this characteristic of differential uptake by tumor and normal cells and have occasionally been of use in demarcating index tumors and identifying and sometimes treating multicentric disease. Because of the significant incidence of synchronous primary tumors, evaluation of the rest of the upper aerodigestive tract may be useful. The value of triple endoscopy—bronchoscopy, esophagoscopy, and direct laryngoscopy—has been debated, but it is advisable for the ideal work-up of an advanced head and neck cancer in a patient with a long history of tobacco and alcohol use.

Since treatment planning depends a great deal on the clinical stage of disease, and since therapies vary greatly in morbidity to the patient, precise and accurate staging is essential. Extent of resection and volume of radiotherapy depend on a reliable assessment of the degree of invasion of the primary tumor. The use of adjuvant chemotherapy or radiotherapy or the inclusion of neck dissection in the surgical plan usually is determined after careful analysis of the neck. Various radiologic studies may help to define the extent of the disease.

Radiologic evaluation of intraoral disease usually involves assessment of the mandible. If tumor abuts the mandible in the lingual or buccal sulcus and is not easily mobile, it may penetrate the cortex. In edentulous patients, tumor descends through the cortical defects left by the previous teeth along the occlusal ridge. Once the medullary canal has been invaded, spread may occur locally or by way of the inferior alveolar nerve. In dentulous patients, direct invasion of the lingual plate occurs, and spread usually does not extend within the bone beyond the extent of the soft-tissue tumor. Dental films taken directly at the site of tumor contact show cortical invasion or widening of the periodontal membrane. Panoramic films and mandibular series show the extent of disease within the mandible by cortical disruption and widening of the inferior alveolar canal or mental foramen. Simple plain-film radiographic analysis of the mandible, however, may be inaccurate, with one series of 111 patients revealing a false-negative rate of 44 percent and a false-positive rate of 9 percent. Technetium-99m radionuclide bone scans have been advocated for diagnosis of mandibular invasion. Although their sensitivity is greater in general than plain films, their specificity is not great, with false-positive rates of 53 percent and false-negative rates of 12 percent. In large tumors bone scans are even more likely to be false-positive because of surrounding inflammation. CT is a highly sensitive method of diagnosing cortical invasion, but the presence of metal tooth fillings interferes with its accuracy (Fig. 15-15). MRI is probably the most accurate and useful method of evaluating the mandible in suspicious but not definitive cases, because it can accumulate and reformat data in any plane without repositioning the patient, it offers superior separation of cortical and marrow images, and it is unaffected by dental fillings.

Analysis of the extent of intraoral and paranasal sinus tumors in difficult-to-examine areas such as the parapharyngeal space, larynx, and nasopharynx is assisted by both CT and MRI. Correlation between extent of disease determined at operation and preoperative CT scanning was found in 78 percent of 26 patients evaluated by high-resolution CT, and 94 percent when MRI with contrast enhancement with gadolinium DTPA was used. Particular advantage is obtained in evaluation of muscle and bone invasion, but mucosal detail is difficult to assess with both CT and MRI (Fig. 15-16).

The administration of agents such as gadolinium with MRI provides better soft-tissue contrast and can suggest the differences between tumor and surrounding inflammation. By reformatting images in several planes, a three-dimensional characterization of the tumor is possible that can better demonstrate tumor relationships to nearby normal structures and can aid in more accurate staging. MRI-directed stereotactic surgical techniques have been used in neurosurgery and in the administration of radiotherapy and may have a role in surgical extirpation of carcinoma of the head and neck.

Since survival seems to depend on eradication of regional as well as local disease, evaluation of the neck for the risk of metastatic disease is crucial to the overall treatment plan. Surgery or radiotherapy may be used to remove regional disease in the lymph nodes if there is disease that is clinically apparent (palpable N+) or if there is high risk of disease in a clinically negative neck (N0). As has been discussed, the size of the primary tumor (T stage), its histologic parameters, and its site all may suggest an increased likelihood of subclinical disease in the neck. Unfortunately, physical examination of the neck may be faulty, with false-negative rates ranging from 16 to 60 percent. Although of some importance, the rates of false positivity are generally lower and usually not as clinically relevant. Lymphangiography has been unrewarding in the evaluation of the neck. CT has recently been evaluated and found in most studies to be slightly more sensitive than physical examination: 82 percent sensitivity versus 75 percent, 93 percent versus 70 percent, and 90 percent versus 82 percent in three studies. Implicit in this advantage is the ability of CT to upstage disease from N0 to N+ and suggest the need for neck dissection. Gadolinium-enhanced MRI appears even more sensitive—92 percent sensitivity versus 82 percent for CT scanning and 75 percent for physical examination—and is also useful in upstaging disease. Although these technologies are not necessary in every case of clinical staging, they are of significant value in decision-making when physical examination of the neck is difficult because of obesity, anatomic variation, or previous surgical or radiation therapy. Micrometastases of less than 3 mm, however, are still undetectable by any available technology, although position emission tomography using fluorine 18-labeled 2-deoxyglucose (PET-FDG) has demonstrated exquisite sensitivity in early studies of squamous cell carcinoma and melanoma metastatic lymphadenopathy.

Therapy

Decision-making on therapy is based almost completely on the clinical stage (TNM) of the tumor at the time of presentation. Although histologic characteristics may to a slight

degree influence the treatment, the size of the local tumor and the extent of invasion of adjacent structures (T stage); the presence or absence, the number, and the laterality of cervical metastases (N stage); and the presence or absence of distant metastatic disease (M stage) dictate therapy. Definitive or curative treatment methods all are oriented toward total extirpation of local and locoregional disease, with the expected subsequent decrease in disseminated disease. Palliative procedures, which may produce relief of pain, relief of airway obstruction, or improvement in local function and hygiene, similarly require complete macroscopic tumor removal at the primary site and in the neck and may occasionally be justifiable in the presence of distant metastases. Subtotal resection of local or locoregional disease is unlikely to be of any benefit in any situation.

Definitive therapy may consist of surgery alone, radiotherapy alone, surgery with radiotherapy as a preoperative or postoperative adjuvant, or chemotherapy delivered either systemically by intravenous route or locally by intraarterial infusion before either or both modalities as neoadjuvant therapy. For small tumors (less than 2 cm, T1), surgery or radiotherapy, well planned and appropriately executed, in most cases will have equivalent local control and survival rates (Fig. 15-17). The choice of the method, then, depends on patient compliance, volition, associated disease, expense, interference with normal function, and available facilities. As the size of the tumor increases to T2 stage or greater, the likelihood of local control and ultimate cure with radiotherapy alone decreases, so surgery or surgery with adjuvant radiotherapy becomes preferable (Fig. 15-18). In the larynx, where staging is more a function of invasion of adjacent structures than of tumor size, and where function is completely dependent on preservation of structure, radiotherapy is usually the first choice for T1 and T2 lesions and is occasionally used as definitive therapy for T3 lesions. Advanced disease has a poor prognosis even with combined treatments of surgery and radiotherapy.

Unfortunately, postoperative adjuvant chemotherapy has been unsuccessful in prolonging life in controlled randomized trials. Despite high response rates to preoperative administration of cisplatin and 5-fluorouracil there has not been the survival advantage that had been demonstrated with chemotherapy for breast disease. One possible role for chemotherapy, however, is as a predictor of successful therapy and thus an indicator of which patients may be treated with radiotherapy instead of surgery with better preservation of function. In several prospective studies patients with T3 or T4 laryngeal tumors were treated with cisplatin, 100 mg/m² body surface area, and 5-fluorouracil, 100 mg/m²/day × 4 for three courses. Patients with complete response (disappearance of all tumor) were then treated by either radiotherapy alone or the more conventional approach of surgery (laryngectomy) with postoperative adjuvant radiotherapy. Patients with less than complete responses were treated with surgery plus radiotherapy. Results show no difference in survival between patients who respond completely to chemotherapy whether they had been treated primarily with surgery, radiation therapy, or not. Patients with incomplete response to chemotherapy had much poorer survival than either group of complete responders. These findings may indicate that the natural history of tumors that respond completely was more benign and that they can be treated less aggressively, preserving function. Approximately 30 to 40 percent of patients treated with chemotherapy and radiotherapy can survive with a preserved larynx. Nonresponders probably demonstrate greater tumor heterogeneity or clonal resistance to therapy and thus are unlikely to have their tumor controlled by any therapy. The results of these trials on carcinoma of the larynx have stimulated investigators to study other head and neck sites for the possibility of organ preservation. Given a 40 to 50 percent complete response to chemotherapy, improved techniques in patient selection (genetic analysis of tumors), and radiotherapy, there is speculation that higher cure rates with preservation of function are possible. Salvage surgery in these previously irradiated patients, however, is difficult, with major complication rates approaching 40 to 50 percent.

The basic principle of solid tumor therapy is en bloc treatment, either resection or radiotherapy of the primary tumor and the regional disease in the neck. The decision of whether to treat the neck or not depends on the presence of clinically discernible metastatic disease or the risk of micrometastases to the neck. As primary tumor size increases, risk of neck disease increases at a greater rate for some primary sites than others (Fig. 15-19). When palpable lymph nodes are present in the neck, confirmation of metastatic disease may be obtained with fine-needle aspiration and cytologic examination, or the decision to proceed with therapy may be made on purely clinical grounds. Palpable or radiologically positive lymph node metastases require surgical therapy in the form of some type of neck dissection, usually performed in continuity with the resection of the primary tumor. Subclinical disease or micrometastases may be treated by a modification of neck dissection or radiotherapy, depending on the modality chosen for treatment of the primary site.

The lymph nodes draining the head and neck are contained in a fascial envelope, between the superficial and deep layers of the investing fascia of the neck. Within the layers of the investing fascia lie the sternocleidomastoid muscle and the spinal accessory nerve (CN XI). The cervical lymph nodes lie between the investing fascia and the deep cervical fascia (Fig. 15-20) or prevertebral fascia and can be removed in toto with the jugular vein, the sternocleidomastoid muscle, and the accessory nerve, as in the classical radical neck dissection described by Crile in 1906 (Fig. 15-21). The lymphatic connections between the tumor and the cervical metastases should remain intact with en bloc resection, and all the lymph- node-bearing tissue of the exposed side of the neck, including the anterior and posterior triangle, is removed.

The major morbidity of neck dissection is secondary to paralysis of the trapezius muscle by resection of the accessory nerve. Because of this, Bocca in 1967 suggested preservation of the nerve. Other surgeons have demonstrated techniques in which the jugular vein and the sternocleidomastoid muscle also can be left intact. These methods are particularly useful in prophylactic (elective) neck dissections when there is no clinical evidence of metastatic disease to the neck. Although there is some debate about the quality of function after modified neck dissection, it seems likely that long-term function is improved when the spinal accessory nerve is preserved, and the likelihood of survival is not significantly impaired by the lesser procedure.

Since the site of potential neck metastasis can be fairly accurately predicted by the location of the primary squamous cell carcinoma, and since the lymph-node-containing areas of the neck have been described as discrete anatomic areas demarcated by muscles, fascial condensations, and the triangles of the neck, it has been suggested that in addition to leaving the accessory nerve and the sternocleidomastoid muscle intact, only those lymph nodes at risk should be resected. The concept of *selective neck dissection*, customized to the site of primary disease, has gained support. Thus for carcinoma of the lip, anterior tongue, floor of the mouth, and buccal mucosa, *supraomohyoid neck dissection* might be used, removing the submental, submandibular, upper and midjugular lymph nodes (Levels I, II, and III). If a nasopharyngeal lesion or posterior scalp melanoma was the primary site, the suboccipital, retroauricular, upper and midjugular, and posterior triangle lymph nodes (Levels II, III, IV, and V) would be removed via a *posterolateral neck dissection*. In primary sites of the pharynx or larynx, a *lateral neck dissection* including upper, middle, and lower jugular nodes (Levels II, III and IV) might be appropriate. In thyroid disease, removal of the paratracheal, perithyroidal, and precricoid nodes in an *anterior compartment neck dissection* might be appropriate (Fig. 15-22). Although it seems obvious that control of disease in the neck by its surgical removal with some form of neck dissection should improve survival, there is no well-designed prospective study to prove the equivalency or superiority of neck dissection or radiotherapy for the clinically negative (N0) neck or to suggest which operation of the several described should be used.

Approximately 70 percent of patients in the United States who present with squamous cell carcinoma of the upper aerodigestive tract have advanced disease, stage III or stage IV. Even with aggressive therapy late- stage disease has a high recurrence rate. Recurrent disease may occur at the primary site in the previously treated neck and in the contralateral neck. Just as with primary disease, the therapy of recurrent disease is based on wide local resection and clearance of the involved or at-risk regional lymph nodes. The extent of resection depends on involvement by tumor and the proximity to the great vessels and the central nervous system. If tissue tolerance to ionizing radiation allows, postoperative adjuvant radiotherapy may be useful, even as a second course. For unresectable recurrent disease, combinations of chemotherapy and radiotherapy or chemotherapy alone as well as laser or other noninvasive surgical methods of tumor reduction may provide some palliation in reducing bulk, alleviating pain and airway compression, but long-term survival does not appear to be improved.

The therapy of distant metastases of squamous cell carcinoma of the upper aerodigestive tract has been unsuccessful. The lung is a common site of metastasis for head and neck cancer, but it must be realized that a solitary lesion in the lung of a patient with a previous squamous cell carcinoma of the head and neck is more likely to be a primary lung cancer than a metastatic deposit. The patient should be aggressively evaluated and treated for that primary cancer. Resection of metastatic disease to the lung, however, has not proved salutary in most cases. Second primary tumors or metachronous tumors are a significant problem in head and neck cancer. Prevention of recurrent disease and of new primary disease and reversion of possible premalignant entities may be possible with chemoprevention. Isotretinoin, a vitamin A analogue, and beta carotene have reduced concomitant premalignant changes, and have possibly reduced the incidence of second primary malignancies in patients with epidermoid carcinoma of the upper aerodigestive tract.

Immunotherapy

Because of prolonged alcohol abuse and coincident malnutrition, many patients with squamous cell carcinoma of the head and neck present with various manifestations of immunodeficiency. There is evidence that anergy and deficient cellular immunity result in poorer survival rates for squamous cell carcinoma of the head and neck as well as for other malignancies.

Cell-mediated immunity has been demonstrated to be depressed in head and neck cancer patients. The cause of this depression is uncertain, however, and surgery, radiotherapy, the malignancy itself, and suppression of natural killer cell function by circulating immune complexes all have been implicated. Natural killer cells have been shown to function as a recognition and defense mechanism against metastatic disease from primary squamous cell carcinoma. The effects of humoral or B-cell-mediated immunity on the progression of head and neck carcinoma are not as clear. Unlike some malignant cells, the squamous cell carcinoma tumor cultures have not been particularly immunogenic, and cell-specific immune manipulation has not yet been achieved.

Nonspecific cellular immunostimulation with various agents, such as levamisole, thymosin, and interferon, however, has shown some promise both as an adjuvant and as palliative method. Natural killer cells from patients with head and neck cancer treated in vitro with interleukin-2 (IL-2) show increased activity after therapy, suggesting that IL-2 negates some of the suppressive agents in the serum. The use of IL-2 has been extended to the clinical arena, with infusion of this agent with intramuscular interferon- α reversing in vivo depressed natural killer cell activity. Perilesional injection of IL-2 in recurrent inoperable head and neck carcinoma resulted in temporary but dramatic response in 65 percent of 20 patients treated. Systemic recombinant interferon- α given intramuscularly resulted in one complete response, one partial response, and two stabilizations of disease in 14 patients treated with a second cycle, with salutary effect being attributed to a rise in natural killer cell activity.

Although the benefit is as yet uncertain, it appears likely that more specific characterization of the antigenic identity of squamous cell carcinoma and more precise manipulation of effector cells through natural or synthetic lymphokines will ultimately help in the adjuvant therapy of head and neck cancer.

Reconstruction

In the treatment of head and neck cancer, as in any other disease, there is a hierarchy of priorities. *Survival* obviously is of first concern and depends a great deal on adequate surgical or radiotherapeutic ablation and possibly adjuvant chemotherapy or immunotherapy. *Freedom from pain* fortunately is accomplished in most cases of successful ablative therapy. Preservation or restoration of *function* as well as *appearance* is the next consideration and relies largely on the ability of the surgeon to repair the created defect with local, regional, or distant tissues. Finally, but perhaps most practically important to the patient, is the *efficiency* of the treatment regimen. Efficiency dictates that the therapy be delivered in a time period commensurate with the natural history of the disease.

Advanced squamous cell carcinoma has a high recurrence rate. Eighty-nine percent of patients whose local or regional disease recurs have their recurrence within 2 years of therapy. Thus restoration of function and appearance in this group of patients with a high failure rate and short disease-free intervals should be performed as quickly as possible. A multiple-stage method of reconstruction carried out over several months may prolong recovery. The most efficient methods are single-stage reconstructions performed at the time of the ablative surgery.

In the past there was considerable reluctance to perform immediate reconstruction after resection of malignancy. Fear of cloaking persistent or recurrent disease, coupled with realistic assessment of the poor results from multiple procedures, led to recommendations that the patient be observed for 1 to 5 years after surgery. The functional disability inherent in major oral or oropharyngeal resection usually mandated permanent gastrostomy for feeding and tracheostomy for safe maintenance of the airway. The attendant loss of taste, speech, swallowing, and other corollary functions made the postoperative state of the patient with head and neck cancer miserable.

Improved methods of reconstruction, better pathologic analysis at surgery, and a more comprehensive understanding of the natural history of the disease have made single-stage reconstruction at the time of the initial surgical resection the current standard of care in most instances. Resection of the primary disease and regional metastatic disease, confirmation of disease clearance by frozen-section examination of the margin of resection, and immediate reconstruction are usually possible for squamous cell carcinoma of the upper aerodigestive tract. In malignancies for which frozen-section analysis may be inaccurate, when bone is involved, in recurrent disease with previous radiotherapy, or when there is uncertainty about other aspects of the resection, secondary reconstruction may be more appropriate.

The upper aerodigestive tract is a complex mixture of cutaneous cover, epithelial lining, bone, and cartilaginous framework all joined in a complex arrangement by muscle and driven in intricate synergy to facilitate the main vegetative functions of the organism, alimentation and respiration. This complex mobile structure and the heavy colonization of the mucosal surfaces by bacteria as well as the deleterious effects of adjuvant radiotherapy (acute inflammation in the early stages and fibrosis and vasculitis in later stages) all make reconstruction extremely difficult.

The basic needs presented by surgical resection are restoration of continuity of the alimentary tube with epithelial lining, provision of reliable external coverage for protection of the great vessels and bony structures, and separation of the central nervous system and upper aerodigestive tract. Restoration of oral continence, facilitation of the coordinated motions of the tongue and larynx, and maintenance of an open passage for swallowing while separating the oral, oropharyngeal, and nasal cavities are refinements on the basic demand that are necessary for a reasonable quality of life. Accurate *analysis of the wound* created by the surgical resection is the first element required for successful reconstruction. Size, exposure of the central nervous system, mobility of the removed parts, presence of bacterial colonization or invasive infection, type of tissue removed (mucosa, bone, cartilage), history of previous surgery or radiotherapy, likelihood of subsequent surgery or radiotherapy, exposure of the carotid or jugular vessels, and effect of external appearance all are important wound characteristics that affect the choice of reconstructive techniques. Whether to attempt to satisfy only the basic reconstructive needs or to restore as many missing elements as possible is a difficult decision involving patient desires and compliance, surgical skill, consideration of disease stage, and many other factors.

The past 15 years have brought enormous advances in reconstructive techniques that have affected mainly the *efficiency* of therapy in restoring the patient to reasonable function and appearance rapidly. Indirectly, however, improved reconstructive techniques have an impact on survival, since more aggressive resections, salvage of radical radiotherapy, and decrease in postoperative complications with attendant infection and malnutrition are accomplished with relative safety. The fundamental improvement has been the ability to transfer large volumes of well-vascularized tissue to the area.

Although flap reconstruction of the external surface of the nose was taught by the early Hindu surgeon Sushruta, and the Renaissance saw a number of reconstructive attempts at nasal and other external defects (by distant flaps attached to the defect and divided after parasitizing their blood supply from local tissues), major reconstructions have been a relatively recent phenomenon. Throughout the development of surgical technique that followed the introduction of general anesthesia in 1846, the principle of random flaps of skin being attached and divided remained the mainstay of the reconstructive effort. Large segments of tissue from the chest and back were moved in multiple stages to the oral cavity, face, and pharynx.

In 1965 Bakamjian described the deltopectoral flap, which possessed an *axial* or direct arterial blood supply from the perforating vessels of the internal mammary artery and vein to the skin of the chest and shoulder. This provided a relatively reliable large segment of tissue that was particularly useful in the reconstruction of the pharynx. The forehead flap, another axial pattern flap, based on the superficial temporal vessels, was described in 1963 and, despite its disfigurement of the donor site, became a useful method of reconstruction of the oral cavity. With the realization in the 1970s that the blood supply to the skin came not only from the randomly oriented subdermal plexus vessels and axial cutaneous vessels but also from perforating vessels from the subjacent muscles, the *musculocutaneous concept* transformed reconstruction, particularly that of head and neck defects. Large flat muscles of the thorax could be rotated on their long vascular pedicles to supply a large volume of well-vascularized tissue in a single operation on the oral cavity, pharynx, or soft tissues of the face. Moreover, if recurrence of disease mandated a subsequent resection, another thoracic musculocutaneous flap was available to repair the defect. The pectoralis major, latissimus dorsi, trapezius, sternocleidomastoid, and platysma muscles all are useful, either alone or with their overlying skin. Now much larger and more complex wounds could be addressed at a single operation, returning the patient to reasonable function and appearance promptly.

Some of the problems with the thoracic musculocutaneous flaps are the effect of gravity on bulky flaps, additive morbidity to the shoulder girdle when neck dissection is performed, and variable blood supply to the skin, particularly in the pectoralis major musculocutaneous flap, depending on where the skin portion of the flap is located. Furthermore, there is no reliable method of transporting vascularized bone with a regional musculocutaneous flap.

Despite these disadvantages, the musculocutaneous concept has made immediate reconstruction of the head and neck resection the accepted procedure in most cases.

An offshoot of the success of reconstructive efforts with musculocutaneous flaps has been the increased interest in vascular anatomy (Fig. 15-23). Subsequent research and improvements in microscope and instrument technology have resulted in the ability to transfer tissue of many different types from various sites of the body by separating arterial supply and venous drainage of the tissue and reattaching it to blood vessels in the head and neck. Bone, muscle, skin, fascia, and combinations of these are available for various sites, as are intraabdominal viscera. Microvascular reconstruction or free tissue transfer has made it possible for the surgeon to close virtually any defect in the head and neck, no matter how large or complex. Even more important, however, is that the large number of methods available allows the reconstructive surgeon to choose the method or methods most suitable to a specific site and analyze the results (Table 15-1). Just as each wound has its own characteristics, so does each flap. Vascular pedicle length, bulk, type of epithelium, presence and durability of bone, and thickness of soft tissue all can be evaluated to select the most appropriate replacement for the individual problem (Fig. 15-24).

Complications

The therapy of squamous cell carcinoma of the head and neck usually requires two potent modalities—surgery and radiotherapy—directed at an area that is heavily contaminated with saprophytic and pathogenic organisms in a patient who is frequently malnourished and immunodeficient, and may be noncompliant. Since many patients present with advanced disease, the failure rate even for combined therapy is high as a result of disseminated disease as well as local or locoregional failure. It is therefore imperative to deliver the therapy in a form that is efficient and results in the patient's rapid return to function, reasonable appearance, and whatever social situation the patient may be able to recover. Recognition of the inevitable sequelae of therapy—complications—and rapid resolution or, preferably, prevention are thus of great importance.

Complications specific to head and neck cancer therapy can be categorized as *anatomic*, injury to nerves or blood vessels within the field of surgery; *physiologic*, the results of interference with blood or lymphatic supply to the area secondary to surgery or radiotherapy; *technical*, surgical rearrangements that result in secondary problems; and *functional* derangements of normal behavior secondary to therapy. All of these can then be grouped into catastrophic or noncatastrophic complications, which will in large part dictate the surgeon's approach to them, whether preventive or therapeutic.

The most appropriate approach to complications is prevention. Restoration of positive nitrogen balance, preoperative pulmonary hygiene, control of diabetes mellitus, and weaning from alcohol and tobacco are important nonspecific measures. Preoperative antibiotics decrease the likelihood of wound infection and its sequelae. Numerous studies have shown that previously administered radiotherapy, particularly if given in definitive therapeutic doses, increases the risk of complication. Dental hygiene or rehabilitation before definitive surgery is important in patients whose mandible has been previously irradiated to prevent subsequent osteoradionecrosis. Patient education is crucial to ensure cooperation in what may be a difficult postoperative rehabilitation.

Other than injury to the thoracic duct, which may result in significant fat and protein loss through a chylous fistula, most *anatomic* complications are nerve injuries, either purposeful or otherwise secondary to primary tumor resection or radical neck dissection. The accessory, marginal mandibular, mylohyoid, and cervical plexus sensory branches are frequently sacrificed in neck dissection. Injuries caused by traction, electrocautery, or other technical misadventure may occur to any structure but are most likely to affect hypoglossal, lingual, mandibular, vagus, phrenic, facial, recurrent laryngeal, motor branches to the cervical plexus and cervical sympathetic chain. Careful technique during surgery, adequate hemostasis to allow good visualization, and knowledge of normal and pathologic anatomy will decrease the likelihood of anatomic complications.

Previous surgery, the planned surgery, and radiotherapy all may interfere with blood supply to the head and neck, resulting in local and systemic problems. Irradiation alone or combined with surgery result in a 22 to 30 percent incidence of clinical hypothyroidism, particularly in laryngeal surgery. Hypoparathyroidism, transient or permanent, may result in up to 10 percent of cases of thyroidectomy and must be considered after laryngopharyngectomy. Obstruction of one or both jugular veins, particularly when combined with lymphadenectomy, results in lymphedema of the face and may result in intracerebral edema, particularly if excessive fluid is administered during surgery. Head elevation, diuretics, and judicious fluid management, however, usually allow collateral flow through the vertebral veins to resolve these problems.

Surgical misadventure, poor planning or execution, or the presence of infection will result in technical complications. Respiratory problems can result from pneumothorax precipitated while operating in the mediastinum or supraclavicular fossa. Hematoma may cause acute upper airway obstruction. Tracheostomy may cause subcutaneous emphysema, tracheoinnominate fistula, or subglottic stenosis. The combination of infection and local ischemia of skin or mucosa may result in wound infection, suture line breakdown, flap necrosis, osteomyelitis, and osteoradionecrosis. Exposure of a previously irradiated carotid artery usually results in a bacterial infection and rupture and must be treated as a surgical emergency. Careful planning, meticulous attention to watertight closure of the pharynx and oral cavity, and provision of adequate independently well-vascularized tissue for reconstruction will minimize the likelihood of *technical* complications.

Restoration of function is one of the main goals of head and neck cancer therapy. Although some dysfunction is inherent in all therapy, proper selection of reconstructive methods, attention to intrinsic function of the structures resected and those remaining, and careful and patient postoperative rehabilitation will decrease the severity of the common functional complications: chronic airway obstruction, aspiration pneumonia, dysphagia, dysphonia, and mental depression.

The potential for catastrophe is great in head and neck cancer surgery. Recognition of a complication and realization of the likelihood of rapid deterioration are important to prevent *catastrophic* complications. Tetany from hypoparathyroidism, acute airway obstruction from hematoma or a dislodged tracheostomy tube, tracheoinnominate fistula, and carotid hemorrhage all can lead to rapid death. Any complication that occurs in a patient who has been previously irradiated must be aggressively resolved. Carotid artery exposure, oropharyngocutaneous fistula, or skin flap necrosis in the irradiated neck could result in subsequent invasive infection of the great vessels and death (Fig. 15-25). Oropharyngocutaneous fistula is relatively common, ranging from 6 to 38 percent of head and neck cancer cases, and can be treated expectantly if salivary flow can be diverted and the carotid arteries protected by well-vascularized tissue. The pectoralis major musculocutaneous flap is a useful method of closing major fistulas and covering the great vessels at the same time. Pharyngocutaneous fistulas of moderate size may be treated by the sternocleidomastoid musculocutaneous flap (Fig. 15-26).

Oral Cavity

Anatomy and Physiology

The anatomic borders of the oral cavity are the mucosal surfaces of the lip externally and the anterior tonsillar pillar posteriorly. The oral cavity usually is considered as a number of distinct entities: the lips, the buccal mucosa, the gums (mandibular and maxillary, including the retromolar trigone), the floor of the mouth, the mobile tongue, and the hard palate. Thus the oral cavity is the aditus of the long seromuscular-mucosal tubular conduit of food and liquid that allows the organism to obtain nourishment. As such it is modified in several ways to facilitate the initiation of alimentation. The lips are a sphincter that allows the oral cavity to be sealed after intake of food and liquid. The buccal surfaces and the buccinator muscles help to collapse and expand the oral cavity to facilitate passage of food back to the pharynx. The hard palate provides a stable platform against which the mobile tongue can push and separates the oral and nasal cavities. The floor of the mouth and the gums are structural components of the reservoir and aid in the preparation functions of the oral cavity in the process of eating; the mobile tongue is the main propulsive agent in the oral cavity. Each site also has a distinct contribution to the modulation of air expelled from the lungs that results in speech.

Therapy

Carcinogenesis in the oral cavity and the natural history of subsequent disease are generally similar independent of anatomic area, although TNM staging has not been a perfect prognosticator for all sites, and histology and site play an important role in outcome. The consequences to the patient of therapy, particularly surgical therapy, however, are very different and depend very much on the function of the area involved and the success with which it can be reconstructed.

Lip Etiology

In the United States and Canada, carcinoma of the lip is common, with a marked male predominance of 20:1. The lower lip is by far the most common site, being involved in approximately 95 percent of cases, and has squamous cell carcinoma as the most common histology. Basal cell carcinoma predominates in the upper lip. The habit of pipe smoking, with chronic thermal injury, was for many years believed to be the carcinogenic stimulus. Recently, however, it has become clear that the protuberant lower lip is exposed to higher doses of ultraviolet radiation, resulting in malignancy that behaves similarly to UV-induced skin cancer. Farmers and other outdoor workers in their sixth to ninth decades, of Celtic and northern European origin, who live in areas of high sunlight exposure are at highest risk. Chronic exposure to sun results in loss of the anatomic vermilion border, the junction between the skin and mucosa of the lip, followed by leukoplakia or carcinoma in situ and by invasive malignancy (Fig. 15-27).

Pathology

Well-differentiated stage I lesions comprise the majority, 60 to 80 percent, of most clinical series. Local recurrence after surgery or irradiation occurs in 10 to 20 percent of cases, but salvage therapy (subsequent surgery with or without radiation or chemotherapy) may be successful. Nodal metastases, usually to the submental or submandibular nodes, are present in 10 to 15 percent of cases and occur in tumors of all histologic differentiation. Treatment by local excision or radiation results in cure rates of approximately 90 percent for stage I disease and 55 to 80 percent in stage II locoregional disease. Aggressive carcinoma of the lip seems to follow a pattern of perineural invasion, down the mental nerve to involve the mandible and pterygoid space, and ultimately via the trigeminal nerve to the base of the skull.

Therapy

Surgical therapy requires resection of the disease with a clear margin of normal tissue around it. If lymph nodes are involved, ipsilateral or bilateral neck dissection is indicated. Elective or prophylactic node dissection for the patient with the N0 neck usually is not recommended in epidermoid carcinoma of the lip. The functional goal in lip reconstruction is restoration of oral continence and reasonable appearance. Because the lower lip is longer and more protuberant than the upper lip, primary closure of defects of approximately 25 percent of the upper lip and 35 percent of the lower lip will result in satisfactory appearance and function. The lip opposite the resection is an important donor site for larger defects. Cross-lip flaps from either the upper or lower lip of the Abbé or Estlander type and advancement of lateral labial and buccal elements are useful, depending on the size and site of the defect. Preservation of sensory and motor function of the remaining orbicularis oris muscle also is possible, to maintain the most efficient sphincter mechanism, using the Karapandzic principle. When the entire lower lip is resected, the damming function can be restored by a radial forearm free flap including palmaris longus tendon, which is inset to the adjacent oral musculature to serve as a dynamic sling. Since the tissue at risk is oral mucosa on the protuberant area of the lip, preneoplastic changes of leukoplakia or dysplasia should be treated by mucosal resection, vermilionectomy (lip shave), and advancement of the labial mucosa to the sun-exposed margin of the skin (mucosal advancement).

Buccal Mucosa Anatomy and Physiology

The buccal mucosa extends from the commissures of the lips to the pterygomandibular raphe and from the maxillary (upper) to the mandibular (lower) alveoli on both sides. The subjacent structures include the buccal fat pad and buccinator muscle, and the surface of the buccal mucosa permits entry to the oral cavity of Stensen's duct from the parotid gland. This area modulates speech and oral capacitance.

Pathology

Buccal mucosa cancer comprises about 5 percent of all oral cancers, and, as with other sites, there is a significant male predominance (3:1). There is a high incidence of advanced disease on presentation, with 18 percent stage I, 36 percent stage II, and 44 percent stage III, with a 56 percent incidence of nodal metastasis. Large cancers of the buccal mucosa are less likely than tumors of other oral sites to have subclinical metastases to the neck. A subset of lesions arising in the buccal mucosa is *verrucous carcinoma*, which presents as an exophytic mass that has the cellular histology characteristic of malignancy but lacks the invasive aspects. Verrucous carcinoma is more common in females than in males and may be related to human papillomavirus. The buccal mucosa is the most common site of this variant of squamous cell carcinoma, which shows warty dense keratinization, sharply circumscribed deep margins, pushing borders, and inflammatory infiltrate. There is a high incidence of multicentricity of malignancy in patients with verrucous carcinoma, with up to 40 percent having other sites of invasive carcinoma, so it probably represents a part of the spectrum of field cancerization (see Fig. 15-10 E). There is some concern that radiation of this lesion may result in dedifferentiation or change to a more malignant histology, but this observation may be a manifestation of tumor cell heterogeneity and clonal resistance to radiotherapy rather than actual malignant degeneration. Carcinoma of the buccal mucosa, both verrucous and infiltrative, occurs commonly in chronic tobacco chewers and snuff dippers in the United States and in people who use *pan* in India and Southeast Asia.

Therapy

Surgical resection with or without adjuvant radiotherapy results in survival rates of 50 to 60 percent, with 60 to 75 percent for localized disease and 25 to 45 percent for locoregional disease. The route of invasion of epidermoid carcinoma of the buccal mucosa is through the buccinator muscle and buccal fat pad dorsally toward the pterygoid musculature or lateral to the skin. In either case significant limitation of oral motion and discomfort on chewing (trismus) occurs.

Surgical resection frequently creates a full-thickness defect of mucosa, muscle, and skin with or without adjacent mandible or maxillary tuberosity. For extensive lesions, restoration of function requires replacement of internal lining as well as external skin coverage. Although the forehead flap based on the superficial temporal vessels and the deltopectoral flap from the shoulder were the standard in the 1960s and 1970s, recent methods have included the combined use of pectoralis major musculocutaneous flap for lining and deltopectoral skin flap for skin coverage or the latissimus dorsi musculocutaneous flap folded on itself to provide both internal and external surfaces (Fig. 15-28). Fasciocutaneous flaps such as the scapula and radial forearm can provide ample tissue for both defects and vascularized bone as well when transferred as a microvascular free flap. For smaller defects, intraoral flaps of mucosa and muscle such as tongue flaps, palate mucoperiosteal flaps, or advancement flaps have been useful. Split-thickness skin grafts, though successful in the short term, ultimately result in fibrosis and difficulty with chewing. For superficial lesions that extend over the mandibular alveolus and require supple coverage, the platysma musculocutaneous flap is an excellent choice (Fig. 15-29).

Hard Palate Pathology

The roof of the mouth, bounded by the soft palate posteriorly and the teeth anteriorly and laterally, is not a common site of intraoral squamous cell carcinoma. More common in this area are tumors, both benign and malignant, of the minor salivary glands (Fig. 15-30). Squamous cell carcinoma in the United States usually is a disease of elderly male smokers and remains superficial for prolonged periods before extending through periosteum and bone and spreading either cephalad into the nasal cavity or maxillary antrum or dorsally through the pterygopalatine fossa. Epidermoid carcinoma of the hard palate is more common in India and Venezuela, where reverse smoking (with the lighted end of the cigarette inside the mouth) is practiced.

Therapy

Treatment is surgical resection with or without adjuvant radiotherapy. Because of the underlying bone, definitive radiotherapy is rarely useful. Cervical metastases are relatively rare in disease of the hard palate, with only 10 to 25 percent of patients presenting with disease at either the prevascular facial nodes or the jugulodigastric nodes. Occult metastases are rare, so elective neck dissection is not part of the therapeutic regimen. Five-year survival rates range from 33 to 75 percent, depending on stage, with an average of 55 to 60 percent for all patients.

Small to moderate-sized defects of the hard palate are best treated with a dental prosthesis in both the dentate and edentulous patient; massive defects may require temporalis muscle flap or local flaps and skin grafting or free tissue transfer, since support for the prosthesis may not be available. The inflammatory process of the minor salivary glands of the hard palate, necrotizing sialometaplasia, which produces an ulcerative lesion with erythematous borders similar to squamous cell carcinoma, may be confused with malignancy. Biopsy examination, however, will show no evidence of malignancy, and the disease is self-limited, usually requiring no therapy.

Floor of Mouth Anatomy and Physiology

The floor of the mouth is the horseshoe-shaped area between the mobile tongue and the lingual surface of the mandible. The papillae that allow Wharton's ducts to empty into the oral cavity lie at the anterior border of this area; posteriorly, the floor of the mouth blends into the glossopalatine fold and the retromolar trigone. In this natural reservoir there may be prolonged contact of the floor of mouth mucosa with carcinogenic agents dissolved in the saliva after oral ingestion or inhalation. This area provides capacity, allowing the tongue to sit low in the mouth, thus increasing the volume of the oral cavity and preventing obstruction of the direct route between the lips and the pharynx.

Pathology

Approximately 13 to 17 percent of oral lesions arise in this area, the third most common site after the lip and mobile tongue. Because of the proximity of the mucosa to the hyoglossus and mylohyoid muscles of the submandibular triangle, and because of the rich lymphatic supply, direct extension of tumor into the neck and bilateral cervical metastases are frequent, especially in anteriorly located lesions. Medial growth at the primary site also invades the ventral surface of the tongue, and lateral growth invades the mandible. Advanced-stage disease is common, with 46 to 52 percent of patients presenting with stage III or stage IV disease. Subclinical disease in the neck or micrometastases in the clinically negative neck are common, with overall neck involvement increasing with primary tumor size and ranging from 15 percent for tumors less than 2 cm (T1) to 50 percent for tumors greater than 4 cm (T3).

Therapy

Because the tumor may abut or actively invade the mandible, resection with an adequate margin of normal tissue in this area frequently requires removal of the mandibular periosteum or actual resection of a segment of mandible. Uncertainty about invasion of the mandible or proximity may require removal of the medial cortex (marginal mandibulectomy) or a segment of the mandible. Because of the proximity of the mandible, therapeutic doses of radiation may result in ischemic necrosis of the bone, osteoradionecrosis. Combined therapy, surgical resection of the primary tumor with neck dissection en bloc or pull-through resection followed by adjuvant radiotherapy, is appropriate for advanced disease. Survival rates are strongly dependent on stage, ranging from 68 to 91 percent for stage I to 35 to 46 percent for stage III disease. Recurrence at the primary site frequently involves the mandible or the suprahyoid complex of muscles and may require laryngectomy because of invasion of the preepiglottic space.

Reconstruction of the floor of the mouth presents a number of challenges. The watertight seal of the oral cavity, the continuity of the mandibular arch, and the mobility of the tongue all depend on this area. For superficial lesions with normal well-vascularized muscle beneath them, split-thickness skin grafts have been advocated. Although they may be successful initially in covering the area, the contracture inherent in this method may result in tethering of the mobile tongue to the subjacent muscle or to the mandibular periosteum. If the neck is not involved by tumor, a musculocutaneous island flap of platysma muscle based on the facial artery provides supple skin coverage. Bulky thoracic musculocutaneous flaps, such as pectoralis major, provide skin for lining but may be compressed by the intact mandible, resulting in ischemic necrosis, or may push the tongue back in the oral cavity, limiting movement and obstructing the pharynx and airway. The thin pliable skin of the volar forearm, based on the radial artery and its venae comitantes as the radial forearm free flap, provides an excellent lining for this area that can drape over the mandible and allows free movement of the tongue while providing a watertight seal. Vascularized segments of bone and skin of varying size also are available as free tissue transfers from the fibula, scapula, dorsum of the foot, and other sites with high success rates if the mandible must be resected (Fig. 15-31).

Gums, Gingivae, Alveolar Ridge Pathology

Squamous cell carcinoma arising on the gums constitutes 10 to 17 percent of oral cavity malignancies. Eighty percent of lesions arise on the mandibular alveolus. Although the male predominance persists at this site, there seems to be a less direct relationship with tobacco and alcohol, and the causation may be related in some cases to chronic trauma from poorly fashioned dentures or jagged teeth. The thin layer of mucosa allows invasion of the underlying mandibular or maxillary bone in 35 to 50 percent of cases. Direct invasion through the periosteum is most common in patients with teeth, but spread through the empty sockets along the occlusal ridge and subsequent perineural invasion are common in edentulous patients. Cervical metastases occur in 30 to 45 percent of the cases, depending largely on the size of the primary.

Therapy

Surgical resection requires removal of the subjacent bone by partial or total maxillectomy or total mandibulectomy. Maxillary defects of moderate size can be treated with dental prostheses, and mandibular defects reconstructed with various combinations of skin and bone as previously described.

Oral Tongue Anatomy and Physiology

The tongue is a complex muscular structure covered with mucosa and receiving motor innervation from the hypoglossal nerve. The bulk of the tongue is made up of the superior and inferior longitudinal muscles joined by the vertical and transverse intrinsic muscles. Inferoposteriorly and laterally the tongue is connected to the hyoid bone by the hyoglossus muscle, and superiorly and anteriorly to the mandible by the genioglossus muscles. The styloglossus and palatoglossus muscles attach the tongue superiorly to the base of the skull. These junctions allow the tongue great mobility, promoting synergy with the larynx and the pharyngeal and palatine muscles. Beneath the smooth ventral surface of the tongue, along the floor of the mouth, are the numerous openings of the sublingual ducts. On the dorsal surface are papillae, with specialized sensory organs for taste at the base. Also opening onto the tongue are the ducts of the minor salivary glands. Sensation is provided to the tongue by the lingual nerve carrying fibers of cranial nerve V, and taste by the glossopharyngeal and chorda tympani of VII. Differentiation of sweet, sour, salt, and bitter taste relies a great deal on intact function of the tongue. As a mucosa-covered muscle, the tongue is the major propulsive force in the oral cavity. It initiates and continues movement of the food bolus to the pharynx, generating pressure of up to 120 mmHg. Numerous investigators have demonstrated that dysfunction after surgery for oral cancer depends almost exclusively on the amount of tongue resected.

Pathology

In most series of oral cancers, the oral or mobile tongue is second only to the lip as the most common primary site. Tobacco and alcohol are the most common associated factors, but chronic irritation from jagged teeth or dental appliances also may be involved. Although the sixth and seventh decades are the peak periods, sporadic occurrence of squamous cell carcinoma in the tongue has been described in patients under thirty and in renal transplant and other immunosuppressed patients and may not be linked to the usual carcinogenic stimuli. In India, submucosal fibrosis seems to be a predisposing influence, and in Scandinavia, Plummer-Vinson syndrome, glossitis, iron-deficiency anemia, and achlorhydria may be related.

Malignancy of the mobile tongue occurs most frequently at the midportion of the lateral tongue and is frequently asymptomatic (Fig. 15-32). Radial spread through the tongue may extend submucosally to the base of the tongue and across the midline or laterally to the floor of the mouth. Because of the rich lymphatic supply, ipsilateral metastases are common to the submandibular and submental nodes. Clinical evidence of cervical metastasis is present in 40 to 61 percent of patients, and subclinical disease in the N0 neck is found in 25 to 31 percent. As in other sites, the presence of lymph node metastases seems to be the most important prognosticator, with survival rates of 73 to 92 percent for localized disease (T1, T2, N0) and only 31 to 45 percent with regional metastasis.

Therapy

Definitive therapy for carcinoma of the oral tongue can be attempted with either external beam radiotherapy or interstitial radiotherapy. External radiation in doses to 65Gy may be useful, but with the implantation of afterloading devices (tubes into which iridium-192 or radium needles can be placed) doses in the range of 100Gy to 150Gy can be delivered over a small area with greater effect.

The surgical therapy of carcinoma of the tongue consists of resection of the tumor with a margin of normal tissue and en bloc removal of the regional lymph nodes. Unfortunately, evaluation of the extent of local disease is difficult in tongue cancer. Although several authors have demonstrated respectable 5-year survival rates (48 to 62 percent) with partial glossectomy with or without en bloc or discontinuous neck dissection, a generally accepted uneasiness about the ability to obtain clear margins has led to the common use of adjuvant radiotherapy. Hemiglossectomy, or resection to the median raphe, has been advocated by some for lesions involving any part of the lateral tongue.

Extensive lesions of the tongue may extend posteriorly to involve the larynx. Even in those patients who do not demonstrate invasion of the larynx, widespread involvement of the tongue or resection of the base of the tongue may predispose the patient to aspiration and ultimate respiratory failure. Despite skepticism on the part of some surgeons, total glossectomy with or without laryngectomy has been shown to be a valuable procedure for both cure and palliation. A 3-year survival of 53 percent has been achieved in one series, with 80 percent of patients demonstrating intelligible speech if the larynx is preserved and 93 percent regaining the ability to maintain their nutritional status by oral alimentation.

Although attempts have been made to innervate muscle of various types transplanted to replace the tongue, there is no satisfactory way to reconstruct the tongue. Denervation of the tongue by resection of, or injury to, both hypoglossal nerves usually renders the patient incapable of swallowing or of effective speech. After surgical resection of a portion of the tongue, the reconstructive goal is to allow free mobility of the remaining tongue while providing a watertight seal to the oral cavity. If the floor of the mouth is not involved in the resection, simply skin grafting of the raw surface may suffice. Suturing the edge of the resected tongue to alveolar or buccal mucosa usually tethers the tongue, impeding its mobility. Advancing the posterior mobile tongue or setting back the excess anterior tongue may provide the optimal solution. Although the pectoralis major flap has been advocated for intraoral reconstruction, its bulk tends to push the tongue back or pull it down into the neck, interfering with its motion and the subsequent elevation of the larynx necessary for effective swallowing and speech. Furthermore, its thickness and weight effectively fix the tongue to the adjacent mandible, further interfering with its motion. The lateral arm or radial forearm free flap provides lightweight supple tissue more appropriate for restoring tongue, floor of mouth, and mandibular alveolar epithelial lining.

The defect of total glossectomy consists of the tongue, floor of the mouth, and sometimes pharyngeal and laryngeal mucosa. Restoration of oral continence usually requires significant amounts of soft tissue. The pectoralis major flap serves well to replace the entire floor of the mouth, as does the jejunal free flap, which also can replace the pharynx and cervical esophagus. If the larynx is preserved in total glossectomy, the radial forearm free flap serves as an excellent diaphragm to pull the hyoid bone anteriorly toward the mandible and assist in swallowing.

When a portion of mandible must be resected for carcinoma of the oral tongue, the urgency of reconstruction depends on what part of the mandible has been resected. Although any mandibulectomy results in some dysfunction, partial mandibulectomy lateral to the mental foramen usually is well tolerated, the main morbidity being weakness of chewing and malocclusion of the remaining teeth. Resection of the symphysis or anterior segment of the mandible, however, is a much more devastating problem and requires immediate reconstruction. Vascularized bone from scapula, fibula, iliac crest, radius, or metatarsal is excellent for reconstruction. If appropriate soft tissue is not available, two free tissue transfers can be performed to satisfy the individual needs of the wound.

Pharynx

Anatomy and Physiology

The pharynx is the continuation of the muscular tube that constitutes the alimentary tract. It is anatomically divided into three sections, each with a slightly different function: the nasopharynx, the oropharynx, and the hypopharynx (see Fig. 15-4). An important role of the pharynx is separating the respiratory and the alimentary tracts, and its specialized structures reflect this function. The nasopharynx is unique in the pharynx in that it is a rigid cavity bounded on three sides by bone—superiorly by the base of the skull and posterior sphenoid sinus, anteriorly by the posterior rim of the ethmoid plate and the choanae, passages from the nasal cavity into the nasopharynx, and posteriorly by the pharyngeal tubercle of the occipital bone and the atlas and axis, with their prevertebral fascial and muscular coverings. The inferior surface is the nasal side of the soft palate. The lateral sides of the nasopharynx give entry to the eustachian tubes (to decompress the middle ear), and the roof is the site of a collection of lymphoid tissue, the pharyngeal tonsil.

The oropharynx is the muscular tube that serves as transit area from the propulsive oral cavity and the recipient nasal cavity to the alimentary and respiratory tracts. This mucosa-lined muscular tube contains the base of the tongue (from the circumvallate papillae back), the tonsils, the oral soft palate, the lateral pharyngeal walls, and the posterior pharyngeal wall. The dominant muscular entity that receives the propulsive energy of the tongue is the superior constrictor muscle, attached on both sides to the pterygomandibular raphe and wrapping 270 degrees to constitute the posterior and lateral walls. Contraction of this muscle closes the palatopharyngeal sphincter, or Passavant's ridge, elevating the palate. This action closes the nasopharynx and pushes the bolus of food into the hypopharynx. Up-and-down motion of the palate is regulated by the tensor veli palatini and levator veli palatini muscles (attached to the base of the skull) and the palatopharyngeus muscle (attached to the lateral pharyngeal wall). Lack of synergy in these muscles is seen in patients with cerebrovascular accidents and hypoxia neonatorum and markedly interferes with speech and swallowing.

The anatomic boundaries of the hypopharynx are reflections of the anatomy of the larynx. The posterior pharyngeal wall runs from the tip of the epiglottis to the inferior border of the cricoid cartilage. The anterior border is the postcricoid mucosa, and the lateral surfaces are the mucosal cavities on both sides of the larynx known as the pyriform sinuses. The middle pharyngeal constrictor muscle, with its attachments to the hyoid bone and prevertebral fascia, and the inferior pharyngeal constrictor, with its distal condensation of the

cricopharyngeus muscle attached to the lateral surfaces of the thyroid and cricoid cartilages, serve as the pharyngeal sphincters. When they contract they close off the entrance of the cervical esophagus to air and direct it through the larynx. When they relax they allow food through the pharynx and into the cervical esophagus. The inferior constrictor and cricopharyngeus muscles serve as the upper esophageal sphincter and may become hypertonic, a condition known as cricopharyngeus spasm. Zenker's diverticulum, a lateral outpouching of the pharyngeal wall that may collect undigested food and result in chronic aspiration, is the consequence of chronic hypertension of the upper esophageal sphincter (Fig. 15-33). Resolution of this upper esophageal sphincter hypertension has been obtained temporarily by injection of botulinum toxin, but permanent treatment requires transection of the muscle (cricopharyngeus myotomy) with or without resection and closure of the mucosal diverticulum.

Base of Tongue Pathology

Carcinoma arising behind the circumvallate papillae in the base of the tongue frequently remains asymptomatic and undiagnosed until late-stage disease has developed. Even when patients complain of pain, either local or referred, as with otalgia, difficulty in examination or reluctance of primary care physicians to perform indirect laryngoscopy and palpation of the base of the tongue results in misdiagnosis and prolonged therapy for pharyngitis, tonsillitis, and other less serious problems. The central location gives rise to cervical lymphatic metastases in up to 70 percent of patients, and there are bilateral metastases in 17 to 25 percent of cases.

Histology and gross morphology in this region predict to some degree the behavior of the lesion and the appropriate therapy. In addition to epidermoid carcinoma, minor salivary gland lesions are also seen. Exophytic lesions, with cells resembling lymphocytes and absence of keratin pearls, arise in the tissues of Waldeyer's ring, the tonsils (lingual and palatine), and the base of the tongue. These lymphoepitheliomas behave like nasopharyngeal carcinoma and have been characterized as undifferentiated carcinomas with lymphocytic infiltration. Such lesions are more radiosensitive, both at the primary site and as cervical metastases, than most other infiltrative keratin-producing squamous cell carcinomas, with 2-year local control rates of 75 percent for T1 lesions and 67 percent for T2 lesions.

Carcinoma arising in the base of the tongue spreads anteriorly into the oral tongue, superiorly up to the tonsillar pillar, and inferiorly into the lateral pharyngeal wall and into the vallecula, preepiglottic space, and larynx. Because of the proximity and functional relationship of the base of the tongue to the larynx, interference with laryngeal elevation and closure of the epiglottis, with the attendant aspiration pneumonitis, is a hallmark of carcinoma of the base of the tongue.

Therapy

Advanced disease at the primary site or disease with cervical metastases requires surgical therapy. If the lesion is lateral enough, partial glossectomy may be adequate. Since resection of the base of the tongue usually removes the hypoglossal nerve to the tongue in that area, subtotal or posterior glossectomy is unlikely to leave functional tongue. Radical resection may require total glossectomy with or without laryngectomy (Fig. 15-34). In most patients, if the oncologic requirements of the resection do not dictate removal of the larynx, reconstruction with the larynx in situ is appropriate. Surgical therapy or surgery combined with postoperative adjuvant radiotherapy results in 5-year survival rates of 50 to 60 percent for stage III disease and 20 to 25 percent for stage IV disease. Advanced primary disease or disease in the neck is unlikely to be successfully controlled locally with radiotherapy alone, and surgical salvage after radiotherapy has been dismal, with a high incidence of osteoradionecrosis of the mandible.

Reconstruction of defects arising from resection of the base of the tongue should effectively close the pharynx and oral cavity with tissue that will heal primarily and withstand subsequent irradiation and still not interfere with the function of tissues left intact by the curative resection. Elevation of the larynx, the normal motion that occurs during the early pharyngeal phase of swallowing to close the epiglottis, cannot occur if the tongue is tethered to the side of the pharynx or oral cavity or if bulky tissue such as the pectoralis major flap is interposed into the area of the resection. The provision of sensate tissue into the area surrounding the larynx, to prevent aspiration, is also an important consideration. Occasionally local tissue can be mobilized for closure without tension to avoid fistula formation and to provide sensate mucosa. More commonly, however, mucosa or skin to line the tongue, mandible, and pharyngeal wall is necessary. Buccal and palatal flaps have been described for small and moderate-sized defects, but free tissue transfer of skin from radial forearm or lateral arm or lateral thigh flaps are more appropriate for the more common extensive defects. Neurotized free tissue transfer reconstructions of the pharynx and oral cavity have shown two-point discrimination and other measures of sensitivity that are characteristic of the recipient site (upper aerodigestive tract) rather than the donor site (arm, back), suggesting that cerebral cortical integration of the reconstruction favors the recipient site.

Tonsil Pathology

Squamous cell carcinoma of the tonsil may arise in the tonsil, the tonsillar bed, or the tonsillar pillars (Fig. 15-35). As one of the Waldeyer's ring structures, the tonsil shows a higher incidence of lymphoepithelioma than other sites. Whatever the histology, it is second only to the larynx in frequency as a site of upper aerodigestive tract primary malignancy, with 12,000 new cases per year in the United States. Like most upper aerodigestive tract tumors, tonsillar carcinoma presents with predominately late-stage disease, with 28 to 32 percent of patients presenting with stage I or stage II disease, 35 to 40 percent stage III, and 40 to 45 percent stage IV. Cervical metastases at the time of presentation are seen in up to 67 percent of patients, and subclinical disease in the N0 neck in 10 percent of patients. Determination of the extent of local disease in advanced tumors is of considerable importance in decision-making and execution of therapy. Growth of the tumor upward into the soft palate occurs in 60 percent of patients, downward to the base of the tongue in 56 percent, into the nasopharynx in 9 percent, and down the lateral pharyngeal wall to the epiglottis in 27 percent of cases. The site of local extension is particularly important, because in treating patients with radiotherapy, geographical misses as a result of underestimating the local extent of disease were a common reason for failure.

Therapy

Carcinoma of the tonsil appears to be more radiosensitive than other primary-site squamous cell carcinomas. The usual approach to disease originating at this site is to treat for curative intent with radiotherapy ranging from 55Gy to 70Gy to the primary site and bilateral cervical lymph node drainage areas (Fig. 15-36). If there is bulky neck disease or extension of the primary tumor into adjacent bone or pterygoid muscles, surgical resection, reconstruction, and postoperative adjuvant radiotherapy are safer and more effective. Risk of local recurrence is directly related to size of the primary tumor, and recurrence is even more likely when the predominant spread of tumor was into the base of the tongue.

The challenge of reconstruction in tonsillar disease is a function of the dimensions of local growth of the lesion and the history of previous radiation. The soft-tissue defect in the lateral wall of the pharynx created by a superficial lesion can easily be resurfaced with a skin graft, a deltopectoral flap, or a fasciocutaneous free flap. Radial spread up to the soft palate or down into the pharynx demands that the surface of the flap be contoured in several planes without obstructing the pharynx, usually requiring a free tissue transfer of relatively pliable skin, such as the radial forearm, lateral arm, or lateral thigh. Invasion of the ascending ramus of the mandible can be treated by addressing the soft-tissue defect alone by transposing a pectoralis major musculocutaneous flap into the lateral pharynx, using the skin for internal lining and the muscle for coverage of the carotid vessels in the neck. In the patient with teeth, soft-tissue reconstruction alone will lead to malocclusion of the dental arches and other problems. Skin and bone together are available with a number of methods. The scapula and the deep circumflex iliac artery free flaps are particularly suited to the lateral defect created by radical resection of the tonsil and mandible.

Soft Palate Anatomy and Physiology

The soft palate is the mucosa-lined fusion of the inferiorly based palatopharyngeus and palatoglossus muscles and the superiorly based levator and tensor veli palatini muscles. Extending backward from the posterior margin of the palatine bone, directly opposite the superior constrictor muscle and its bulge, which creates Passavant's ridge, its main function is to open and close the passageway between the nasal cavity and nasopharynx and the oral cavity and oropharynx. This obturating and modulating effect has obvious importance in speech and in swallowing.

Pathology

Isolated carcinoma of the soft palate is rare, the disease usually occurring in combination with other frank malignancies or premalignant entities such as leukoplakia or erythroplakia and not showing the usual male predominance.

The oral side of the soft palate is by far the most common site for malignancy, which usually extends down the tonsillar pillars to the base of the tongue. Superior and posterolateral spread to the nasopharynx and posterior pharyngeal wall is less common. Ipsilateral cervical lymph node metastases to the jugulodigastric nodes occur in 40 to 50 percent of cases, and, because tumors often pass over the midline, bilateral metastases are seen in 15 percent of cases.

Therapy

Treatment of soft palate carcinoma follows the usual principle that small primary lesions are effectively eradicated by radiotherapy in the range of 65Gy and that combined therapy, surgery with adjuvant radiotherapy, is necessary for best results in larger tumors. The importance of radiotherapy as a treatment was particularly emphasized in the past because of the inability to reconstruct the soft palate and the devastating functional result. Nonrandomized studies of surgery and radiotherapy have shown 5-year survival rates of 31 to 44 percent, with the size of the primary lesion, the absence of a synchronous primary upper aerodigestive tract tumor, the absence of cervical metastases, and moderately to well differentiated histology being favorable prognostic features.

Resection or dysfunction of the soft palate results in escape of air and oral contents into the nasal cavity and ultimately out the nares. Advances in prosthetic technology have made the use of a dental prosthesis a possible solution in some cases. For defects that do not include the lateral and posterior pharyngeal wall, a superiorly based flap of pharyngeal mucosa and muscle sutured into the margin of the palatal resection will obturate the opening between the nasopharynx and oropharynx. More extensive defects require the introduction of epithelium-lined soft tissue; the lateral arm, lateral thigh, and radial forearm free flaps are ideal when soft tissue alone is needed, and the scapula free flap when both skin and bone are necessary (Fig. 15-37).

Posterior Pharyngeal Wall Anatomy

The posterior and lateral pharyngeal walls extend from the oropharynx down into the hypopharynx, where the lateral pharyngeal walls end as the lateral walls of the pyriform sinus and the posterior wall extends to the cervical esophagus. The lymphatic drainage of the posterior pharyngeal walls is to the jugulodigastric, midjugular, and juguloomohyoid nodes but also directly to the retropharyngeal group of lymph nodes.

Pathology

Because of its location and the nonspecific symptoms of mild dysphagia andodynophagia, squamous cell carcinoma of the posterior pharynx usually is detected at a late stage, with 39 to 55 percent of patients presenting with palpable cervical metastases. Local spread of disease is cephalad toward the nasopharynx and lateral to the lateral pharyngeal walls and larynx.

Therapy

Using combinations of surgery and radiotherapy, 3- to 5-year survival rates of 25 to 32 percent overall have been reported.

The location of the posterior pharyngeal wall as the farthest border of the aerodigestive tract and its proximity to the larynx have presented some problems both of access for surgical resection and of potential for appropriate reconstruction. Accurate resection of local disease requires visual access. Midline division of the lip, mandible, and tongue, the

median labiomandibular glossectomy, allows visualization of the posterior pharynx. If disease is limited to the posterior wall, the surface can be relined with a split-thickness skin graft or allowed to epithelialize. Circumferential disease requires more complete pharyngeal resection and sometimes pharyngolaryngectomy. Reconstruction of the complete or partial circumferential defect may be accomplished with a free autograft of jejunum or a radial forearm free flap. If the larynx can be preserved, aspiration will be less likely if free mobility of the larynx is maintained and sensate epithelium is restored, either by sensory-innervated free tissue transfer or by skin grafting.

Hypopharynx Pathology

In addition to the posterior pharynx, the hypopharynx contains the pyriform sinuses and the postcricoid area. Tumor growth in this area, the lateral and posterior mucosal border of the larynx, is intimately related to the function of the larynx, a facet recognized in the staging systems used in assessing the extent of hypopharyngeal disease (Fig. 15-38). Thus the T stage of the pyriform sinus lesion increases not with tumor size but with extent into the medial wall or with fixation of the vocal cord caused by direct extension of disease. Another consideration in hypopharyngeal carcinoma, as in disease of the cervical esophagus, is the problem of submucosal extension of tumor. Spread of disease into the cervical esophagus discovered at surgery may simply represent clinical understaging of disease, but it also could be multifocal disease or field cancerization. Local spread of disease is cephalad toward the nasopharynx or distal into the cervical esophagus, as well as medial and lateral toward the larynx. Advanced-stage local disease is common, with only 10 to 15 percent of cases confined to only one site in the hypopharynx.

The lymphatic drainage of the area is copious, with primary nodal stations in the midjugular, juguloomohyoid, and retropharyngeal lymph node chains. Even with small lesions there is a likelihood of lymph node metastasis, with 55 to 64 percent of patients presenting with palpable lymphadenopathy, and 41 percent of patients with clinically negative N0 necks demonstrating metastatic disease after elective neck dissection. Even with small primary lesions (T1) localized to one part of the hypopharynx, the risk of cervical micrometastases in the clinically negative neck is high (40 percent). Distant metastases at presentation and with treatment of disease appear to be more common than at other primary sites, occurring in up to 47 percent of the cases.

Therapy

Since it is unusual for hypopharyngeal lesions to present at an early stage, the treatment usually is combined, consisting of surgery followed by adjuvant radiotherapy. The extent of the operation depends on the proximity to the larynx, and the laryngopharyngectomy with bilateral modified neck dissection is the procedure that is most frequently necessary. With such an approach, survival rates of 20 to 40 percent have been achieved.

When the larynx can be saved, primary closure of the surgical defect is the most effective method. Preservation of the superior laryngeal nerves, the sensory innervation to the area, is an important consideration to allow swallowing to proceed without aspiration. When the larynx is removed with the hypopharyngeal lesion, however, the goals of reconstructive surgery are rather simple restoration of alimentary continuity with the least likelihood of fistula formation or other devastating problems and with the ability to restore esophageal speech. Primary closure of the pharyngeal mucosa after partial laryngopharyngectomy for pyriform sinus lesions results in a high likelihood of fistula formation (43 percent) and stenosis (48 to 73 percent), both of which interfere with swallowing and esophageal speech (Fig. 15-39).

Although use of local tissue, including the anterior wall of the larynx and the base of the tongue, and skin of the neck has been described, the extent of surgery, the history of previous radiotherapy, or the likelihood of subsequent adjuvant radiotherapy usually require the importation of distant tissue. The deltopectoral flap from the chest was the mainstay of pharyngeal reconstruction after laryngectomy but is at least a two-stage procedure with a fairly high risk of failure or persistent fistula. The pectoralis major musculocutaneous flap has been advocated for both circumferential and partial defects of the pharynx. The effect of gravity, the inhomogeneous blood supply, and the bulk of the flap result in a high rate of fistula formation and dysphagia, making a free autograft of bowel or skin a preferable method. When a circumferential defect is present, the problem of bulk and gravity can be circumvented somewhat by skin grafting the prevertebral fascia as the posterior wall of the neopharynx (new pharynx) and using the pectoralis major muscle and its overlying skin as a 270- degree reconstruction to complete the pharyngeal conduit.

When total esophagectomy is part of the treatment for carcinoma of the hypopharynx, transposition of the stomach or colon through the thorax can reconstitute the alimentary canal. Gastric pull-up—in which the stomach, based on the right gastric and gastroepiploic vessels, is brought through the chest, and the fundus or cardia sutured to the base of the tongue or pharyngeal remnant—is a fairly reliable technique but has a mortality rate of 10 to 20 percent in most series. Right or left colon interposition usually is reserved for caustic strictures but occasionally is useful in hypopharyngeal or cervical esophageal lesions.

Reconstruction of the circumferential defect resulting from pharyngolaryngectomy by transfer of a free autograft of bowel and revascularization of microvascular anastomosis was first performed in the early 1960s. Although colon, stomach, and jejunum have all been used, the greatest experience has been with jejunal free autograft. Successful reconstruction using the jejunum, either in a tube reconstruction for circumferential defects or as a patch for partial defects, has been achieved in 92 percent of patients treated with this method, allowing 83 percent of the total group to achieve total oral alimentation. Mortality with this method has been less than 5 percent, with few abdominal or thoracic problems. Complications in the neck, though frequent, usually resolved without further surgery, and the segment of bowel was able to withstand radiotherapy without major problems (Fig. 15-40). Although segments of jejunum may be used in the unusual situation in which the larynx has been left in situ, the secretory nature of the mucosa sometimes makes aspiration a problem, and another method may be preferable. In addition to segments of bowel, fasciocutaneous free flaps from the radial forearm, lateral arm, lateral thigh and posterior thigh have been used as either a tube or a patch for hypopharyngeal reconstruction.

Squamous cell carcinoma of the cervical esophagus presents the same reconstructive demands as that of the hypopharynx; when disease is localized, the resected portion esophagus may be replaced with a patch of skin as either a vascularized transposition flap or a free tissue transfer. The mode of spread of the local disease, however, may involve submucosal skip areas and thus require total esophagectomy. The lymphatic drainage of the cervical esophagus is oriented more toward the mediastinum and parapharyngeal nodes than laterally into the neck, and hence a different approach is required for lymphadenectomy.

Nasopharynx Etiology

Carcinoma of the nasopharynx is seen intermittently in the Western world, comprising about 0.25 percent of new cancers in the United States. It is endemic in southeastern Asia, however, particularly in southern Chinese populations such as those originating from Kwantung province and constitutes 21 percent of cancers in Taiwan, 18 percent in Hong Kong, and 14 percent in Indonesia. The incidence is much lower in Asians who have immigrated to North America but still seven times higher than in the white population, suggesting genetic susceptibility to an environmental carcinogen. The sporadic and genetically linked cases of nasopharyngeal carcinoma behave in the same manner. There is a high association of the Epstein-Barr virus with this malignancy.

Pathology

The nasopharynx is a small, mucosa-lined, boxlike cavity at the base of the skull containing the pharyngeal tonsil and the openings of the eustachian tubes and the sphenoid sinus. Tumors arising in this area present local symptoms that vary with their pattern of growth and spread. Lesions that are exophytic and grow out into the cavity may obstruct the eustachian orifices or the choanae, leading to hearing loss (15 percent), nasal stuffiness or obstruction (30 percent), and epistaxis (22 percent). Infiltration and bony erosion of the base of the skull into the cavernous sinus results in cranial nerve palsies in 16 to 25 percent of cases, the most commonly involved being the abducens nerve, followed by the trigeminal nerve and the oculomotor nerve, resulting in paresthesias and diplopia. The most common presenting sign of nasopharyngeal carcinomas is a mass in the neck secondary to cervical metastasis (60 percent). The site of cervical metastasis may be the jugulodigastric nodes or the posterior triangle. Mirror examination of the nasopharynx is an important part of any evaluation of suspicious cervical adenopathy.

Staging of nasopharyngeal carcinoma in the past has been relatively inaccurate because of difficulty with examination. CT has helped to delineate invasion of both the paranasopharyngeal fascial planes and the bony skull in the absence of cranial nerve palsy. Recent staging systems have included histology, multiple symptoms, time from onset of symptoms, location of cervical adenopathy (supraclavicular), and local extent of disease. Variance of histology also has been related to survival and thus is important in prognosis. Keratinizing squamous cell carcinoma has the worst 5-year survival rate at 21 percent; the rate for spindle cell is 41 percent, round cell 51 percent, and mixed nonkeratinizing 54 percent. These histologic predictors of therapeutic response are consistent with the subset of tumors arising as variants from Waldeyer's ring, the lymphoepitheliomas.

Therapy

Despite the frequent presence of cervical lymph node metastases, nasopharyngeal carcinoma is a curable disease. Radiotherapy in doses varying from 50Gy to 84Gy to the primary site with 50Gy to 70Gy to both sides of the neck results in 5-year survival rates varying from 100 percent for stage I disease to 34 percent for stage IV, with overall survival rates ranging from 29 to 49 percent. The total dose of radiotherapy has an effect on survival, with patients receiving lower total doses having poorer survival. Even the presence of skull-base invasion and cranial nerve dysfunction is not a sign of incurable disease. Sixty-two percent of cranial nerve defects can be reversed by radiotherapy, with an overall 31 percent survival rate in this subset of patients. Residual disease at the primary site occasionally may be resected by craniofacial technique, and residual neck disease may be eradicated by radical neck dissection. Distant metastatic disease in nasopharyngeal carcinoma is common, particularly in patients who have bulky cervical metastases. At present, however, adjuvant chemotherapy has not been particularly successful in improving survival.

Nasal Cavity and Paranasal Sinuses

Etiology

The nasal cavity and paranasal sinuses (maxillary, ethmoid, frontal, and sphenoid) are the aditus to the respiratory tract and function to filter impurities from the inspired air, regulate its temperature, and humidify it. As such they are exposed to the many carcinogens in the air, and yet malignancies in these sites are rare. In the Western world they constitute 0.3 percent of all malignancies, with a slightly higher incidence in men than in women. In the Orient this is a more common primary site, comprising 1 percent of new cancers and 23 percent of head and neck cancers in Japan. Endemic areas are found in South Africa, where inspired snuff has a high concentration of nickel. There also may be an increased risk of adenocarcinoma of the paranasal sinuses in woodworkers secondary to inspired wood dust. The disease presents most frequently in the fifth to seventh decades.

Pathology

Squamous cell carcinoma is the most common histology (60 to 80 percent), although minor salivary gland lesions, adenocarcinoma, adenoid cystic carcinoma, and mucoepidermoid carcinoma make up about 20 percent of the tumors. Other malignancies, such as teratocarcinoma, lymphoma, osteogenic sarcoma, schwannoma, fibrous dysplasia, carcinosarcoma, and melanoma, occasionally are encountered. Lymph node metastases are uncommon, with only 15 percent of cases presenting in the neck, since the retropharyngeal lymph nodes are the first station of drainage. There is a questionable relationship between nasal and sinonasal malignancy and the inverting papilloma of the nasal cavity.

Late-stage disease is common, since symptoms usually are rather diffuse, including nasal obstruction (35 percent), local pain (16 percent), epistaxis (12 percent), and cheek swelling (29 percent). Loosening of maxillary teeth or paresthesias also may occur. Diagnosis is made by intranasal biopsy through a speculum or by antrostomy through the lateral nasal wall or labial buccal sulcus (Caldwell-Luc procedure). The maxillary sinus is by far the most common site of origin of the disease (62 percent), followed by the nasal cavity (26 percent), ethmoid sinus (10 percent), and sphenoid sinus (2 percent).

Therapy

Treatment of paranasal sinus tumors is by a combination of radiotherapy and surgery. Radiotherapy alone provides poor palliation and unacceptably low survival rates. Preoperative radiotherapy to 60Gy combined with radical surgery has been the usual approach, resulting in 3-year survival rates ranging from 13 to 32 percent for all sites, 35 to 40 percent for nasal cavity, and 38 to 53 percent for maxillary sinus. Failure of therapy is most commonly manifested by local recurrence (70 percent) and only rarely disease in the neck (5 percent).

Surgical resection involves en bloc removal of the affected sinus and the surrounding involved structures. Total maxillectomy with or without orbital exenteration may be required for adequate clearance. The introduction in 1963 of the intracranial-extracranial approach to tumors, the craniofacial technique, has improved the ability to safely remove tumors of the ethmoid and other paranasal sinuses and has increased 5-year survival rates from the range of 30 percent to 58 percent.

Reconstruction of the postoperative defect that has been previously irradiated is a difficult problem. Unnatural passageways among the paranasal sinuses, oral cavity, nasal cavity, and external environment interfere with the normal flow of air, drying the mucosa and causing discomfort and bleeding as well as interfering with alimentation and speech. Small defects can be obturated with nasal or dental prostheses. Larger defects, however, require three-dimensional reconstruction with free tissue transfer with the intention of restoring pathways of air and ingested substances to as normal a state as possible to improve function and appearance (Fig. 15-41).

Larynx

Anatomy and Physiology

The larynx is a complex, mucosa-lined, bony and cartilaginous box whose function depends on motion coordinated with the adjacent tongue and pharynx. The larynx is divided into three anatomic areas: the supraglottic larynx, from the epiglottis to the ventricle, including the preepiglottic space, hyoid bone, arytenoid processes, and false vocal cords; the glottic larynx, including the true vocal cords and the anterior commissures; and the subglottic area, surrounded by the cricoid cartilage. The different clinical behavior of supraglottic lesions comes in part from their separate embryologic origin. The supraglottis is a derivative of the pharyngobuccal anlage, and malignancy arising in this area behaves similarly to pharyngeal carcinoma, with early metastatic disease to the neck. The glottis arises from the tracheobronchial anlage, and malignancy is more likely to be indolent, with lower incidence of cervical metastases.

The predominant function of the larynx is the modulation of air inspired through the nose and nasopharynx and expired from the lungs. Coordination of respiration with swallowing is a complex function that prevents food from entering the respiratory tree and air from entering the digestive tract. As air enters the pharynx it is shunted to the larynx, which sits in its neutral position with the epiglottis open. Air is then drawn through the larynx by negative pressure from the thorax and prevented from entering the esophagus by the sphincter action of the middle and inferior constrictor muscles. Saliva and other oral contents are prevented from entering the larynx by sensitive reflexes mediated by the sensory component of the superior and recurrent laryngeal nerves and the intrinsic and extrinsic muscles of the larynx. In the act of swallowing, the food bolus is prevented from entering the larynx by the simultaneous relaxation of the middle constrictor muscle and the elevation of the larynx toward the relatively fixed epiglottis, which seals the aditus. This elevation is initiated as the food passes through the faucial arch by the intrinsic and extrinsic muscles of the tongue pulling through the hyoid bone. Speech is produced by air expired from the lungs through the larynx, mouth, and nose. The thickness and degree of abduction or adduction of the cords produce tone and volume, and the actions of the tongue, palate, lips, and buccal mucosa modify the expelled air into an articulated pattern, speech. Although the respiratory tract can be short-circuited and separated from the digestive tract by tracheostomy, either with or without laryngectomy, the maintenance of normal speech and swallowing depends on adequate synergy in the pharyngolarynx and oral cavity of tongue, larynx, and pharyngeal musculature and normal central nervous system afferent and efferent signals. Perhaps more than at any other body site, therapeutic decision-making in laryngeal carcinoma by both patient and physician has been influenced by the desire to preserve the voice.

Etiology

Carcinoma of the larynx is the most common malignancy of the upper aerodigestive tract in the United States, with about 10,000 new cases presenting each year. As with most other head and neck cancers the risk for development for laryngeal cancer is directly proportional to the amount of exposure to tobacco, with a lesser relationship to alcohol intake. Changing mores have decreased the male-to-female ratio for incidence of laryngeal carcinoma from 11:1 in 1960 to 4.5:1 in 1990. Furthermore, women with carcinoma of the larynx are more likely to have supraglottic than glottic laryngeal cancers. Increasing age also is a risk factor, although disease in young patients has been reported and likewise is related to tobacco use. Black patients in the United States have an increased risk of developing laryngeal carcinoma at a younger age and a slightly lower male- to-female ratio (3.5:1) than the general population. Other risk factors suggested have been metal dust (particularly nickel), asbestos, hair dyes, and wood dust, but these are likely cofactors to the insult created by tobacco smoke. The relationship of papillomas of juvenile or adult onset in the causation of the larynx is unclear.

Diagnosis

Because of the obvious symptoms of hoarseness, other voice changes, tickling in the throat, and coughing, carcinoma of the glottic larynx is more likely than many upper aerodigestive tract malignancies to present with early-stage disease. Evaluation of the patient includes indirect (mirror) laryngoscopy as well as physical examination of the neck (Fig. 15-42). The geographic site and extent of the lesion, whether confined to the larynx or spreading beyond, are of importance, but the mobility of the vocal cords also is crucial, since it predicts invasion of the tumor into the deep structures of the larynx. CT and MRI are useful in determining whether there is invasion of the thyroid cartilage in the staging process (Fig. 15-43). Definitive staging and diagnosis to clearly define extent usually require direct laryngoscopy and biopsy, which can be performed with a rigid laryngoscope with the patient anesthetized or with a flexible fiberoptic scope with the patient awake and able to inspire and vocalize.

Pathology

The biologic behavior of malignancy in the various areas of the larynx is somewhat different and plays a major role in the choice of therapy. Supraglottic laryngeal carcinoma or extrinsic laryngeal carcinoma is more likely than glottic cancer to present with advanced-stage disease. There is an abundant lymphatic supply that extends anteriorly through the preepiglottic space as well as laterally to the midjugular and paratracheal lymph nodes. Involvement of the preepiglottic space occurs in 29 percent of clinically unremarkable supraglottic tumors. Micrometastases in the clinically negative neck also are common, with 26 to 33 percent of necks showing metastases and 25 percent bilateral metastases. In patients with inadequate primary therapy for supraglottic carcinoma recurrence in the neck and distant metastases is more likely than local recurrence alone.

The vocal cords (glottis) or intrinsic larynx is a small area that is in constant use and thus regularly exposed to carcinogenic stimuli. When tumors are confined to the small cartilaginous box with its paucity of lymphatics, they usually behave indolently. When tumors extend outside the glottis—transglottic lesions—a more aggressive pattern is noted. Unlike supraglottic cancer, early changes are noticeable and appear to have a long natural history, making them amenable to less radical therapy. Keratosis or thickening of the keratin layer of the squamous mucosa results in transformation to carcinoma in situ in 14 percent of cases and microinvasive cancer in 28 percent of cases over a period of 2 to 24 months. Keratosis with atypia progresses to carcinoma in situ in 18 percent and invasive carcinoma in 25 percent in 1 to 24 months, and carcinoma in situ progresses to invasive carcinoma in 38 percent of cases in 8 to 48 months. Recognition and treatment of these early changes in laryngeal mucosa is crucial to prevention of disease and prolonged survival.

Therapy

This rather indolent behavior and the low likelihood of clinically evident or micrometastatic cervical disease (1 to 7 percent in T1 and T2 lesions, 13 percent with T3 lesions) make glottic cancer more suitable for less radical therapies, with a high probability of salvage after failure of primary therapy. Because of the propensity for submucosal extension, however, advanced- stage disease with cartilage invasion requires initial radical therapy.

Subglottic carcinomas are rare, comprising only 1 to 2 percent of all laryngeal carcinomas, and tend to spread by submucosal extension down the trachea and through the cricoid cartilage and cricothyroid membrane into the soft tissues of the neck. Because of their rarity, they frequently present with advanced-stage local disease, although cervical metastasis is not as common as in supraglottic disease. Radical combined therapy, surgery and radiotherapy, is necessary.

Despite the peculiarities of growth of squamous cell carcinoma of the glottic larynx, prognosis for carcinoma at all sites is related to similar variables predicting other sites. Unifactorial and multifactorial analyses have identified the presence of lymph node metastases, advanced local disease, vocal cord fixation, histologic grade, ulceration, site within the larynx, male gender, and 3H-thymidine labeling index as significant factors in survival after therapy with both surgery and radiotherapy.

Because of the importance of preserving the voice and because of the low incidence of subclinical cervical metastases in glottic cancer, radiotherapy has become the accepted treatment for most early squamous cell carcinomas of the larynx. Conventional radiotherapeutic techniques that deliver doses of 60Gy over 30 fractions over 6 weeks result in cure rates of 80 to 90 percent for T1,N0 carcinoma of the glottis and 70 to 90 percent for T2,N0. Furthermore, salvage of radiotherapy failures with voice preservation by subtotal laryngectomy is possible in 75 percent and overall surgical salvage in 92 percent of T1 patients and 82 percent of T2 patients. More advanced T3 lesions also have been approached with standard radiotherapy and hyperfractionation (more than one course per day) in doses of 65Gy to 76.8Gy through ports that include the cervical lymphatics, resulting in local control rates of 36 to 67 percent over 2 to 4 years and surgical salvage of 67 to 83 percent. Such high-dose radiotherapy does, however, result in significant complication rates, such as persistent laryngeal edema and chondronecrosis.

Radiotherapy as treatment for supraglottic carcinoma is somewhat more controversial because of the late-stage presentation and the tendency for bilateral cervical metastases to occur. Although control rates of 74 percent for T1 and T2 supraglottic lesions have been obtained, and 40 percent for T3 and T4 lesions with radiation and surgery held as a salvage method, the possibility of preserving the voice with supraglottic laryngectomy and treatment with preoperative or postoperative adjuvant radiotherapy has made this combined approach more popular for late-stage disease or disease with ominous histologic features.

The surgical approach to the larynx is varied, with the unifying principle that local and locoregional disease may be eradicated with preservation or restoration of whatever function is possible. Early lesions of the glottis, such as keratosis with atypia, carcinoma in situ, or minimally invasive carcinoma, that involve the mobile cord and do not extend to the anterior or posterior commissures may be treated with reasonable success by removal of the mucosa over the cord by vocal cord stripping. Performed transorally, by either laser diathermy or sharp dissection, this technique has minimal morbidity, leaving good voice quality. T1 lesions of more bulk but confined to one cord may be removed by cordectomy, occasionally performed transorally but usually by opening the larynx through laryngofissure. Although this interferes with voice quality, most patients are cured, and surgical salvage by total laryngectomy is a successful backup procedure.

The conventional therapy for cancers that cross the anatomic boundaries of the supraglottis and glottis or the glottis and subglottis, and the so- called transglottic lesions, is total laryngectomy. Glottic lesions that cross the anterior or posterior commissure and are not felt to be amenable to radiotherapy similarly require total laryngectomy. Depending on the presence of palpable lymph nodes or suspicion of metastatic disease because of the site of primary disease, radical or modified radical neck dissection may be indicated. Removal of all disease is the aim of total laryngectomy, although whole-organ pathologic examinations have shown presence of microscopic disease at the margins in up to 30 percent of cases, with anterolateral (19 percent), posterolateral (11 percent), postcricoid (7 percent), and superior margins being most commonly involved. Free margin status is the most

reliable predictor of local control, with 48 percent of patients with positive margins developing local recurrence. When total laryngectomy is used as definitive therapy for advanced-stage disease, 3- to 5-year survival rates of 65 to 69 percent for T3,N0 and T3,N1 lesions and 45 to 54 percent for T4,N0 lesions have been reported.

The goal of voice preservation when surgical extirpation is a necessary part of therapy is met in two ways: by conservation surgery, removing only part of the larynx, or by reconstruction of a voice-producing mechanism after total laryngectomy.

In conservation surgery lesions confined to one part of the larynx are removed with the goal of preserving enough larynx to allow speech that is superior to esophageal or mechanical speech. Standard supraglottic laryngectomy or horizontal hemilaryngectomy is appropriate in patients where lesions are smaller than 3 cm, the vocal cords are mobile, a margin of 5 mm is possible at the anterior commissure, there is no cartilage or preepiglottic space invasion, the pyriform sinus apex and postcricoid and interarytenoid areas are clear, and tongue mobility is normal. The approach to the larynx is through the thyroid cartilage, removing the epiglottis, false cords, and ventricular mucosa en bloc with the hyoid bone. If posterior extension involves the arytenoid cartilage on one side, this may be included with the specimen. Although postoperative aspiration for a limited period of time is frequently seen, good voice quality is usually obtained and the temporary tracheostomy can almost always be removed. Total laryngectomy as a salvage procedure for intractable aspiration, fistula, or recurrent disease is successful in 33 to 50 percent of cases. Five-year survival rates in the range of 70 to 85 percent for supraglottic carcinoma make it superior to radiotherapy, although no prospective randomized studies have tested this thesis.

Vertical hemilaryngectomy is the standard conservation procedure carried out for invasive carcinoma of one vocal cord or a vocal cord and the anterior commissure that does not invade the thyroid cartilage. Total resection of the cord and its subjacent muscle and adjacent cartilage is necessary. Access to the larynx is obtained posterior to the cord by cutting through the thyroid cartilage, leaving the external perichondrium sutured to the sternothyroid muscle. The cord with or without the ipsilateral arytenoid cartilage is resected, and if the commissure is involved, the keel of the thyroid cartilage is resected as well. The strap muscles and perichondrium or a cartilage graft provides a buttress against which the normal cord can be apposed. Five-year survival rates of 75 to 87 percent for stages I to III disease treated in this fashion have been reported, again superior to radiotherapy alone. As with supraglottic laryngectomy, removal of the temporary tracheostomy and resumption of oral nutrition without aspiration usually is possible, with voice preservation. Extended hemilaryngectomy has been described for T3 lesions, where a small segment of vocalis muscle with its overlying cartilage, with sensory and motor innervation left intact, is used with the adjacent hypopharyngeal mucosa to create a phonatory shunt through which air from the lungs can be forced to provide speech. Although permanent tracheostomy is necessary, voice quality has been described as superior to esophageal speech and obtainable in a high percentage of cases.

Voice restoration after total laryngectomy is a difficult problem. Despite widespread availability of speech therapy and extensive preoperative teaching and aggressive postoperative rehabilitation, only 24 to 45 percent of alaryngeal patients acquire esophageal speech. Inability to expel air from the stomach, cricopharyngeus spasm, and other physiologic reasons have been cited, but depression and social and emotional withdrawal play an important part as well. Mechanical devices inserted into the airway have been remarkably unsuccessful, although external vibrators do allow intelligible speech of low volume. Various attempts have been described to create a neolarynx or neoglottis, a mucosal or epithelial diaphragm which fits over the transected end of the cricoid or trachea and serves as a pathway for air inspired through the tracheostomy to exit through the oral cavity as speech. With these various methods, good speech quality may be obtained in 50 to 75 percent of cases, but there is a significant risk of aspiration, requiring takedown of the reconstruction. A simple tracheoesophageal puncture described by Singer, maintained patent by a small tube, allows pulmonary air to enter the esophagus and thus the pharynx, to be modulated by the tongue, lips, and buccal mucosa. This method has allowed fluent speech restoration in 71 to 88 percent of patients with little risk of aspiration or other serious complication. This procedure may be performed either at a second stage after healing of the wound or at the time of laryngectomy, whether closed by primary approximation or reconstructed with a pectoralis major flap, gastric pull-up, or free tissue transfer. Transplantation of the larynx has as yet been unsuccessful, but research is continuing with this method.

Extensive surgery on the larynx is associated with complications. Pharyngocutaneous fistula occurs in up to 38 percent of laryngectomies and has been associated with previous radiotherapy, malnutrition, and cell-mediated immunodeficiency. Prophylactic metronidazole has been suggested as a method to prevent fistula, but prompt recognition and closure with well-vascularized tissue will prevent prolonged drainage and rupture of the adjacent irradiated carotid vessels. Problems specific to the method of reconstruction also occur, such as regurgitation of acid and food after pharyngogastrostomy after gastric pull-up or hypersecretion of mucus after jejunal autograft. The problem of stenosis after laryngectomy, particularly in the patient who receives adjuvant radiotherapy, is similar to that after pharyngolaryngectomy. Combined with the alterations in cricopharyngeus function (upper esophageal sphincter mechanism), stenosis greatly impedes the acquisition of esophageal speech. Interposition of well-vascularized soft tissue, such as the radial forearm or lateral arm free flap, can decrease the likelihood of postlaryngectomy speech and swallowing problems (Fig. 15-44).

Nonepidermoid malignancies of the larynx are uncommon. Adenosquamous carcinoma is a poorly differentiated aggressive lesion that behaves in a manner similar to advanced squamous cell carcinoma. Rhabdomyosarcoma, chondrosarcoma, schwannoma, and other mesenchymal malignancies have been reported in both children and adults. Carcinoid or moderately differentiated neuroendocrine carcinoma behaves in a fashion similar to malignancy in other sites and may be treated by partial or total laryngectomy, depending on its stage. Adjuvant radiotherapy and chemotherapy do not seem to contribute to the cure of the patient. Anaplastic small cell carcinoma (oat cell) also occurs in the larynx rarely and, like its counterpart in the lung, is associated with smoking and increased age. Prognosis is also similar to lung disease, with average survival of 10 months after diagnosis. Multimodal therapy with chemotherapy as the basis is the most appropriate approach.

One of the most difficult problems associated with laryngectomy is recurrence of disease at the stoma. Occurring in approximately 5 percent of patients, it may be the result of submucosal subglottic extension or unresected pretracheal or paratracheal lymph node metastasis. Resection of disease is possible in some patients by mediastinal dissection, repositioning the stoma lower and removing the soft tissues of the superior mediastinum to the innominate vessels. Esophageal reconstruction may be performed by gastric pull-up or jejunal free autograft, and interposition of vascularized pectoralis major muscle usually is necessary to separate the great vessels from the stoma. With this aggressive surgical approach, 17 to 28 percent of the patients can be treated successfully.

Benign lesions of the larynx may present a diagnostic problem but usually are seen in different age groups and rarely affect mobility of the vocal cords. Tuberculosis of the larynx, though rare, may mimic malignancy. Papillomatosis may present in children, possibly secondary to contamination of the fetus at delivery by the human papillomavirus, a DNA virus. This persistent problem may require tracheostomy, since repeated excisions by laser or other means may result in laryngeal or subglottic stenosis. Adult-onset papillomatosis of the larynx may come with immunosuppression of various types as recurrent disease or as an opportunistic infection. Immunotherapy with the bacille Calmette-Guérin (BCG) has been suggested as an adjuvant of surgery. Vocal nodules or singer's nodules and polyps and sebaceous cysts also may occur in the mobile cords and may be treated by simple excision.

CONNECTIVE TISSUE NEOPLASMS

Soft-Tissue Sarcomas

General Considerations

Soft-tissue sarcomas arise within tissues that develop embryologically from the mesoderm, such as skeletal and smooth muscle, adipose and fibrous tissue, and endothelial cells. They are uncommon tumors overall, with fewer than 5,000 new cases diagnosed each year in the United States. Less than 10 percent of all sarcomas arise in the head and neck; they represent less than 1 percent of all head and neck malignancies. This low incidence, with the attendant difficulty in developing a comprehensive management plan, may partially account for the poor survival rates in adults with soft-tissue sarcomas. Unlike epidermoid carcinoma, the most important determinant of the biologic behavior of sarcomas is histologic grade. Low-grade tumors tend to recur locally, while high-grade lesions often metastasize early to distant sites, including lung, liver, and bone. Other prognostic indicators include the histologic type, the location, the size and the degree of local invasion of primary lesions, and the presence of locoregional or distant metastatic disease.

Typically soft-tissue sarcoma of the head and neck presents as a painless mass. The more advanced or more deeply invasive lesions may cause pain, obstructive symptoms, or cranial nerve deficits. Evaluation of local extent of disease in soft-tissue sarcomas is similar to that of squamous cell carcinoma and includes indirect laryngoscopy, palpation endoscopy, and CT or MRI. Depending on histology and stage of disease, a metastatic work-up should be performed. Pleomorphic sarcomas of the head and neck, like those arising in other areas, frequently demonstrate cellular changes consistent with indolent disease in one part of the tumor and characteristic of very aggressive disease in another part. A generous incisional biopsy often is necessary to confirm the diagnosis. The mainstay of treatment for the majority of these lesions in adults remains extensive local resection with clearly negative margins (tumor no less than 1 mm from the surgical margin). Regional node dissections usually are performed only if nodal disease is clinically evident. In adults, adjuvant radiation therapy and chemotherapy may be of some worth, whereas in children their therapeutic value has been clearly established.

Fibrosarcoma

This is the most common type of soft-tissue sarcoma in adults. There is a bimodal age distribution, with the first peak in infants under 2 years of age and the second in adults 40 to 60 years of age. Fibrosarcoma often presents as a painless mass of the face or neck. Histologically, many fibrosarcomas are low grade. The rate of metastasis to regional nodes is 5 percent, but 50 percent metastasize to lungs, liver, or bone. With aggressive local resection, 5-year survival rates around 45 percent have been reported.

Malignant Fibrous Histiocytoma

The reported incidence of these tumors in the head and neck has grown in recent years because of the reclassification of other high-grade sarcomas as malignant fibrous histiocytomas. Two histologic variants exist, myxoid and inflammatory, with the latter carrying a poorer prognosis. Whereas local recurrence remains the most common manifestation of treatment failure, the incidence of nodal metastasis may approach 50 percent. Therefore, in addition to aggressive local resection, prophylactic neck dissection is advocated in some centers, depending on stage of disease. The 5-year survival rate is approximately 50 percent.

Angiosarcoma

These rare cutaneous malignancies most often present as an ulcerated painless mass of the scalp or face in middle-aged men. They are highly aggressive tumors, with perivascular extension most often cited as the cause of local failure as well as of distant metastasis. Radiation therapy and chemotherapy have both been used as adjuvants to local resection, but 5- year disease-free survival is rare.

Rhabdomyosarcoma

This is the most common type of soft-tissue sarcoma in children, and after lymphoma is the second most common malignancy of the head and neck in children. Two important histologic subtypes have been identified. The embryonal variant occurs in young children (average age four) and carries the better prognosis. Seventy-nine percent of orbital rhabdomyosarcomas are of the embryonal type. Most of these neoplasms are now treated successfully with a combination of radiation therapy and chemotherapy (vincristine, doxorubicin, and cyclophosphamide), yielding survival rates of 77 to 90 percent. Alveolar rhabdomyosarcoma occurs in older children (average age ten) and in the cervical soft tissues. While the initial response rate to radiation therapy and chemotherapy is comparable to that of embryonal rhabdomyosarcoma, the rate of local recurrences and of distant metastasis is significantly higher, decreasing survival rates to approximately 50 percent at 5 years. The Intergroup Rhabdomyosarcoma Study has established specific multimodal treatment protocols for each type and stage of disease. To date the improvements in management of rhabdomyosarcoma in children have not yet been realized in adults.

Neural Tumors

Esthesioneuroblastoma (Olfactory Neuroblastoma)

This rare tumor arises from the supporting cells of the olfactory neuroepithelium at the apex of the nasal cavity. In its earlier stages it may present as a fleshy upper nasal mass causing epistaxis, unilateral nasal obstruction, or rhinorrhea. In its more advanced stages the tumor may extend throughout the cribriform plate into the anterior cranial fossa, invading the frontal lobe and leading to anosmia, rhinorrhea, headaches, or even personality changes. While this tumor may arise in all age groups, it has higher incidence in the third and fourth decades of life and a slight male predominance. When esthesioneuroblastoma is suspected on the basis of the history or CT scans showing bony erosion, a biopsy procedure of the nasal mass should be performed in the operating room because its highly vascular nature. Metastasis occurs in 20 percent of cases, most commonly to the cervical lymph nodes (often bilateral), lungs, and bone, but the majority of patients succumb to aggressive intracranial extension. Therefore, surgical treatment is directed at complete local control. Early lesions may be resected through a lateral rhinotomy and complete ethmoidectomy. More extensive tumors require a combined neurosurgical-craniofacial effort, with exposure and complete resection of the cribriform plate and ethmoid sinuses from both a superior and inferior approach. Adjuvant radiation therapy may improve local control and survival, and inoperable tumors often are treated with radiation therapy alone.

Paraganglioma

Paragangliomas or chemodectomas are neoplasms that arise from neural crest cells and histologically resemble their adrenal gland counterpart, the pheochromocytoma. They are classified by their location: carotid body, jugular (arising from the glomus jugulare), vagal body, orbital, and laryngeal. Although these extraadrenal paraganglionic cells do contain small amounts of catecholamines, it is rare for them to produce a clinically significant excess of catecholamines. The most common of the paragangliomas, the carotid body tumor, usually presents as an asymptomatic neck mass. Chemodectomas at other sites often produce compressive symptoms resulting in hearing loss, tinnitus, or facial nerve paralysis. Only 6 percent of these tumors are malignant, and only 4 percent metastasize. Malignancy is determined by clinical behavior rather than histology. They are highly vascular and have a characteristic appearance on angiography, with carotid body tumors separating the internal and external carotid vessels by a mass effect. When technically feasible, treatment should include complete resection, which in the carotid body tumor requires subadventitial dissection. The poor-risk patient with an asymptomatic mass may be managed more conservatively.

Osteogenic Sarcoma

Ten percent of osteogenic sarcomas occur in the head and neck. Most arise in the mandible or maxilla in the third or fourth decade of life. Risk factors include prior radiation therapy, fibrous dysplasia, and retinoblastoma. These malignancies present most frequently as a hard mass with radiographic changes of cortical destruction as well as new bone formation. These tumors are highly infiltrative not only of bone but also of secondary mucosa and muscle and must be resected with a wide zone of adjacent soft tissue. Nodal metastasis is uncommon; however, distant metastasis to lung occurs in some patients if local control is not achieved. There has nonetheless been some enthusiasm for chemotherapy administered after surgical resection. Radiation therapy is generally reserved for curative situations. As with rhabdomyosarcoma, the use of neoadjuvant or adjuvant chemotherapy has been effective in children with osteogenic sarcoma but not in adults.

AIDS-RELATED DISORDERS

HIV infection and its consequence, AIDS, produce a large number of abnormalities in the head and neck region, both neoplastic and nonneoplastic.

Symbiotic Infections

The generalized immunosuppression allows normally symbiotic or latent organisms to cause serious disease and morbidity. Oral candidiasis is sometimes one of the earliest manifestations of AIDS. The combination of xerostomia, which may be related to the chronic fibrosis of the salivary glands or benign lymphoepithelial lesions seen in HIV-positive patients, and the disturbance of normal flora of the oral cavity results in the proliferation of yeast and other fungi to pathogenic levels. Herpes simplex virus infections, resulting in painful ulcerations of the lips, oral mucosa, and oropharynx, as well as varicella-zoster, with its painful distribution along the fifth cranial nerve, also are common in patients afflicted by AIDS, again probably a result of generalized immunosuppression. Other infectious disorders less commonly seen are human papillomavirus, cytomegalovirus, cat-scratch disease, tuberculosis, and opportunistic bacterial infection localized to the oral cavity and oropharynx.

Leukoplakia

Oral hairy leukoplakia is similar to the leukoplakia seen in chronic smokers except that it has a shaggy or more nodular appearance and is more likely to present on the lateral border of the tongue, frequently as a bilateral lesion. Cellular hyperplasia with parakeratosis is seen, and there is a high correlation, both serologically and by electron microscopy, with the Epstein-Barr virus. This lesion is sometimes an early sign of ultimate HIV positivity, again as an expression of decreased immunosurveillance.

Lymphoproliferative Disorders

Lymphoproliferative disorders are characteristic of AIDS and frequently occur in the head and neck. Obstruction of the nasopharynx secondary to overgrowth in the Waldeyer's ring area has been reported. Cervical lymphadenopathy is commonly a part of the AIDS-related complex (ARC), and involvement of the intraparotid and submandibular lymph nodes, with cystic degeneration and symptomatic accumulation of fluid within them, causes enlargement of the nodes and is an early indicator of the full-blown disease. Malignant lymphomas, most commonly of the B-cell type, also have been reported and are second only to Kaposi's sarcoma in frequency, occurring in up to 10 percent of AIDS patients. These lymphomas usually are extranodular, with an unusual frequency of primary central nervous system disease. Lymphoma also has been reported in the neck, oral cavity, and paranasal sinuses. Although the tendency to disseminate is lower than in other lymphomas, the disease course is rapid and lethal. Lymphoma occurs more commonly in intravenous drug users than in patients who have developed AIDS from other sources. Again there is a high correlation of lymphoma with elevated antigen titers for Epstein-Barr virus and electron microscopic evidence of its presence. T-cell lymphomas also occur but are less common. For disease localized to one area, radiotherapy may be an effective treatment. Multidrug systemic chemotherapy also is effective but places the already immunosuppressed patient at higher risk for disseminated infection.

Kaposi's Sarcoma

Kaposi's sarcoma is the most common malignancy in AIDS, arising in 15 percent of patients. There is probably a sexual mode of transmission, since it is much more common in homosexual AIDS patients than in those acquiring the disease from intravenous drug use or contaminated blood products. Unlike the classical form of Kaposi's sarcoma, which occurs in elderly males usually on the lower extremities, this form of the disease occurs in oral or perioral mucosa in 55 percent of the cases. The palate is the most common site, although the tongue, pharynx, and larynx have been described as sites as well. Usually multifocal within the region, it may arise in or metastasize to the cervical lymph nodes or the salivary glands. The tumor initially presents as a flat blue to purple patch and may appear to be a submucosal hematoma secondary to trauma. Later in the course of growth it becomes nodular. Biopsy examination reveals endothelial cell and fibroblast proliferation, with increased capillary growth, prominent spindle cells, few mitoses, and extravasation of erythrocytes. The causation of Kaposi's sarcoma is unclear, although there is some evidence linking it to cytomegalovirus infection or viral induction of local angiogenic factors.

As with all malignancies in AIDS patients, the use of conventional multidrug systemic therapy is somewhat hazardous because of chronic infection. Single-drug treatment with vinblastine or etoposide, systemically or by intralesional injection, results in partial regression of tumor and significant palliation in up to 30 percent of patients. Interferon- α , systemically or as an intralesional injection, also has been moderately effective. Local radiotherapy with an energy source, such as electrons, that has low penetration, in total doses of 800 to 1,500 cGy, delivered in ten fractions, also is effective in shrinking lesions and allowing significant palliation. With all therapies, median survival of patients is 2 years, with a 5-year survival rate of only 10 percent.

Carcinoma

Squamous cell carcinoma of the upper aerodigestive tract has been reported in the mouth, tongue, larynx, and other sites in AIDS patients. Although the true risk is uncertain, squamous cell carcinoma has appeared in patients of earlier age and without the usual risk factors of smoking and ethanol intake. This pattern is suggestive of a defect in immunosurveillance similar to that seen in patients having undergone organ transplantation. In AIDS patients, the median age of presentation of squamous cell carcinoma is 32 years, in contrast to 60 years in the noninfected patient, and the clinical course seems to be somewhat more aggressive.

SALIVARY GLANDS

Physiology

The production and excretion of the mixture of mucus, water, and electrolytes known as saliva into the entrance to the upper aerodigestive tract are the function of the salivary glands. The major salivary glands are the symmetrically paired parotid, submandibular, and sublingual glands, which discharge saliva into the oral cavity via Stensen's ducts, Wharton's ducts, and the numerous small orifices in the floor of the mouth, respectively. Clustered primarily in the soft and hard palates but also found in the sinuses and other sites of the upper aerodigestive tract are the minor salivary glands, which produce mucus for lubrication. Saliva consists of varying concentrations of numerous substances. To some degree its composition reflects the osmolality and electrolyte concentration of the extracellular fluid, but its specific functions require other components as well.

The normal volume of salivary secretion in the adult male ranges from 1,000 to 1,500 mL per day, mainly as serous fluid from the parotid and submandibular glands (95 percent) but also composed of mucus (5 percent) from the sublingual and minor salivary glands. Active water resorption, sodium and potassium ion exchange, and bicarbonate production couple with passive diffusion of urea and uric acid to determine the inorganic content of saliva. Immunoglobulins A, G, and M, albumin, lysozyme, and other enzymes also are secreted. In addition to its lubricating properties, which allow food to be moved through the mouth, saliva has antibacterial and antiviral properties, which protect the soft tissues of the oral cavity as well as the teeth.

The neurogenic control of salivary secretion depends on reflex arcs that carry afferent stimuli from the specialized sense organs via cranial nerves I, V, VII, and IX to the brainstem and hypothalamus, and efferent signals, both sympathetic and parasympathetic, along the branches of the external carotid artery, the superficial petrosal nerve, and the chorda tympani.

Anatomy

Parotid Gland

The parotid gland is shaped like a flattened pyramid, with its apex in the parapharyngeal space behind the mandible adjacent to the pterygoid muscles and its base extending into the preauricular area from below the angle of the mandible in the neck, sometimes as low as the mid–sternocleidomastoid muscle, to just below the zygomatic arch. The medial extent of the gland usually reaches over the masseter muscle and the vertical ramus of the mandible. The parotid gland is surrounded by the continuation of the investing layer of the cervical fascia, the superficial layer of which is continuous with the platysma. The parotid gland parenchyma is arbitrarily divided into the deep and superficial lobes by the facial nerve, which exits the skull via the stylomastoid foramen and passes through the substance of the gland. Approximately 70 percent of the gland lies superficial to the plane of the nerve and 30 percent deep to it.

The lymphatic drainage of the midface, forehead, and anterior portion of the scalp empties into lymph nodes that lie superficial to the fascia of the parotid and within the gland itself before sending efferent channels to the jugulodigastric area. Sensory supply to the gland derives from branches of the trigeminal nerve and from the great auricular nerve, carrying cervical plexus axons. Stensen's duct, the parotid duct, condenses from the larger intralobular ducts and passes adjacent to the buccal branch of the facial nerve on a line between the tragus of the ear and the midline of the upper lip. It enters the oral cavity adjacent to the second maxillary molar tooth.

The facial nerve, which supplies motor innervation to the muscles of facial expression and several other muscles, exits the base of the skull from the stylomastoid foramen, passing deep to superficial and caudad to cephalad before branching at the pes anserinus into an upper and lower division. Although there is some variability in branching patterns, particularly of the buccal branch, the upper division usually includes the temporal, zygomatic, and buccal branches, and the lower division, the mandibular and cervical branches.

Submandibular Gland

The submandibular salivary gland is surrounded by a condensation of the cervical fascia that lies beneath the platysma muscle, connecting to the parotid fascia and known as the *pars interglandularis*, separating the parotid and jugulodigastric areas from the submandibular triangle. The anatomic boundaries of the submandibular triangle are the anterior belly of the digastric muscle medially, the posterior belly laterally, and the mandible superiorly. The gland lies on the hyoglossus muscle and wraps both superficial and deep to the mylohyoid muscle medially. The submandibular ganglion of the lingual nerve carries efferent nerve supply via the chorda tympani. Wharton's duct conveys the secretions of the submandibular salivary gland into the oral cavity.

Sublingual Glands

The sublingual salivary glands lie immediately beneath the mucosa of the floor of the mouth, intimately related to the lingual artery, and release their mucous secretions into the oral cavity through numerous orifices along the alveololingual sulcus.

Inflammatory and Infectious Disorders

Inflammation usually presents as diffuse enlargement or firmness of the gland, unilateral or bilateral, associated with tenderness and erythema. Bacterial infection usually is the result of duct obstruction and retrograde infection with oral bacteria. Acute bacterial parotitis may be seen in the elderly postoperative patient who becomes dehydrated and is usually caused by *Staphylococcus aureus*. Although rehydration and antibiotic therapy may be successful, drainage of localized abscesses and even total parotidectomy may be required for resolution. Because of the septate nature of the parotid, simple drainage of the gland is frequently unsuccessful. Acute sialadenitis of the submandibular gland also may necessitate gland resection.

Mumps, coxsackievirus, and echoviruses also may cause acute parotitis, usually panglandular but occurring as unilateral gland enlargement. Tuberculosis, actinomycosis, and cat-scratch disease also may present with enlargement of the salivary glands or of their adjacent lymph nodes. Systemic disorders such as sarcoidosis, Sjögren's syndrome, and cirrhosis with liver failure also result in salivary gland enlargement. Mikulicz's syndrome, enlargement of the gland with histologic changes reflecting loss of acinar epithelium and replacement with chronic inflammatory cells, is a nonspecific accompaniment of several diseases, such as leukemia, lymphoma, and tuberculosis.

Tumors

Pathology

The clinical problem most frequently presented to the surgeon is that of a discrete mass in the salivary gland, particularly the parotid gland. Of salivary gland tumors, 70 to 80 percent present in the parotid gland. Of parotid gland tumors, 70 to 80 percent are benign and, of the benign tumors, 80 percent are pleomorphic adenomas. Although lymph nodes, lipomas, cysts, or other benign entities may present as solitary masses, the benign pleomorphic adenoma is by far the most common mass. Occurring most frequently in the fifth decade, with a slight female predominance, the benign mixed tumor or pleomorphic adenoma is the proliferation of epithelial and myoepithelial cells of the ducts as well as an increase in the stromal component, which histologically may appear similar to cartilage. A true epithelial benign neoplasm, it may grow to a large size without causing facial nerve symptoms (Fig. 15-45). Pleomorphic adenomas usually present as a solitary painless mass in the superficial lobe of the gland. Deep lobe growth may present as an intraoral pharyngeal mass as growth extends into the pharynx via the parapharyngeal space. Malignant degeneration of pleomorphic adenomas occurs in 2 to 10 percent of adenomas followed for long periods, with carcinoma ex pleomorphic adenoma occurring most frequently as adenocarcinoma.

The second most frequent benign neoplasm of the salivary glands is the papillary cystadenoma lymphomatosum or Warthin's tumor. With a marked male predominance, it usually occurs in the tail of the parotid gland and presents with a lymphocytic infiltrate as well as cystic epithelial proliferation. There is a 10 percent incidence of bilaterality and multicentricity. A subset of the group of monomorphic adenomas, Warthin's tumors comprise 4 to 8 percent of all parotid tumors. Other benign monomorphic adenomas include oxyphilic adenoma, oncocytoma, basal cell adenoma, sebaceous adenoma, sialadenoma papilliferum, and canalicular adenoma, all of which are rare. Pleomorphic adenoma is the most common benign tumor of the parotid, submandibular, and minor salivary glands.

Malignant tumors of the salivary glands occasionally (4 percent) present as a diffuse enlargement of a gland, almost always as a discrete mass. Pain is associated with malignancy in 12 to 24 percent of cases. Other symptoms include formication (a paresthesia that is described as the feeling of ants crawling on the skin), facial nerve dysfunction (8 to 26 percent), or complete paresis of the nerve (7 to 9 percent). All of these symptoms and signs carry a poorer prognosis than asymptomatic disease (Fig. 15-46). Facial nerve palsy is almost never seen with benign disease, so the presence of nerve palsy, even without a discrete mass, must be considered a possible sign of malignancy.

Fixation to the masseter muscle or pterygoid muscles occurs in approximately 17 percent of cases, and skin ulceration in 9 percent. The risk of clinical or subclinical metastases to the cervical lymph nodes from salivary gland cancers depends on the histology and grade of the primary tumor. High-grade mucoepidermoid adenocarcinoma and squamous cell carcinoma have a high risk of metastatic disease (Fig. 15-47), whereas adenoid cystic acinic cell and lower grades of mucoepidermoid and squamous cell have a low risk of metastasis. Approximately 20 percent of parotid gland neoplasms are malignant. The risk of malignancy is inversely proportion to gland size, with the submandibular glands having 40 percent malignant tumors and the minor salivary glands 60 percent.

Diagnosis

The discrete mass in the salivary gland must be considered a possible malignancy. History and physical examination may provide some indication that the lesion is malignant. Complete resolution after 10 days of antibiotics and clinical features consistent with inflammation may constitute an adequate therapeutic trial. Definitive histologic diagnosis, however, is ultimately necessary. Fine-needle aspiration may be helpful in planning surgery, but any uncertainty should be treated with adequate surgical excision. CT scan of the parotid is helpful in determining extension of the tumor into the deep lobe. MRI can be formatted in both axial and coronal planes, giving even better anatomic information. When carefully performed and analyzed MRI can provide information about the facial nerve and its anatomic relationship to the tumor. Contrast-enhanced studies may allow discrimination between the gland and metastatic lymph nodes, particularly in the submandibular area, differentiating metastases from the head and neck site to an intraglandular or epiglandular lymph node from intrinsic malignancy of the salivary gland. Sialography, or injection of contrast material into Stensen's or Wharton's ducts, is useful in demonstrating the chronic stenotic changes of a benign lymphoepithelial lesion or chronic parotitis and in showing complete occlusion from stones. Eighty percent of parotid duct stones are radiolucent. Eighty percent of submandibular gland stones are radiopaque.

Treatment

The surgical approach to a salivary gland mass is predicated on the assumption that it is malignant. The major confounding factor is the presence of the facial nerve in the parotid gland. Since the facial nerve passes through the gland, the usual surgical oncologic approach to head and neck malignancy—en bloc resection—would require excision of the nerve in all cases. This approach, however, is not appropriate in most situations. If there is no evidence of nerve involvement, the tumor should be excised by superficial lobectomy, removal of the parenchyma above the nerve with a margin of normal tissue, preserving the nerve.

If the tumor is malignant, total parotidectomy with preservation of the nerve is indicated, though it is a piecemeal procedure (Fig. 15-48). Involvement of a branch of the nerve requires removal of that branch. In young patients, nerve graft should be used to replace the resected nerve segment in the hope of avoiding the long-term sequelae of facial nerve palsy. In the event of invasion of the nerve by tumor, any proximal extension of the malignancy to the base of the skull should be evaluated, and in some cases resection of the nerve to clear margin in the stylomastoid foramen or the facial canal may improve survival. If the facial nerve is not involved by malignancy but preservation of the nerve would result in gross disruption of the tumor, the nerve should be removed and replaced with a nerve graft.

When clinical examination, with or without fine-needle aspiration cytology, does not clearly define the problem, a biopsy specimen should be obtained by superficial lobectomy, with identification and preservation of the main trunk of the facial nerve and its branches. Benign tumors of the superficial lobe should be removed with a clear margin by superficial lobectomy. If the deep lobe is involved, total parotidectomy may be required even for benign disease, although partial parotidectomy is sometimes possible.

Treatment of the neck in patients with malignant disease of the parotid depends on the histologic type and the grade of the tumor and its risk of metastatic disease or the presence of the metastatic disease itself. Node- positive necks are treated by the appropriate neck dissection—radical neck dissection if there is involvement in the sternocleidomastoid muscle or the jugular vein, or modified or selective neck dissection depending on the site of metastasis. Although elective or prophylactic neck dissections are not as frequently necessary as in mucosal malignancy, they are indicated in high- grade mucoepidermoid carcinoma, squamous cell carcinoma, and high- grade adenocarcinoma.

Neoadjuvant (preoperative) or adjuvant chemotherapy has not been effective in treating malignancy of the salivary glands—parotid, submandibular, or minor glands. Adjuvant postoperative radiotherapy, however, is effective. Radiation portals should include the entire site of surgery, the foramen ovale, the base of the skull, including the mastoid process and the stylomastoid foramen, and the ipsilateral neck depending on the risk of metastasis.

The risk of recurrent disease and the pattern of recurrence also depend on histology and grade. Mucoepidermoid carcinoma, squamous cell carcinoma, and high-grade adenocarcinoma have a high frequency of cervical and distant metastasis as well as local recurrence. Adenoid cystic carcinoma is characterized by an indolent course but relentless local progression and perineural invasion, with disease eventually extending to the base of the skull and the brain. Both adenoid cystic and mucoepidermoid carcinoma may demonstrate extensive pulmonary metastasis that remains asymptomatic for relatively prolonged periods, and the patient's death may be caused by locoregional disease rather than by disseminated disease. Just as adjuvant chemotherapy has been unsuccessful, so is treatment of disseminated disease, although both doxorubicin and cisplatin have demonstrated finite response rates without prolongation of survival. Local recurrence may be successfully treated by radical resection with or without adjuvant radiotherapy.

The treatment of submandibular gland abnormalities follows the same basic rule as for the parotid gland. If definitive diagnosis cannot be made before surgery, total excision of the gland, with preservation of the uninvolved marginal mandibular, hypoglossal, and lingual nerves, is indicated and is adequate therapy for benign tumors and inflammatory or autoimmune disorders. Radical resection of the nerves, platysma, skin, and underlying muscle is reserved for extensive local invasion. Adenoid cystic carcinoma is the most common malignant histology of the submandibular gland, and pleomorphic adenoma the most common benign tumor. Adjuvant postoperative radiotherapy appears to be helpful in the malignant tumor.

Minor salivary gland disorders reflect the spectrum seen in the major salivary glands. Mucocele is a cystic enlargement of the intraoral glands usually seen in the lip or the floor of the mouth. Sjögren's syndrome of keratoconjunctivitis sicca is diagnosed by lip biopsy and histologic confirmation of the chronic inflammatory changes seen in the minor salivary glands. Necrotizing sialometaplasia, an ulcerative but self-limited disorder affecting the junction of the hard and soft palates, clinically mimics malignancy but eventually heals; biopsy examination reveals no evidence of malignant change.

Tumors of the minor salivary glands, either benign or malignant, may occur in any of the mucosa-lined areas of the upper aerodigestive tract but are most common in the hard and soft palates. Their presentation may be as a submucosal or an ulcerative mass. Pleomorphic adenoma is the most common benign tumor, and mucoepidermoid and adenoid cystic carcinoma (Fig. 15-49) occur with roughly the same frequency as malignancies. Therapy is wide local resection, including subjacent bone if the hard palate is involved, with adjuvant radiotherapy for malignancy. Since subclinical metastases to the neck are rare, cervical lymphadenectomy is reserved for patients with histologically proved lymph node metastases. Palatal defects usually can be rehabilitated with dental prostheses, though local or distant tissue transfer may be necessary for extensive disease.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 16 - Chest Wall, Pleura, Lung, and Mediastinum

Valerie W. Rusch
Robert J. Ginsberg

INTRODUCTION AND HISTORY

Thoracic surgery focuses primarily on the organs that support the delicate sequence of events that move air to blood and blood to tissues. The cardiorespiratory system functions to ensure that those events occur dependably, the margin of error being extremely small. The analysis and management of surgical problems involving the chest and its contents, whether relating to congenital anomalies, tumors, trauma, or infection, all center on maintaining the mechanical transport of oxygen to the vital organs and the necessary exchange of gases. Air with adequate oxygen content must pass through the upper airway, the trachea, and the bronchi to reach the alveoli, properly warmed and humidified, for movement across alveolar membranes. Those membranes must allow efficient diffusion of oxygen and carbon dioxide. Blood with sufficient oxygen-carrying capacity must circulate through the alveolar capillaries in adequate volumes and at the proper speed to take up oxygen and discharge carbon dioxide; it also must be at the proper pH and temperature and must have the proper biochemical characteristics for optimal exchange. The vascular system must have the appropriate integrity, pressure gradients, volume, and flow dynamics to traverse the pumps and conduits from alveolar capillaries to vital organ capillaries and back. At the end-organ interface, the characteristics necessary for release from the blood to the tissues of oxygen, and recapture of carbon dioxide must be present. Irreparable damage to vital organs may occur in minutes if any part of the system fails.

The early history of thoracic surgery was limited to the management of trauma and was closely linked to the history of weaponry. Management of chest wounds received in battle was recorded in ancient writings, including the Iliad (ca. 950 b.c.). Galen described a patient who recovered after partial excision of the sternum and pericardium for recurrent abscess due to an injury. Writing about chest wounds in the thirteenth century, Theodoric noted that “the stitches should be placed. . . so that the natural heat cannot escape in any way nor the air outside be able to enter.”

The introduction of firearms in the fourteenth century complicated the management of chest wounds. The proper care of the open pneumothorax (sucking chest wound) remained unsettled for centuries. Consistent with many of his other revolutionary insights, Napoleon's surgeon, Baron Larrey, confirmed the sporadic observations of other surgeons about the lifesaving value of closing an open wound of the thorax. His description of the cardiopulmonary effects of an open chest wound can hardly be improved on:

A soldier was brought to the hospital of the Fortress of Ibrahyn Bey, immediately after a wound penetrated the thorax, between the fifth and sixth true ribs. It was about 8 cm in extent. A large quantity of frothy and vermilion blood escaped from it with a hissing noise at each inspiration. His extremities were cold, pulse scarcely perceptible, countenance discolored, and respiration short and laborious; in short, he was every moment threatened with a fatal suffocation. After having examined the wound, the divided edges of the part, I immediately approximated the two lips of the wound, and retained them by means of adhesive plaster and a suitable bandage around the body. In adopting this plan, I intended only to hide from the sight of the patient and his comrades, the distressing spectacle of a hemorrhage, which would soon prove fatal; and I, therefore, thought that the effusion of blood into the cavity of the thorax, could not increase the danger. But the wound was scarcely closed, when he breathed more freely, and felt easier. The heat of the body soon returned, and the pulse rose. In a few hours he became quite calm, and to my great surprise grew better. He was cured in a very few days, and without difficulty.

The development of elective thoracic surgery has followed closely the history of airway management, particularly techniques for tracheal intubation and mechanical ventilation. During the first third of the twentieth century, surgeons devised imaginative techniques to control the open pneumothorax associated with chest operations. These included hand bellows and tracheostomy devices as developed and modified by Fell (1883), O'Dwyer (1896), and Matas (1900), as well as various packing techniques, preoperative pneumothorax for “conditioning,” and suturing the lung to the parietal pleura. Improvements in laryngotracheal intubation techniques, in design and materials of endotracheal tubes, and in anesthesia gradually displaced these improvisations.

In 1904 Sauerbruch developed a negative-pressure chamber in which the operating team and the patient could be housed. Under these conditions, the lung would not collapse when the chest was open. Although animal experiments produced some success, operations on patients were not rewarding. Orotracheal intubation with metal tubes for the treatment of croup and for the prevention of aspiration during oral surgical procedures provided the early experience that led to positive-pressure endotracheal anesthesia for thoracic surgery.

World War I saw the development of techniques for managing empyema. Almost two-thirds of all deaths in that war were related to pneumonia and subsequent empyema. The Empyema Commission of the United States Army, led by Evarts Graham, developed the principles of management of empyema in an era long before antibiotics. The mortality rate of this complication dropped remarkably after the commission established the standard of tube thoracostomy instead of open drainage for acute empyema.

During the same era, thoracic surgeons began developing techniques for managing “the white plague”—tuberculosis. This included the concepts of cavitory collapse and thoracoplasty. Pulmonary resection for tuberculosis, although attempted in the late 1800s, was not successful until the techniques of pulmonary resection were developed in the twentieth century. It wasn't until 1932 that the first successful pneumonectomy for cancer was performed by Evarts Graham in St. Louis. Thereafter, surgeons rapidly elaborated sophisticated techniques of hilar dissection and individual ligation of vessels, thereby allowing safe resection of lungs, lobes, and segments.

During World War II the Army Medical Corps, faced with many severe chest injuries, developed standards for the surgical management of hemothorax, and subacute and chronic empyema that remain accepted to this day.

The advent of antibiotics greatly reduced the hazards of pulmonary infections and the need for surgical intervention. After World War II the incidence of cigarette smoking rose dramatically, leading thoracic surgeons to focus on the surgical management of lung cancer. As lung cancer became epidemic and improvements in radiotherapy and chemotherapy occurred, combined-modality therapy for locally advanced tumors became integral to the treatment of lung cancer.

With increasing knowledge of pulmonary physiology and the emergence of intensive care units in the past half century, thoracic surgery has become safer, with rapidly decreasing postoperative complications and mortality.

The introduction of flexible fiberoptic equipment allowed flexible bronchoscopy to replace, for most indications, the use of rigid equipment and largely eliminated the need for general anesthesia in diagnosing pulmonary problems. These small-caliber instruments allow examination not only of the major airways but also of the segmental and subsegmental airways for diagnostic and therapeutic purposes.

The past decade has seen another revolution in thoracic surgery. Improved optical and video equipment has led to the development of minimally invasive video-assisted techniques (video-assisted thoracic surgery, or VATS). Surgeons now are able to perform intrathoracic diagnostic and therapeutic procedures without the debilitating effect of a thoracotomy. In many instances, this has decreased postoperative pain and shortened hospital stays.

Improved surgical techniques also have allowed diseases that previously were managed non-surgically to be treated curatively by surgical resection. The work of Grillo and Pearson in developing mobilization techniques for the trachea now routinely permits tracheal resections for benign and malignant conditions.

Surgery for non-infectious benign conditions of the lung previously was limited to the resection of compressive bullae. The successful development of lung transplantation by the Toronto group during the 1980s now offers patients with end-stage pulmonary disease a new lease on life. Spurred by the pioneering work of Brantigan and Cooper, surgeons are focusing attention on lung volume-reduction surgery as a bridge or an alternative to lung transplantation for patients with diffuse severe emphysema.

Throughout its history, the demise of pulmonary surgery has been predicted, first with the successful management of empyema, next with the introduction of antibiotics to treat pulmonary infection, and finally with the decreasing incidence of smoking in developed countries. Instead, the specialty has evolved rapidly as improved medical therapies obviated the need for surgical intervention in some diseases, but new diseases or technologies made thoracic operations necessary and appropriate. The spectrum of problems managed by thoracic surgeons today is broader and more complex than a decade or two ago.

ANATOMY OF THE THORAX AND PLEURA

Thoracic diseases often can be localized by physical examination to underlying anatomy because the bony parts of the thorax are palpable and cardiac and breath sounds are transmitted through the chest wall. Several anatomic factors can be misleading. The “squaring off” effect of the shoulder girdle gives the chest the physical appearance of a rectangle, obscuring the fact that the chest wall is conical in shape, tapering quite sharply in the upper chest. The lung apices rise well above the level of the clavicles anteriorly and the scapula posteriorly. The diaphragm rises as high as the level of the nipple, and the upper part of the abdomen, including the liver, spleen, stomach, distal pancreas, and kidneys, is overlapped by six of the ten anterior ribs and the lower four posterior ribs. These easily overlooked anatomic facts can lead to serious errors in patient management, especially in patients with penetrating trauma (Fig. 16-1).

The framework of the thoracic cage consists of the sternum, twelve thoracic vertebrae, ten pairs of ribs that end anteriorly in segments of cartilage, and two pairs of floating ribs. The thoracic inlet has a rigid structural ring formed by the sternal manubrium, the short, semicircular first ribs, and the vertebral column. As a result of its articulation with the manubrium and the attachment of the costoclavicular ligament, the clavicle helps protect the subclavian vessels and brachial plexus, which traverse the thoracic inlet. The same rigidity that provides protection from trauma leaves little room for pathologic swelling, enlarging masses, or age-related postural adjustments.

The cartilages of the first six ribs have separate articulations with the sternum; the cartilages of the seventh through the tenth ribs fuse to form the costal margin before attaching to the lower margin of the sternum. Since there is significant flexibility of the chest wall in children, serious trauma can be transmitted to the intrathoracic structures with little injury to the bony framework. Although this flexibility decreases progressively with age, surprising damage can occur occasionally in the chest of adults without evidence of skeletal injury.

The pectoralis major and minor muscles are the principal muscles covering the anterior thorax, and the lower margin of the pectoralis major forms the anterior axillary fold. Auscultation of the chest in the axilla often allows the best determination of breath sounds because the thoracic cage is covered only by the origins of the serratus anterior muscle in that location. The long thoracic nerve and vascular pedicle pass vertically on the anterior surface of that muscle—a point to remember when doing a thoracentesis or tube thoracostomy. The latissimus dorsi and teres major muscles converge to form the posterior axillary fold on each side. The auscultory triangle can be palpated near the inferior medial border of the scapula. Posterior and inferior to this, the latissimus dorsi, trapezius, rhomboid, and other shoulder girdle muscles form a strong muscular coat for the posterior thorax.

The sternal angle is easily palpable and allows quick identification of the second rib because of its articulation with the sternum at this location. A plane that is parallel to the floor and passes through the sternal angle of an upright patient also passes through the fourth or fifth thoracic vertebra. The tracheal bifurcation lies in this same plane, whereas the apex of the aortic arch is located slightly higher. There is a gradual increase in the length of ribs from the first to the seventh and a progressive lateral displacement of the rib–costal cartilage junctions. Because of the radiolucency of the cartilages, standard anteroposterior chest x-rays may not show an injury to the thoracic cage even though a severe blunt trauma to the chest has disarticulated and fractured multiple costal cartilages.

The pleura is a serous membrane of flat mesothelial cells overlying a thin layer of connective tissue in which a vascular and lymphatic network is distributed. The part covering the lungs is referred to as the *visceral pleura*, and it is continuous over the pulmonary hilus and the mediastinum with the parietal pleura, which covers the inside of the chest wall and the diaphragm. As shown in Fig. 16-2, the line of pleural reflection extends slightly beyond the lung border in each direction. While it is convenient to consider the pleura as a closed sac around the pleural cavity, this concept misrepresents a highly dynamic structure. The pleural surfaces behave more like a flowing syncytium across which fluids actively move (from visceral pleura to parietal pleura), phagocytosing cells and debris and sealing air and capillary leaks. It is this physiologically active membrane that contributes to the general resistance of the pleural space to infection and the lung's remarkable ability to tolerate the trauma of surgery or injury with such a low frequency of persisting air-leak problems. With normal lung expansion, the pleural cavity is completely filled, and only a potential space exists. When lung volume is decreased, such as with lobar atelectasis, the normally negative pressure in the pleural space becomes positive, allowing fluid to be drawn into the pleural cavity. Other disease processes such as infection or malignancy, which either increase production or inhibit clearance of pleural fluid, also lead to the accumulation of fluid in the pleural space.

There is no communication between the pleural cavities, but the anteromedial borders of the two pleural sacs come nearly into apposition behind the sternum. The interior border of each pleural cavity is located at the ninth rib in the midaxillary line, and the borders continue posteriorly in the eleventh intercostal space. Occasionally, the pleural sac extends as low as the twelfth rib. Posteriorly, the margins of the two pleural sacs lie on the anterolateral surfaces of the vertebrae separated by the esophagus. A retroesophageal recess is formed when the pleural margins are in near apposition, and pulmonary lesions arising in the recess can be mistaken for mediastinal tumors or cysts. At the inferior margin of the lung hilus on each side, a double layer of mediastinal pleura is formed (the inferior pulmonary ligament).

The structures occupying the intercostal spaces are important in relation to thoracic function, disease, and diagnostic procedures. The parietal pleura, for example, is well supplied with nerve endings for pain, whereas the visceral pleura is insensitive. Only when pulmonary disease extends to involve the parietal pleura or chest wall is pain produced. Fig. 16-3 shows the structures in an intercostal space and emphasizes the layering effect of the muscles and fascia. Three layers of intercostal muscles are present in a major part of the thoracic wall, but some anatomists consider the intercostals to be a single muscle entity. With quiet respiration, the ribs are elevated by synchronous contraction of the intercostal muscles. Because the ribs of each side move as a unit in respiration, a localized painful lesion may disrupt effective function of the entire side. During quiet respiration, however, movements of the diaphragm provide approximately 75 percent of pulmonary ventilation, and temporary loss of unilateral intercostal muscle function is not a threat to breathing. With labored breathing or in individuals who have impaired diaphragmatic muscle function, the muscles of the upper extremity and the cervical muscles attached to the chest wall assist in elevation and expansion of the thorax.

The endothoracic fascia is a layer of light areolar tissue subjacent to the parietal pleura. At the apex of each hemithorax, it is thickened into a more substantial layer referred to as *Sibson's fascia*.

The vein, artery, and nerve of each interspace are located deep to the external and internal intercostal muscles and lie just behind the lower margin of the rib. For most interspaces, a smaller collateral artery runs along the top border of the rib below. There is significant overlap of neural supply by adjacent nerves, and complete anesthesia in an interspace generally will not occur unless the intercostal nerve of the adjacent space above and below and the space in question are anesthetized. To minimize the risk of lacerating the intercostal artery, a thoracentesis needle or a clamp used to perforate the pleura for insertion of a catheter should be passed across the top of the lower rib of the selected interspace.

The lymphatic drainage of the chest wall extends in both anterior and posterior directions. Lymph draining from the anterior region of the first four or five intercostal spaces passes to lymph nodes along the internal thoracic arteries. These nodes may be connected by cross-anastomoses before draining into a single or double trunk that joins the thoracic duct, a right lymphatic duct, or a bronchomediastinal trunk. Lymphatics that drain the posterior and lateral regions of the intercostal spaces are tributary to lymph nodes that lie near the vertebral ends of each interspace. In the lower part of the thorax, these nodes join the drainage from the posterior mediastinum to contribute to the cisterna chyli. The posterior lymph nodes of the upper thorax drain into the thoracic duct or a right lymphatic duct.

A musculofibrous floor is provided for the thorax by the diaphragm. The peripheral muscular portions of the diaphragm arise from the lower six ribs and costal cartilages, from the lumbar vertebrae (right and left crus), and from the lumbocostal arches. Additional fibers arise from the xiphoid cartilage, and all the muscular elements converge into the central tendon. The central part of the tendon underlies the pericardium, whereas the right and left divisions extend posteriorly. Some of the lower intercostal nerves are thought to contribute to the sensory innervation of the diaphragm, but motor innervation is supplied by the phrenic nerve on each side.

Of the three major openings in the diaphragm, the aortic hiatus is most posterior. The aorta, azygos vein, and thoracic duct pass through this opening. The esophageal hiatus transmits the esophagus and the vagus nerves, and only the inferior vena cava goes through the foramen of that name.

Contemporary imaging techniques (including computer-assisted tomographic and nuclear magnetic resonance scanning) have increased the clinician's ability to identify anatomic relationships and their clinical significance. They have dramatically altered the preoperative assessment of pulmonary and mediastinal lesions.

Fig. 16-4 shows the cross-sectional anatomy at four different levels in the thorax associated with identifiable topographical landmarks. These studies provide considerable anatomic clarification of intrathoracic problems.

THORACIC INCISIONS

In recent years, an increasing variety of incisions has been used to diagnose, stage, and treat thoracic disease. This reflects an evolution in the types of procedures performed and technological advances, such as videothoracoscopy, that permit some procedures to be performed through a minimally invasive approach. Some incisions are used primarily for diagnostic operations, whereas others are used for therapeutic procedures.

Cervical mediastinoscopy and parasternal mediastinotomy incisions are used mainly for diagnostic purposes, particularly for staging lung cancers or determining the cause of mediastinal adenopathy. Mediastinoscopy is performed via a 3-cm incision in the suprasternal notch, which is carried down to the trachea in the midline between the strap muscles. The pretracheal fascia is incised, and blunt digital dissection along the anterior tracheal wall to the level of the carina then permits insertion of the mediastinoscope for biopsy of the mediastinal lymph nodes. These nodes are accessed by blunt dissection (using a suction cautery), back through the pretracheal fascia into the mediastinal soft tissues. A standard staging procedure for lung cancer includes biopsies of the right and left paratracheal and the subcarinal lymph nodes.

A parasternal mediastinotomy, which usually is performed on the left side, is also known as the *Chamberlain procedure* (after Maxwell Chamberlain, the surgeon who popularized this approach for staging left upper lobe cancers). A 6-cm incision is made over the left second costal cartilage, which is excised. The internal mammary vessels are ligated or preserved, and dissection is carried medial to the pleura posteriorly into the mediastinum along the aortic arch. This permits direct biopsy of the aortopulmonary window lymph nodes and an assessment of the resectability of the primary tumor with respect to the pulmonary artery. This approach also is used frequently to biopsy anterior mediastinal lymphomas, which are usually located just beneath the second and third costal cartilages.

For therapeutic procedures, the most frequently used incision is the posterolateral thoracotomy, but the anterolateral thoracotomy incision also is used for many general thoracic operations. Both these incisions require division of one or more major shoulder-girdle muscles and this results in voluntary restriction of shoulder motion in the early postoperative period. Patients must be encouraged to begin active shoulder and arm motion after operation, but elderly patients are especially likely to develop a restricted range of shoulder motion if not supervised carefully. The distal parts of the transected muscles lose their nerve supply and can atrophy to a significant degree postoperatively. Patients usually note a zone of reduced sensation in the skin on the caudal side of the incision for several months after operation.

The posterolateral thoracotomy incision traditionally has been used for most pulmonary resections (except lung biopsy), for esophageal operations, and for the approach to the posterior mediastinum and vertebral column (Fig. 16-5). The skin incision is begun at the anterior axillary line just below the nipple level in the male and at the corresponding position in the female. The incision extends posteriorly below the tip of the scapula and ascends midway between the vertebral border of the scapula and the spinous processes of the vertebrae. To expose the thoracic cage, it is necessary to divide the latissimus dorsi and divide or retract the serratus anterior, trapezius, and rhomboid major muscles. The pleural cavity is entered by dividing the intercostal muscles in the chosen interspace, most often the fifth intercostal space, which provides good general intrathoracic exposure. Division of the rib posteriorly before the mechanical rib spreader is put in place helps avoid accidental fracture of one or more ribs or a costochondral separation by the instrument. Injury to the rib or cartilage may increase postoperative incisional pain and prolong restricted motion of the chest cage.

The anterolateral thoracotomy has some advantages in trauma victims who are hemodynamically unstable. The incision allows rapid entry into the chest with the patient supine or semisupine on the operating table. This is tolerated better than the lateral decubitus position and gives the anesthesiologist maximum control over the patient's cardiorespiratory system. The incision may be used for mediastinal operations, for some cardiac procedures, and for wedge resections of the upper and middle lobes of the lung. The preferred approach is a submammary skin incision starting at the sternal border overlying the fourth intercostal space and extending to the midaxillary line. The pectoralis major muscle and part of the pectoralis minor are divided at the level of the fourth or fifth intercostal space, and the incision is extended into the serratus anterior. By extending the intercostal muscle incision posteriorly along the top of the subjacent rib, it is possible to obtain a wider opening in the chest than the length of the skin incision would suggest. In an emergency situation, more exposure to the lung and mediastinum can be obtained by transecting the sternum.

The routine use of double-lumen endotracheal tubes makes it possible to use a less extensive midlateral thoracotomy incision. This incision has evolved from the transaxillary approach through the bed of the third rib that was used extensively in some clinics for upper-lobe biopsies for resection of small apical pulmonary blebs and pleural abrasion in patients with recurrent pneumothorax, for upper thoracic sympathectomy, and for biopsy of upper mediastinal lymph nodes or masses. By moving down the lateral chest wall several ribs, and with the advantage afforded by single- lung anesthesia, good exposure can be obtained for most pulmonary resections and hilar dissections. The incision has the advantage that it requires cutting no major muscles, can be made and closed rapidly, and results in less postoperative discomfort. An important requirement for adequate exposure in the incision is proper positioning of the patient. The patient is placed in a straight lateral position with the arm at right angles (to facilitate mobility of the scapula). The skin incision parallels the course of the fifth rib, extending from a few centimeters anterior to the middle of the lateral border of the scapula forward toward the submammary fold. The latissimus dorsi is elevated along its entire anterior border, as is the pectoralis major along its axillary border. The serratus anterior is separated from its insertion into the fifth rib, which is removed after the periosteum is stripped. Two rib retractors are placed at right angles to one another, one retracting the two muscle groups anteriorly and posteriorly and one

retracting the ribs caudad and cephalad (Fig. [16-6](#)).

A hazard that is common to all the lateral thoracotomy incisions is the potential for injury to the brachial plexus and the axillary neurovascular structures from excessive displacement of the shoulder in positioning the patient on the operating table after anesthesia has been induced. By preventing posterior displacement of the shoulder, this complication can be minimized.

The thoracoabdominal incision combines an upper abdominal incision with an incision in a lower intercostal space (sixth, seventh, or eighth) that may be carried as far posteriorly as the posterior axillary line. The costal margin and diaphragm are divided to provide an extensive exposure of the upper part of the abdomen and the retroperitoneal and posterior thoracic structures. Prolonged pain associated with incomplete healing of the costal margin, as well as complicated wound management involving two body cavities if infection occurs, has reduced the enthusiasm for this incision. Although elective use of this incision is becoming less common, it is still appropriate for some operations involving retroperitoneal structures (kidney, thoracoabdominal aorta) and for hepatic or thoracoabdominal trauma under emergency conditions.

A bilateral anterior thoracotomy incision with transection of the sternum ("clamshell" thoracotomy) was the standard operative approach to the heart and mediastinum before experience was gained with the median sternotomy. It is the preferred incision for double-lung transplantation and for resection of bilateral pulmonary metastases. It also is useful in circumstances in which the instruments necessary to perform median sternotomy are not available and there is urgent need to have access to both hemithoraces. It provides some cosmetic advantage in young women because bilateral submammary incisions leave less disfiguring scars than the median sternotomy. A median sternotomy also can be carried out through a submammary incision with large skin flaps, although hypesthetic nipple is a frequent complication of this approach.

The median sternotomy incision provides optimal exposure for anterior mediastinal lesions, and it is the principal incision used for cardiac operations. Either pleural cavity may be entered or incision into the pleural cavity may be avoided if it is unnecessary. Disadvantages of the incision include an increased risk of infection if it is necessary to do a tracheostomy within a few days after operation and the protracted course that occurs with infection because of involvement of the sternal fragments. An occasional patient who develops an acute wound infection also develops a severe mediastinitis associated with dehiscence of the sternal wound. The morbidity of this complication is high but has decreased with the evolution of effective treatment, particularly including rotation of pectoralis or rectus abdominis muscle flaps.

The skin incision extends from just below the suprasternal notch to a point several centimeters below the xiphoid process (Fig. [16-7](#)). A saw or, if not available, a Lebsche knife and mallet are used to split the sternum. A mechanical retractor is used to spread the incision, but care is taken not to fracture the sternal halves or injure the C₈-T₁ component of the brachial plexus by excessive spreading of the retractor blades. Postoperatively, a sternotomy is associated with less pain and interference with pulmonary function than a lateral thoracotomy.

The rigid thoracoscope was used for many years, particularly in Europe, for diagnostic visualization of the pleural space. The recent development of videoendoscopic technology revolutionized thoracoscopy, making it possible to perform therapeutic as well as diagnostic procedures. *Video-assisted thoracic surgery (VATS)* has become an accepted approach to diagnosis and treatment of pleural effusions, recurrent pneumothoraces, lung biopsies, resection of bronchogenic and mediastinal cysts and esophageal myotomy. However, after an initial wave of enthusiasm, thoracic surgeons have largely abandoned VATS for major cancer resections in favor of more traditional open incisions.

VATS is performed via two to four "port" incisions measuring 1 to 1.5 cm in length to allow insertion of the videothoracoscope and various instruments. The location of these incisions varies depending on the type of procedure being performed. Occasionally, a 8-cm non-rib-spreading intercostal "access" incision is added to improve exposure or allow insertion of additional instruments.

At the conclusion of any thoracic operation, the pleural cavity is drained with one or two chest tubes connected to an underwater seal system. Each tube is brought out through a separate stab incision in the chest wall at least two interspaces away from the incision. If the pleural cavity is not entered in operations through a median sternotomy, it is advisable to drain the mediastinum for at least 24 h with an intercostal tube that is brought out through a stab incision in the epigastrium.

PATIENT EVALUATION

Because all operations on the chest result in some short-term respiratory dysfunction and many require partial removal or permanent alterations in function of intrathoracic organs, the surgeon must make a careful assessment of the patient's ability to withstand the contemplated procedure. This assessment includes most components of the patient's overall state of health.

The initial patient encounter is by far the most important step in preoperative assessment. Major predictors of postoperative complications include the extent of resection, pre-existing cardiopulmonary disease, age, and other co-morbid conditions (e.g., diabetes, other systemic illness, immune-depressive diseases and drugs, obesity, and recent weight loss). Most of the first encounter with the patient should be devoted to accurate history taking. This initial step not only suggests the diagnosis but also determines the indications and need for a surgical intervention. It also determines the suitability of the patient to undergo such a procedure, the extent of co-morbid disease, and the need for additional testing.

If the patient is young and in good health, little further evaluation may be necessary. However, because most candidates for thoracic operations are middle-aged or older, are former or active smokers, and have some symptoms or signs of *chronic obstructive pulmonary disease (COPD)* or heart disease, careful inquiry and examination for evidence of preexisting impairment in pulmonary or cardiac reserve is essential. The patient who can climb two flights of stairs at a steady pace without dyspnea or wheezing probably has the strength, endurance, and reserve for an uncomplicated postoperative course. Any history of exercise intolerance, dyspnea on exertion, wheezing, smoking, or productive cough or physical findings of obesity, clubbing, or cardiopulmonary disease require additional evaluation and preoperative preparation.

All patients undergoing thoracotomy experience some compromise in their cardiopulmonary function postoperatively, particularly if lung tissue is resected. For those with preexisting impairment, this additional loss of pulmonary reserve may result in serious postoperative problems unless the patient is carefully evaluated preoperatively in order to give the surgeon reliable guidelines about how much pulmonary tissue the patient can safely lose.

Pulmonary Function

The major consideration in evaluating the potential thoracotomy patient is whether or not the patient's pulmonary function is adequate to tolerate the operation, the stress of the postoperative period, and the long-term functional demands. When the planned operation will result in a loss of functioning pulmonary tissue (lobectomy, pneumonectomy), there is a risk of postoperative respiratory insufficiency if the patient's preoperative pulmonary function is already compromised.

Complete pulmonary function tests including lung volumes, spirometry, diffusion capacity, and arterial blood gas determinations provide the critical information used by thoracic surgeons to determine the feasibility of pulmonary resection. Fig. [16-8](#) illustrates the subdivisions of lung volume in relation to a spirographic tracing and Table [16-1](#) indicates typical values in pulmonary function tests.

Pulmonary Mechanics

Vital capacity (VC), the amount of air that can be forcefully expelled from a maximally inflated lung position, is a useful determination if its limitations are understood. The predicted value decreases with age, and its measurement requires full cooperation from the patient. The VC is an indicator of the volume of air that a patient is able to move, but the ability to move postoperative secretions is more closely related to the velocity of forced air movement. The forced expiratory volume in 1 s (FEV₁) is a dynamic measurement of a patient's ability to move volumes of air during units of time. The FEV₁ usually is reported as a percentage of the VC (FEV₁/VC) as well as an actual volume. It is important to note its value both ways. If the VC is significantly reduced by restrictive lung disease (e.g., pulmonary fibrosis), the FEV₁/VC ratio may be satisfactory, whereas the actual volume exhaled may be markedly abnormal. While the VC is near normal in patients with moderately severe obstructive airway disease, it is reduced in individuals with restrictive pulmonary disease and those weakened by neuromuscular disease. The FEV₁ is reduced in obstructive airway disease, but the degree of reduction may vary from day to day or week to week in the same individual and may be improved by medications and pulmonary rehabilitation programs. The FEV₁ is a useful test to monitor patients with marginal pulmonary function who are being prepared for operation by aggressive respiratory therapy programs (See Fig. [16-8](#)).

Blood-Gas Determination

Measurement of the arterial blood gases and pH is routine in the preoperative evaluation of a candidate for thoracic surgery. An occasional patient is discovered to have hypoxemia or CO₂ retention that was not suspected on the basis of clinical examination or spirometry. A measurement of the arterial carbon dioxide pressure (PaCO₂) provides an immediate indication of the patient's alveolar ventilation; any value above 46 torr means that there is hypoventilation. There are multiple causes for this, and the specific reason should be sought in each patient. The ability of the lungs to excrete CO₂ is remarkable, and any persistent elevation in a patient otherwise considered a candidate for a thoracotomy suggests severe chronic lung disease and serious abnormalities in distribution of ventilation and perfusion. Most operations temporarily increase the ventilation-perfusion abnormality. A mild elevation of the PaCO₂ in a patient with chronic lung disease may be treated aggressively with measures to improve pulmonary function and allow the patient to be considered for operation. If pulmonary resection is contemplated in such an individual, the risk of postoperative respiratory failure is high, and the decision to operate may depend on whether functioning pulmonary tissue (versus nonfunctioning) is to be removed.

The measurement of arterial oxygen pressure (PaO₂) is valuable in the preoperative assessment of pulmonary function but must be viewed with a consideration of the possibilities for error in its measurement. At sea level, the normal PaO₂ is greater than 85 torr. The majority of patients with *COPD* have a PaO₂ of 80 torr or below, and values in the range of 70 to 80 torr are not associated with significant postoperative problems. *If the PaO₂ is less than 70 torr, an attempt should be made to determine the cause and to improve the patient's gas exchange preoperatively. The most frequent cause is uneven distribution of ventilation and perfusion (/ mismatch), but other possibilities include right-to-left shunting as a result of the thoracic disease for which the patient is being considered (e.g., shunting through a lung cancer or a non-functional lobe). More sophisticated pulmonary function tests may be required, including determination of alveolar-arterial oxygen difference, calculation of right-to-left shunt fraction, and split pulmonary function (/ scanning).*

In assessing the pulmonary reserve in patients in whom lung resection is contemplated, determination of the amount of lung tissue that can be removed safely is important. The surgeon must begin the operation knowing the maximum amount of lung tissue that can be removed without leaving the patient a pulmonary cripple. The *split-function test* of lung function is quantitative radionuclide ventilation-perfusion scanning for regional lung function (/ scans) and has been especially helpful in making this decision. The perfusion data are obtained by comparing the counts over each lung during ^{99m}Tc perfusion scanning. Perfusion scans can then be correlated concurrently with ventilation scanning to measure regional function.

The practical value of such split-function studies is that postoperative VC and FEV₁ can be predicted for the patient who may require resection of a portion of lung neoplasm (predicted postoperative FEV₁ = preoperative FEV₁ minus the percentage perfusion in the involved lung). A patient may require only a lobectomy, but the acute effects of a major

thoracotomy reduce lung function to a greater extent than this because of some congestion in the remaining lung, restriction of chest wall motion, and incisional pain. In the early postoperative period, pulmonary function can be reduced by as much as 50 percent. This is particularly true if there is significant preoperative reduction in pulmonary function. An important prospective study by Boysen and associates demonstrated the validity of the split-function concept in a group of patients with impaired ventilatory function (preoperative FEV₁ <2.0 L). If the predicted postoperative FEV₁ exceeded 800 mL, the patients were considered acceptable candidates for pulmonary resection up to and including a pneumonectomy. The perioperative mortality in this series was 15 percent, a figure no longer considered acceptable for major pulmonary resections even in extremely high-risk patients. Other investigators have corroborated their data and demonstrated that the measured values for FEV₁ after a pneumonectomy correlated closely with the predicted values. Fig. 16-9 illustrates the use of a split-function study to decide that a patient with severely decreased pulmonary function is a reasonable risk for pneumonectomy.

Testing patients' ventilation and gas exchange during maximal effort and exercise also is important. Previously, maximum voluntary ventilation (MVV) was used. This test is performed by having a patient inhale as deeply and rapidly as possible for up to 15 s. The MVV measures the status of respiratory muscles, the compliance of the lung-thorax system, and airway resistance. In the past, surgeons depended heavily on the results of the MVV because patients with an MVV less than 50 percent of predicted experienced a prohibitively high operative mortality after pneumonectomy.

The carbon monoxide diffusion capacity (DLCO) is increasingly recognized as an important parameter of pulmonary reserve. The DLCO is decreased by severe chronic obstructive and restrictive pulmonary disease. It also can be decreased by some chemotherapy drugs (e.g., bleomycin, mitomycin, and perhaps paclitaxel) and occasionally by radiation. Measurement of the DLCO is particularly important in cancer patients who have received chemotherapy or radiation preoperatively. A DLCO of less than 50 percent of predicted is associated with high risk of postoperative pulmonary complications, and a DLCO of less than 30 percent of predicted generally precludes consideration of a thoracotomy and any pulmonary resection.

In recent years there has been interest in exercise testing for patients who are candidates for pulmonary resection but have impaired pulmonary function. By combining respiratory gas analysis with ergometric testing, oxygen consumption can be correlated with work capacity. These tests measure maximal oxygen uptake (MVO₂). This is particularly useful in evaluating patients who have reasonable exercise capability despite severe obstructive airway disease. Patients with MVO₂ measurements of less than 10mL/kg/min have prohibitive risk of postoperative complications. By using these tests, patients can be selected who are "good" risks to tolerate thoracotomy and resection even with significant impairment on spirometry.

Unilateral balloon occlusion of the pulmonary artery with right-sided heart catheterization is rarely indicated in the preoperative evaluation of patients who require a major pulmonary resection. Normally, after pneumonectomy, the remaining lung accepts the entire pulmonary blood flow without development of pulmonary hypertension. In an occasional patient, occlusion of one pulmonary artery (simulating a pneumonectomy) results in pulmonary hypertension to levels above 30 torr, and this has been correlated with excessive mortality after pneumonectomy. This test is rarely done and only in circumstances where there is conflicting information from other pulmonary function tests.

Numerous studies have shown that there is no data-analysis technique that absolutely separates the operable patient from the inoperable. Instead, the goal of preoperative evaluation is to separate patients into low- and high-risk groups. Of the standard function tests, surgeons have come to place the greatest reliance on the FEV₁, FEV₁/FVC ratio, DLCO, and MVO₂ as the best determinants of operability for the patient with reduced function from respiratory disease, advanced age, or chronic illness.

Although studies have failed to support a precise correlation between a given level of reduced pulmonary function and surgical outcome, guidelines suggested in a recent review by Miller and colleagues predict which patients being evaluated for resective surgery are at high risk (Table 16-2).

The continuing improvements in facilities and technology for postoperative care have allowed surgeons to become more liberal in recommending thoracotomy to selected patients with compromised pulmonary function. Several recent reports provide encouraging indications that suitably selected and well-managed patients with FEV₁ of less than 70 percent of predicted expect satisfactory results after lung resection if pulmonary function is preserved by aggressive perioperative management and pain control and, when necessary, limiting resections to the wide local removal of tumors (wedge resection or segmentectomy) rather than lobectomy or pneumonectomy.

Another important facet of preoperative management is to provide the patient with whatever physical and mental preparation is needed. Many patients are smokers. They must be persuaded to stop smoking before operation, preferably for 2 weeks or more. The use of nicotine substitutes is helpful. Surgeons agree that the character and amount of bronchial secretions have a major impact on postoperative morbidity. Aggressive attention to reducing the amount and tenacity of the secretions must be made before the operation. Any pre-existing pulmonary infection should be identified and treated intensively using appropriate antibiotics, bronchodilators when indicated, and chest physiotherapy.

It has been shown that the patient's attitude toward their disease, the desire to have a favorable outcome, and confidence in their doctors are all predictive of success. The initial patient encounter is extremely important in developing these favorable conditions.

Cardiac Evaluation

In reviewing the epidemiology of postoperative cardiac events, Mangano and colleagues have pointed out that 1 in 8 of the 25 million patients in the United States who undergo noncardiac surgery have coronary artery disease and that 50,000 of these will have perioperative myocardial infarctions. Half the 40,000 annual postoperative deaths in this country are the result of cardiac disease. Careful preoperative evaluation allows the surgeon to identify those at highest risk and, in collaboration with the anesthesiologist, to protect them by appropriate medications and intraoperative monitoring.

The need for preoperative cardiac evaluation is based on the expected postoperative demand for increased cardiac output, the frequency of coincidental cardiac disease in the elderly smoking population, and the likelihood of cardiopulmonary complications after thoracic operations. Because cardiac symptoms are sometimes masked by the symptoms of the primary thoracic disease for which the patient is being considered, the screening evaluation must be precise. A preoperative electrocardiogram (ECG) should be obtained. If the history, examination, or ECG reveals any abnormality, consultation with a cardiologist may be appropriate. Exercise or pharmacologic stress tests and gated radionuclide angiocardiograms provide excellent noninvasive evaluation of suspected myocardial ischemia and of ventricular functions such as cardiac output, stroke volume, and ejection fraction. Rest and stress echocardiography also is useful in detecting valvular disease, in estimating ventricular function, and in assessing a patient's ability to respond appropriately to the stress of a pulmonary resection. For patients with evidence of significant coronary artery disease, it may be necessary to proceed to coronary arteriography and even angioplasty or coronary stenting before planned thoracotomy (Fig. 16-10).

Other Organ Systems

Any preoperative evaluation should include screening tests for renal and hepatic function even in the absence of historical data that suggest disease of those organs. Under ordinary circumstances, a hospital admission biochemical profile that includes measurements of blood urea nitrogen, creatinine, serum proteins, transaminases, lactate dehydrogenase, alkaline phosphatase, and bilirubin is an adequate initial screening test. The discovery of any abnormality mandates a more detailed evaluation.

Malnutrition increases the morbidity and mortality rate of any major surgical procedure, and an assessment of the preoperative nutritional state is important. In clinical practice it may be difficult to separate the effects of total calorie deficit from a deficiency of protein alone. There is a reduced blood volume and reduced tolerance for intraoperative bleeding in hypoproteinemic patients. Impaired antibody production, reduced host resistance to infection, decreased lymphocyte proliferative response, and depression of the delayed skin reactivity to antigens are associated with weight loss and hypoalbuminemia.

Particularly important in thoracic surgical patients are the adverse effects of protein depletion on pulmonary functions and ventilatory capacity. As skeletal muscle is catabolized during starvation, the muscle groups in the thorax, abdomen, shoulder, and diaphragm involved in respiration and coughing share in the unselective loss of strength that is seen in all muscles, and this results in an increased risk of a major thoracic operation.

Cardiopulmonary Rehabilitation

Patients who are malnourished, chronically ill, or have chronic cardiopulmonary diseases can benefit from cardiopulmonary rehabilitation before surgery. There is no evidence that delaying surgery to improve function adversely affects cancer therapy. One, two, or three weeks of intensive preoperative rehabilitation using modern cardiopulmonary exercise techniques supervised by competent therapists can markedly improve cardiopulmonary function. Surgeons frequently fail to take advantage of these simple maneuvers that can significantly decrease the operative risk.

POST-OPERATIVE CONSIDERATIONS

The complications that develop most often after thoracic operations are pulmonary. These also are the most common causes of postoperative death. Although postoperative cardiac events, particularly supraventricular arrhythmias, are common, they rarely lead to significant mortality after thoracic surgical procedures.

Pulmonary Function Changes

Significant physiologic changes in pulmonary function occur after major thoracic and upper abdominal operations. Each produces similar changes in pulmonary function and has similar initial complications; with lung resections, some of these changes may be permanent. The magnitude of the changes is affected by preexisting pulmonary disease, the length of the operation, the use of postoperative analgesics, and immobilization in bed. Patients without pulmonary disease develop similar changes and are subject to similar complications, although the risks are much lower. The pulmonary changes occur in (1) lung volumes, (2) ventilatory patterns, (3) respiratory gas exchange, and (4) defense mechanisms.

Lung Volume

Total lung capacity and each of its subdivisions are significantly reduced after abdominal or thoracic operations. *VC is reduced by 25 to 50 percent or more, with the maximum reduction occurring during the first 4 days after operation. Similarly, functional residual capacity and expiratory reserve volume are decreased, with a gradual return toward normal beginning in the second week after operation. The reduction in lung volume is often accompanied by an increase in the closing volume (early airway closure) and potentiates the development of atelectasis. If a pulmonary resection has been performed, the magnitude of change is even greater and is proportional to the amount of functioning lung that was removed.*

Ventilatory Pattern

The sedative effect of the anesthetic agent and the postoperative analgesics, combined with the severe pain of the thoracotomy incision, produces sharp reductions in tidal volume after operation. The expected response is an increase in respiratory rate sufficient to maintain minute ventilation. Parenteral narcotics ordinarily used to manage postoperative pain all depress the respiratory center, inhibiting the compensatory rate increase and leading to carbon dioxide retention and hypoxemia. Nonsteroidal anti-inflammatory agents (e.g.,

indomethacin or ketorolac) often can provide pain relief nearly equivalent to narcotics without respiratory depression but also may have significant gastrointestinal and renal toxicities.

Another important effect in changes in ventilatory pattern is the sharp reduction in or elimination of the normal periodic hyperinflations (sighs). Normal adults sigh at the rate of nine or ten times per hour under quiet conditions. With loss of periodic hyperinflations, there is closure of lung units resulting in atelectasis and a reduction in compliance.

Gas Exchange

Decreases in PaO₂ and mild elevations of PaCO₂ are frequent as patients recover from anesthesia. PaCO₂ generally returns to normal or below normal in the early postoperative period, whereas PaO₂ can remain depressed during the first week. The factors responsible for the reduction in the PaO₂ include abnormal ventilation-perfusion relationships and intrapulmonary shunting associated with atelectasis.

Pulmonary Defense Mechanisms

The lung is normally protected against inhaled particulate matter and microbes by several mechanisms. The cough reflex defends the upper airways against inhaled or aspirated material in the tracheobronchial tree. Clearance of inhaled particles and microbes from the lower airways depends on the mucociliary system, and the alveoli are defended by mucociliary transport, lymphatic drainage, and alveolar macrophages. Because coughing is inhibited by several mechanisms in the postoperative period, there is significant impairment of that defense mechanism. Ciliary function is decreased, and multiple factors, including arterial hypoxemia, depress the activity of alveolar macrophages. Finally, the composition and physical properties of mucus are altered in a way that reduces the effectiveness of the mucociliary transport system.

Preventive Measures

Many of these changes in postoperative ventilation can be minimized or prevented by preoperative optimization of the patient with regard to cardiopulmonary function using appropriate drug treatment (antibiotics, mucolytics, bronchodilators, etc.) and rehabilitation. In the postoperative period, pain relief is important. In the past two decades, many pain-relieving measures have been introduced in an attempt to provide the patient with a painless postoperative course, and the routine use of these is critical to preventing postoperative complications.

Pain Control

In the first few postoperative days, effective management of incisional pain is of central importance in the maintenance of adequate ventilation. The pain that accompanies a thoracotomy incision is severe and disabling. Unless well managed, it will cause hypoventilation, retention of secretions, atelectasis, hypoxia, hypoxemia, shallow and ineffective respiratory effort, and pneumonia. It is a constant challenge to find the delicate balance between giving patients enough pain medication so that they are able to cough without giving them so much that they lose their initiative to do so.

Pain is frequently managed by parenteral narcotics administered intramuscularly or intravenously on a fixed schedule (by-the-clock) or on-demand (p.r.n.). Particularly when given on demand, use of these narcotics can be associated with swings in patient alertness ranging from obtundation with respiratory depression and suppression of cough to severe pain and anxiety. When parenteral narcotics are used as the primary means of postthoracotomy pain control, they are best administered by a continuous intravenous infusion carefully regulated by observation or by a patient-controlled analgesia (PCA) device. These devices allow adequate continuous pain relief without undue somnolence. Success with this approach requires careful patient education and close nursing care. Nausea, vomiting, and respiratory depression remain potential side effects.

The search continues for the optimal way to control incisional and pleural pain without the deleterious consequences of systemic narcotics. Either short-acting (lidocaine) or long-acting (bupivacaine) local anesthetic agents can be given as one-shot intercostal blocks before closing the chest or by intermittent or continuous injections through catheters left in place in the pleural space at completion of the thoracotomy. When these methods are used, great care must be taken to avoid inadvertent intravascular or subdural injection. Severe vasomotor hypotension is occasionally reported with this technique, and the patient should be monitored closely whenever it is used. Some advocate use of intercostal nerve cryoanalgesia and have reported excellent pain relief by this approach (each nerve receives one 30-s exposure to the probe). Long-term intercostal neuralgia can be a complication of this technique.

There are opiate receptors in the spinal cord with specific endorphins and enkephalin mediators. The use of continuous epidural infusion of preservative-free morphine, hydromorphone, or fentanyl administered via a thoracic or lumbar epidural catheter provides excellent pain control. Although well-controlled studies have not shown that epidural analgesia improves pulmonary function, it is considered in many centers to be the optimal approach to pain control.

Advances in anesthesia techniques, particularly improvements in the design of double-lumen endotracheal tubes, offer the surgeon other ways to reduce postthoracotomy pain. Wide exposure for delicate hilar dissection was necessary when the dissection was carried out around a retracted but filled, moving lung. The ability to work in the chest with a fully deflated lung encourages surgeons to avoid unnecessary rib spreading and to seek other, less traumatic means for entry into the chest. Urschel has reported a large series of lobar resections in both chest cavities through a median sternotomy. The extensive experience with this incision in the open-heart surgery population has demonstrated that it is much less painful and much better tolerated physiologically. Others have developed muscle-sparing incisions to allow major resections without the need for dividing the muscles to the shoulder girdle. This incision is rapidly opened, rapidly closed, and leaves a chest wall with more functional integrity. Postoperative pain is less. The combination of these less traumatic incisions and intrapleural bupivacaine infusions or epidural analgesia as an approach to pain control is gaining favor worldwide.

Complications

Pulmonary complications of thoracic operations have their origins in these postoperative changes in physiology and usually begin in the operating room or soon thereafter. The principal pulmonary complications include atelectasis and respiratory infections.

Atelectasis

Atelectasis begins as closure of lung units, causing microatelectasis, a diffuse sublobular form not visible on chest x-rays, and macroatelectasis, the collapse of a segment, lobe, or entire lung. The three mechanisms that are considered responsible for atelectasis are accentuations of the postoperative pathophysiologic changes described earlier: (1) retained bronchopulmonary secretions, (2) decreased sighing, and (3) decreased expiratory reserve volume.

Retention of secretions is a major cause of atelectasis in patients with chronic bronchitis. It is more subtle in patients with normal lungs, although they also develop either microatelectasis or macroatelectasis. Decreased sighing and reduced tidal volume contribute to the reduced compliance in the postoperative period. Unless reversed by voluntary efforts at deep breathing, induced coughing, or attentive respiratory care, these changes will contribute to the development of atelectasis. Similarly, the postoperative reductions in lung volumes are related to airway closure that is associated with the changes in ventilatory pattern.

Pneumonia

Postoperative bronchopulmonary infectious complications consist of tracheobronchitis and pneumonitis. While these complications also occur in “normal” patients, their incidence is higher in patients with preexisting chronic airway disease. Decreased cough, atelectasis, reduced mucociliary clearance of inhaled particles and bacteria, pain, and analgesic drugs all contribute to these infectious complications. Interference with the mucociliary clearance mechanism leads to rapid bacterial proliferation distal to obstruction in an area of atelectasis. Of equal importance has been the demonstration that the respiratory tract becomes colonized with gram-negative bacilli, particularly in the presence of tracheal intubation, coma, hypotension, hypoxia, acidosis, and azotemia. Many of these conditions exist in the postoperative period in patients subjected to major thoracic procedures.

Cardiovascular Complications

Postoperative myocardial infarction, cerebrovascular accident, or peripheral arterial embolus or thrombosis occur infrequently after pulmonary resection. These complications can be avoided if patients are carefully evaluated preoperatively for cardiovascular disease and placed on appropriate medications.

The most common cardiac complication is supraventricular arrhythmia, usually atrial fibrillation. This occurs in approximately 17 percent of lobectomy patients, 25 percent of pneumonectomy patients, and 33 percent of extrapleural pneumonectomy patients. The risk of postoperative atrial arrhythmias increases with age, being more common over the age of 60 years. Digoxin, previously used routinely to prevent or treat post-thoracotomy atrial arrhythmias, has proved in prospective studies to be ineffective in this regard. Calcium channel blocking drugs, especially diltiazem, provide effective treatment and may reduce the frequency of this complication.

Deep venous thrombosis (DVT) is another potential complication after thoracotomy. As for the other major operations, the frequency of DVT is related to the type and duration of the operation, prior history of DVT, the patient's ability to resume ambulation postoperatively, and the patient's other medical problems (e.g., obesity). Routine DVT prophylaxis with intermittent venous compression boots or with subcutaneous heparin is an integral part of perioperative care for patients undergoing major thoracic procedures.

Other Complications

Other complications occur less frequently after thoracotomy. Renal insufficiency, usually transient, can develop in patients with underlying renal disease who receive nephrotoxic drugs (aminoglycoside antibiotics, non-steroidal anti-inflammatory agents). Wound infections and empyemas occur in less than 5 percent of patients. Routine perioperative antibiotic prophylaxis with cephalosporins reduces the risk of such postoperative infections.

THORACIC INJURIES

General Considerations

The leading cause of death, hospitalization, and short- and long-term disability for all ages from the end of the first year through the forty-fifth year of life is trauma. Twenty-five percent of all trauma deaths are a result of chest injuries alone; respiratory problems contribute significantly in 75 percent of traumatic deaths. The respiratory system works on a small margin of safety, and physical disruption of the integrity of the system, which is expected in trauma, must be corrected rapidly, or irreversible damage resulting from hypoxia in vital structures will occur. With complete loss of oxygen and with normal tissue oxygen demands (e.g., normal ambient temperature), such damage occurs after only 4 min. Lesser disruptions of delivery extend the tolerance time, but it should be apparent that accurate diagnosis of the failing or disrupted elements of the system and their prompt correction are delayed at progressive peril to the patient. Eighty-five percent of patients with life-threatening thoracic injuries can be managed by simple interventions.

General considerations about the management of the injured patient are discussed in Chap. 6. The preinjury cardiopulmonary functional status must be considered carefully. The rib fracture or modest pneumothorax that is well tolerated without hospitalization in a healthy young patient may lead to pneumonia, empyema, or death in an older patient with chronic airway disease. The patient with preexisting cardiac disease is particularly vulnerable to the development of pulmonary edema and hypoxia that can occur with the rapid administration of intravenous fluids during resuscitation.

A rapid, overall evaluation of the patient must be carried out to determine the severity of the injury. Chest trauma often is associated with other injuries, and the overlap of the upper part of the abdomen by the thoracic cage provides a border zone that is difficult to assess and often is the site of combination injuries. Patients who cannot describe symptoms because of associated head injuries or profound shock are particularly challenging to evaluate.

Types of Injuries

Injuries to the chest are usually classified according to the type of insult that caused the damage. The injury often is influenced by the setting (military or civilian, urban or rural) in which it occurred. Most military injuries are high-velocity penetrating wounds. Low-velocity gunshot wounds are replacing knife wounds as the most common in urban civilian populations. Blunt injuries from motor vehicular or occupational (e.g., logging) accidents make up the majority of nonurban injuries. Penetrating wounds are becoming more frequent in suburban and rural areas as violent crimes increase, and blunt trauma also occurs commonly in urban areas.

The mortality rate of major blunt injuries has decreased steadily during the past quarter century, but complications and death associated with pulmonary contusion, posttraumatic pulmonary insufficiency, and trauma to the heart and great vessels are significant. It is characteristic of blunt injuries that the maximal extent of cardiopulmonary functional loss and often the complete diagnosis require several days to become fully evident.

Penetrating wounds of the lower thoracic region are treacherous. The diaphragm normally rises to the level of the nipples during expiration, and penetrating trauma to this area can injure the subdiaphragmatic viscera. Some surgeons believe that a stab wound of the left lower part of the chest mandates early abdominal exploration, because the knife may have injured the spleen, stomach, or colon. The abdominal findings in these patients may be overshadowed by more obvious chest injuries and can be missed by initial peritoneal lavage. The consequence of error is costly. Because the liver usually is the abdominal structure injured by right-sided stab wounds that penetrate the diaphragm, it may be reasonable to delay exploration until the patient is more stable. Early abdominal exploration is indicated if there is evidence of continuing blood loss. The late consequence of diaphragmatic lacerations justifies operative repair as soon as it is safe to do so.

Whatever the cause, the principles of management should remain focused on the mechanical systems involved: the pump, the hydraulics, and the bellows (the suction-blow system that draws atmospheric air into the alveoli and expels it). The pump must be working, and the vessels must have the integrity and suitable contents to transport the gases to and from the tissues.

Conditions Requiring Urgent Correction

Airway Obstruction

Most patients with major disruption of the airway leading to obstruction will not survive the initial accident; the leading cause of death at the accident site is airway obstruction. At any stage of the early resuscitation and transportation of the patient, correctable airway obstruction may occur. The oropharynx should be cleared of mechanical debris and the chin and neck positioned to facilitate opening the posterior pharynx. Until the stability of the cervical vertebrae has been ascertained, the neck should be positioned only by an anterior chin-thrust motion while applying continuous cephalad traction (Fig. 16-11). If the upper airway remains at risk after clearing and positioning, tracheal intubation is indicated. If it can be done safely, orotracheal intubation with a size n 8-mm or larger endotracheal tube is indicated. If cervical spine injury is suspected and midface soft- tissue damage is not extensive, nasotracheal intubation is preferred, even though the small caliber of the nasotracheal tube may prevent subsequent flexible bronchoscopic examination. Fiberoptic orotracheal intubation performed over the flexible bronchoscope is another safe method of obtaining control of the airway in patients with fractures of the cervical spine or the cribriform plate.

If the equipment or expertise is not available, or if the upper airway injury precludes safe access to the vocal cords from above, cricothyroidotomy should be performed. If high-pressure oxygen is available, catheter jet ventilation may be used (percutaneous passage of cricothyroid membrane catheter, 12-gauge or larger) while the patient is being stabilized and arrangements for tracheal intubation are being completed.

Tension Pneumothorax

When an injury to the lung parenchyma has occurred that allows air to enter the pleural space with each respiratory effort, and when the flap- valve effect of the injury prevents that air from reentering the bronchial tree for egress through the trachea during expiration, tension develops within the pleural space until equilibration with the negative pressures the patient is able to generate is reached; at that time, effective ventilation ceases and venous blood can no longer enter the chest. The mechanics of a developing tension pneumothorax may not be obvious when the patient is first seen. Pain may be the primary complaint, with no evidence of respiratory distress. However, if the lung wound is behaving as a check valve, some air escapes into the pleural cavity with each inspiration or with each cough. Gradually, intrapleural pressure builds up, the lung collapses, and tension pneumothorax can develop. A shift of the mediastinum and compression of the large veins result in a decreased cardiac output that may lead to sudden death.

The diagnosis should be made instantly by observation of a patient with dilated neck veins making respiratory effort but not respiratory motions and unable to move air. It is immediately confirmed by the hyperresonant percussion note over the injured hemithorax and absent or distant breath sounds. The immediate release of the tension by placement of a large-bore needle followed immediately by insertion of a thoracostomy tube is lifesaving.

Open Pneumothorax

The sucking chest wound is one in which a full thickness segment of the chest wall has been destroyed and the negative intrapleural pressure sucks air directly through the chest wall defect rather than through the trachea into the alveoli. It occurs most commonly after shotgun blasts, explosions with flying debris, or impalement injuries. It may or may not be associated with underlying parenchymal damage.

The diagnosis can be made by noting a patient with normal or collapsed neck veins who is making respiratory motions but not moving air. Confirmation is immediate on inspection of the patient's chest and observation of the wound. The patient is stabilized by any mechanical covering over the open wound. As soon as convenient, a watertight dressing should be placed and an intercostal catheter inserted into the pleural cavity. Early debridement and formal closure of the wound should then be performed.

Massive Flail Chest

When severe blunt injury results in two-point fractures of four or more ribs, a segment of the chest wall becomes flail. On inspiratory effort, the negative pressure in the chest pulls the unstable segment of the wall inward in a paradoxical motion. The patient may be unable to develop sufficient intratracheal negative pressure to maintain adequate ventilation, and atelectasis, hypoxia, and hypercapnia occur. A patient who is conscious may splint the segment sufficiently to make it inapparent to cursory examination, but the continuing extra effort in the attempt to move air soon leads to tiring and may result in sudden respiratory decompensation. The progressing failure is aggravated by the developing pulmonary contusion that accompanies blunt trauma sufficient to break multiple ribs. In the unconscious patient, the lesion may be less dangerous because it is more readily recognized and more apt to be treated early.

In the massive flail chest, the diagnosis may be difficult unless the chest wall is visualized during the respiratory effort. If unconscious, the patient ordinarily is making vigorous respiratory motions but moving little air; the paradoxical segment should be obvious. The patient who is awake may exhibit a very rapid shallow breathing pattern at or above 40 breaths/min. Other aspects of the management of lesser flail injuries are discussed below, but when massive flail is diagnosed, endotracheal intubation and positive- pressure controlled ventilation are mandatory.

Massive Hemothorax

When 1500 mL or more of blood is acutely removed from the pleural space as a thoracostomy tube is placed, a high proportion of patients will have a surgically correctable injury requiring urgent thoracotomy. If a patient with penetrating injury or multiple rib fractures is found to have a complete hemothorax dull to percussion in association with hypotension, a chest tube should be inserted. If massive hemothorax is found, the patient should be taken directly to the operating room as blood volume resuscitation is taking place.

Conditions Requiring Urgent Thoracotomy

Continued Intrapleural Bleeding

If bleeding continues from a thoracostomy tube after initial placement at a rate exceeding 100 mL/h for 6 h or more, most surgeons would now agree that a surgically correctable lesion is present. Ordinarily it will be a bleeding intercostal vessel, because bleeding from the lower-pressure pulmonary system almost always stops when the lung is reexpanded after the pleural space is evacuated. In making a decision to proceed with surgical exploration, the rate and pattern of bleeding are more important than the amount.

Massive Air Leak

This uncommon injury usually results from steering wheel compression of the trachea against the vertebral bodies following high-speed head-on collisions. Complete disruption of the trachea or a major bronchus may occur. The injury is often fatal but may be tolerated for a brief period if the surrounding mediastinal soft tissues maintain continuity of the airway. All levels of the trachea and all major bronchi can be involved; however, more than 80 percent of these injuries occur within 2.5 cm of the carina. Patients with intrathoracic tracheal or central bronchial disruption exhibit a variety of signs and symptoms depending on whether there is free communication between the site of injury and the pleural cavity. Important pathognomonic findings are complete unilateral atelectasis in the face of a large air leak or symmetrical downward displacement of both hila. Distal injuries often result in pneumothorax, which is manageable by tube thoracostomy alone because the air leak is small. Lazar reported a case in which a complete tracheal disruption just above the carina with 6 cm of discontinuity was tolerated in a young athlete for 24 h before accurate diagnosis and repair.

Extreme care must be taken in the evaluation of patients with massive air leaks, because overly aggressive diagnostic bronchoscopy or endotracheal intubation and positive-pressure ventilation before the defect is located accurately and preparation for repair are made can result in rapid death. Occasionally, as in other tracheobronchial injuries, an injury may seal itself off and fail to be recognized until severe stenosis develops.

Other Indications

Several other important causes are listed here, but discussed elsewhere.

Acute or rapidly recurring pericardial tamponade

Acute heart failure secondary to valvular or septal injury

Widened or widening mediastinum

Perforation of the intrathoracic esophagus

Dangerous But Less Compelling Injuries

Diaphragm Rupture

Urgent repair of massive diaphragmatic rupture is necessary if high-volume herniation of abdominal contents into the chest prevents adequate ventilation. Ordinarily, however, the acute problems associated with diaphragm rupture are related to the associated injuries to abdominal viscera resulting from the force necessary to rupture the diaphragm. Penetrating trauma to the lower chest or upper abdomen and crush injuries, most often secondary to automobile accidents, are the usual causes of traumatic rupture of the diaphragm. The left hemidiaphragm is ruptured more frequently by blunt trauma than the right, at a ratio of about 9:1. The right hemidiaphragm is protected by two mechanisms: The liver on the right and the heart in the center have a buffering effect that diffuses the sudden increase in intraabdominal pressure; and cadaver studies have shown an inherent weakness in the posterolateral aspect of the left diaphragm. When rupture of the right side does occur, the liver usually is the only abdominal structure that herniates into the chest early, although gradual aspiration of the stomach into the right chest through the diaphragmatic defect can occur over time. With rupture of the left hemidiaphragm, the stomach, spleen, left transverse colon, and omentum in any combination may enter the left pleural cavity immediately. If the diagnosis is delayed for several days or longer, there is progressive displacement of the abdominal viscera into the chest or progressive gaseous distention of the herniated stomach. The latter may occur despite an indwelling nasogastric tube and can precipitate respiratory distress (Fig. [16-12](#)).

Patients with diaphragmatic rupture from blunt trauma usually have associated injuries that demand attention first and prevent a detailed initial evaluation. The first chest x-ray after rupture of either hemidiaphragm may show nothing more than a blurring of the diaphragm with or without evidence of a small hemothorax. In some patients the diagnosis is made very early because the nasogastric tube is seen to lie within the confines of the left pleural space.

Penetrating diaphragmatic injuries rarely produce early symptoms except those related to other structures that may be injured. After several months or years, gastrointestinal obstruction may develop and lead to strangulation of herniated viscera. The hole in the diaphragm is small, and herniation occurs slowly. When the diagnosis is made early, hours to days after the injury, diaphragmatic repair should be performed via laparotomy in order to manage associated intraabdominal injuries. The wound in the hemidiaphragm may vary from a simple radial tear to an extensive and complex laceration. Repair usually can be accomplished by direct suture, but a prosthetic patch of nonabsorbable material occasionally is required. If the diagnosis is delayed, a transthoracic approach is preferred because it provides better exposure to reduce the hernia, free adhesions between the abdominal viscera and intrathoracic structures, and repair the defect in the diaphragm.

Pneumothorax

Pneumothorax usually is the result of injury to the lung or the tracheobronchial tree. Esophageal perforation may be followed by a pneumomediastinum that ruptures into the pleural cavity. Whether the pneumothorax is associated with blunt injury and fractured ribs or is the result of a penetrating wound, there is a variable amount of bleeding into the pleural cavity. When sufficient blood is present to require its removal or transfusion, it is termed *hemopneumothorax*.

Pneumothorax can vary from one that is so slight that it may be missed on the initial x-ray examination to a massive, continuing air leak that displaces the mediastinum, depresses the diaphragm, and compresses the opposite lung, such as the tension pneumothorax discussed previously (Fig. [16-13](#)). A pneumothorax due to a parenchymal lung injury is self-limited because the developing lung collapse combines with blood clotting in the wound for a sealing effect. For some patients, extensive adhesions already present between the visceral and parietal pleura may localize the pleural air and prevent a collapse of the lung (Fig. [16-14](#)).

With any chest injury it is wisest to presume that a pneumothorax is present until proved otherwise. Because of pain and limited chest motion on the injured side, physical examination may be inadequate for diagnosis of a minimal pneumothorax. Since attention may be diverted to the management of other injuries, and because of the risks of tension developing should a general anesthetic with positive-pressure ventilation be given, prophylactic thoracostomy catheters usually should be placed whenever there is significant chest injury. The catheter is best placed in the midaxillary line, just above the fifth rib, after finger exploration has ensured that the pleural space is free at the site of insertion.

Treatment of the more usual pneumothorax depends on symptoms of respiratory insufficiency, the extent of the pneumothorax, and the presence of significant hemothorax. There is a tendency to think of pneumothorax in terms of the two dimensions conveyed by the anteroposterior chest x-ray. Instead, the hemithorax should be considered a modified cone, and when the lung surface is separated from the chest wall by 3 cm or more, the patient may have a 50 percent lung collapse (by volume) rather than the 25 or 30 percent collapse that the chest x-ray suggests. Rarely, when a pneumothorax is less than this amount because of a nonpenetrating injury and is not accompanied by significant blood or fluid in the pleural cavity, treatment may not be required. A decision not to remove the pleural air implies that the patient has had a minor injury and that conditions for observation are ideal. Approximately 1.25 percent of the air will be absorbed each day, with full expansion expected in 3 to 6 weeks.

Aspiration of the air with a needle and insertion of an intercostal catheter only if lung collapse recurs constitute a reasonable method of treatment advocated by some even when the pneumothorax amounts to as much as 50 percent. In all patients with collapse greater than 50 percent, in those with hemopneumothorax, and in those whose pneumothorax is the result of penetrating trauma, an intercostal catheter should be inserted and attached to a water seal with 10 to 25 cmH₂O of negative pressure. In the majority of patients, lung reexpansion and cessation of the air leak occur within a few hours or a few days. If not, a major bronchial injury may be present, and a thoracotomy may be required after appropriate diagnostic procedures.

The use of prophylactic systemic antibiotics in patients with chest trauma is controversial, and their use in cases of nonpenetrating trauma seems unjustified. The simple insertion of an intercostal catheter does not require prescribing antibiotics.

Interstitial Emphysema

Disruption of the respiratory tract at any level results in the passage of air into the surrounding tissues. Mediastinal emphysema occurs when air enters the areolar tissue planes from a tracheobronchial wound or from a perforation of the esophagus. Occasionally, blunt injuries to the chest may disrupt the integrity of a group of bronchioles or alveolar units without disrupting the visceral pleura. As air escapes into the pulmonary interstitium, it dissects centrally along the bronchi and pulmonary vessels to reach the mediastinum. When the mediastinal pleura remains intact, progressive loss of air into the tissue carries the dissection into the neck, where the air escapes the deep tissue planes and spreads in the subcutaneous tissue. The development of subcutaneous emphysema may cause marked distortion of the patient's appearance, but there is no reason to "treat" the condition, except to take whatever steps are appropriate to stop the air leak. The source of the leak must be found, because some potential causes (esophageal perforation or major bronchial injury) require early intervention.

Rib Fractures and Lesser Flail Injuries

The most common injury of the chest is a fracture of one or more ribs, including fracture at the costochondral junction ("separation"). Children seem less liable to rib fractures, but chest x-rays are made less frequently in young age groups with minor trauma. Fractures occur most commonly in the middle and lower ribs with blunt trauma, but the distribution with penetrating wounds varies with the distribution of the penetrating objects.

First Rib Fractures

First rib fractures historically have been associated with high probability of associated upper rib fractures and major vessel injuries. Recent reports by many physicians have demonstrated isolated first rib fractures without other significant injuries in the thoracic outlet in a wide variety of patients. Because of the relative high frequency in association with cranial and maxillofacial injuries and in the "surfer's" rib (an injury occurring in surfers performing the lay-back maneuver), it is probable that isolated first rib injuries are secondary to avulsion of the first rib by its muscular attachments rather than direct trauma to the relatively protected first rib. There is inconclusive evidence that a direct relationship exists between first and second rib fractures and trauma to major vessels at the apex of the hemithorax (Fig. [16-15](#)). Arteriography should be considered in stable patients with first rib fracture who have (1) absent or decreased upper extremity pulses, (2) hemorrhage, especially large extrapleural hematoma or hemothorax, or (3) brachial plexus injury. Additional criteria for angiography include displacement of fragments and multiple thoracic injuries.

Multiple Fractures

The problem of massive flail chest was discussed briefly earlier. Lesser degrees of flail occur whenever there are multiple fractures of the chest wall skeletal structure. Flail chest is present when there is paradoxical respiratory movement in a segment of the chest wall. This requires at least two segmental fractures in each of three adjacent ribs or costal cartilages or other multiple combinations of rib or sternal fractures with costochondral or chondrosternal separations. Posterior flail segments, in the absence of disrupted intrathoracic structures, are easier to manage because of the strong muscular and scapular support and because of patients' natural tendency to lie with their back against the mattress.

Chest wall stabilization, reduction of respiratory dead space, and management of the underlying pulmonary contusion are the major goals of treatment. In the past, endotracheal intubation or early tracheostomy were recommended for the management of flail chest, to allow easy access for tracheobronchial suctioning and to facilitate internal stabilization of the chest wall through mechanical ventilation. Improvements in respiratory therapy, bedside monitoring, and pain management now permit individualization in the treatment of patients with flail chest injuries. The severity of the associated pulmonary contusion and the extent to which vigorous pain management allows adequate pulmonary toilet often dictate whether or not intubation and ventilatory support are necessary. In this regard, epidural analgesia or intercostal nerve blocks should be used liberally to provide pain relief. For some patients with a small flail area and minimal pulmonary injury, close observation may allow satisfactory recovery without the use of assisted ventilation. Intubation is delayed until clear evidence of a need for ventilatory support develops: a respiratory rate of 40 breaths/min, a falling PaCO₂ (evidence of excessive work of breathing to maintain adequate oxygenation), or a PaO₂ below 60 torr on inspired oxygen fractions of over 0.5. Rarely, a patient presents with localized chest wall fractures amenable to direct operative stabilization. When feasible, this approach can shorten convalescence. Other techniques for stabilizing the flail segment, such as the use of external compression dressings or the application of traction by encircling the fractured ribs with towel clips or wire, are historical oddities.

Several long-term follow-up studies have been published analyzing the late consequences of flail chest injuries. Significant disability was reported in 50 to 64 percent of the patients, with pain as the common residual complaint. In one study, 40 percent of patients examined had found it necessary to change their life-style as a result of the chest injury. Attention to the early and continuing rehabilitation of these injured patients is undoubtedly indicated by these findings.

Other Rib Fractures

An inward displacement of the fracture fragments at the time of injury may lacerate the lung parenchyma and produce a pneumothorax with bleeding into the pleural cavity. With a single rib fracture, the incidence of pneumothorax is not high, but there is an increasing likelihood of this complication as the number of fractured ribs increases. The occurrence of pneumothorax may be delayed for some hours or even days after the injury has occurred. Hemothorax of a significant degree occurring with rib fractures usually is a result of laceration of an intercostal artery rather than of bleeding from the lung. Bleeding may be delayed in onset or may recur after an interval of several days, and it may be life-threatening. In the patient who has multiple rib fractures with segmental fractures of one or more ribs, a delayed pneumothorax or hemothorax may coincide with some shift of

the rib fragments demonstrated in serial chest x-rays.

In elderly or chronically ill patients, rib fractures may occur with severe coughing or hard straining. The occurrence of a spontaneous fracture should alert the physician to the possibility of a bone abnormality such as metastatic neoplasm or hyperparathyroidism. Pneumothorax and hemothorax are infrequent with rib fractures that do not result from external trauma.

The diagnosis of a rib fracture should be suspected from the pleuritic type of pain and marked tenderness over the fracture area. A localized contusion of the chest wall structures may mimic the findings, including shallow respirations with chest wall splinting. Green stick fractures are those not associated with separation of the fragments, and they may not be demonstrated by the initial chest x-ray examination. Several weeks may elapse before a suspected fracture is confirmed. When the patient has two or more adjacent fractured ribs, especially if the ribs are broken in more than one place (segmental fractures), the diagnosis is made with greater certainty by examination alone. Cartilage fractures and separation from either the rib or sternum are not demonstrated by chest x-rays. In a review of emergency room practice, Thompson and coworkers emphasized how infrequently management is influenced by rib radiographs and urged higher reliance on physical findings and avoidance of efforts to obtain confirmatory radiographs unless the findings might alter treatment.

The principal goal of treatment for patients without serious injury is relief of pain. If this is accomplished, patients may resume their normal activities except those which require a vigorous work effort. Adhesive strapping of the chest or chest binders to splint the fracture area should be avoided in all but the very young. In the majority of other patients, an adequate oral analgesic or an intercostal nerve block plus oral analgesics provides reasonable pain relief with minimal risk of side effects (Fig. 16-16). Binders and strapping are particular hazards in the elderly and in patients with chronic lung disease. The nerve block may need to be repeated once daily for several days, but a single injection may suffice for individuals whose injury does not require hospitalization. A patient whose rib fractures are accompanied by minimal pneumothorax (<1.5 cm separation between the lung and the inner chest wall by x-ray) may be treated as an outpatient.

Sternal Fractures

Any major blunt trauma to the anterior chest wall may cause a fracture of the sternum. Such fractures may occur alone or in combination with multiple rib fractures. The fractures usually are transverse and are located in the body of the sternum at or near the junction with the manubrium (Fig. 16-17). The injury is painful and can be pinpointed by the conscious patient. The diagnosis should be made on physical examination and confirmed by lateral x-rays of the sternum or tomography. It is essential to rule out significant injury to underlying structures, especially the heart. In the absence of other major injury, the treatment of the fracture is aimed toward relief of pain and observation for signs of respiratory embarrassment. In patients with compromised pulmonary function or obvious instability of the fragments, more vigorous treatment is required, e.g., positive-pressure ventilation or operative reduction and stabilization of the fragments. When the pain of a sternal fracture persists for a long period of time, a nonunion should be suspected. This is the result of persistent displacement of the proximal fragment. In this instance, open reduction is appropriate.

Hemothorax

Intrathoracic bleeding occurs with any form of chest injury that disrupts tissues. Hemothorax usually develops at the time of injury, but occasionally the bleeding is delayed for several days. Rarely, an extrapleural hematoma breaks into the pleural cavity and gives the impression of a delayed hemorrhage.

Bleeding from the lung as a result of rib fractures, stab wounds, or small- missile wounds generally stops before a sufficient volume has been lost to mandate an emergency thoracotomy. Accumulated military and civilian experience indicates that slightly more than 10 percent of patients with traumatic hemothorax require thoracotomy for the control of bleeding or determination of the extent of injury.

Movement of the diaphragm and thoracic structures causes partial defibrination of blood that is shed into the pleural cavity, and clotting usually is incomplete. Sufficient coagulation does occur to interfere with efficient drainage of the pleural blood through intercostal catheters, and the latter often become plugged with blood clot. Pleural enzymes begin to produce clot lysis within a few hours after bleeding stops, and the process of hemolysis with protein breakdown increases the osmotic pressure. Unless the pleural space is drained adequately, the transudation of fluid into the space can produce a significant compression of the lung and a shift of the mediastinum toward the opposite hemithorax.

The main diagnostic concerns in the management of the patient with traumatic hemothorax are how much bleeding has occurred, is it continuing, and if stopped and clotted, when should clot be removed. Consideration of the type and extent of injury, general signs of blood loss, physical signs of fluid in the pleural cavity, and chest x-ray findings are guides to the assessment of the extent of hemothorax. Between 400 and 500 mL of blood may be hidden by the diaphragm on the upright chest x-ray, and 1 L or more may be overlooked on a supine film. A large hemothorax may be missed on the upright chest x-ray unless the observer is aware of the phenomenon of subpulmonary trapping of the blood (Fig. 16-18). A lateral decubitus x-ray can confirm the diagnosis of hemothorax and guide placement of an appropriate drainage catheter.

A small hemothorax that produces little more than blunting of the costophrenic angle on the chest x-ray does not require initial treatment; follow-up x-rays at appropriate intervals will assist with the decision to drain the pleural cavity if there is a progressive accumulation. When the hemothorax exceeds an amount that fills the costophrenic sulcus, or when there is associated pneumothorax, one or more large catheters should be placed in the pleural cavity through the seventh or eighth intercostal space in the posterior axillary line. Low suction applied to the catheters is helpful when combined with active efforts at stripping the tubes of blood clot. If the initial drainage of blood is followed by continued bleeding in the absence of a clotting defect, a decision to operate must be made, with a broad consideration of the possible sources of the bleeding.

With a major hemothorax, the success of tube drainage often is frustrated by extensive clot that obstructs the tubes. A nonfunctioning chest tube represents a liability to the patient because of discomfort and the risk of carrying infection from the skin wound into the pleural clot. Especially with penetrating trauma, a hemothorax that fails to drain adequately through intercostal catheters may develop into empyema. An additional hazard is the organization of residual clot to form a fibrothorax. Coselli and colleagues in Houston reviewed their experience with clotted hemothorax and found that early thoracotomy substantially reduces hospitalization time and empyema rates. Videothoracoscopy has become an effective and less invasive way to evacuate a clotted hemothorax. Thoracotomy is performed only if more extensive exposure is needed to manage an active source of bleeding or if extensive decortication of the lung is necessary.

Tracheobronchial Injury

The management of massive tracheobronchial injuries is discussed above. For small penetrating injuries of the intrathoracic trachea and major bronchi, tracheostomy and effective pleural decompression may provide satisfactory definitive treatment. Those injuries which are associated with an actual defect in the tracheobronchial wall, including partial disruption, require operative exploration and repair. Tracheostomy may be necessary to prevent high intratracheal pressures and to allow tracheal care postoperatively, but positive-pressure assisted ventilation should be avoided.

Penetrating injuries of lobar or segmental bronchi may produce a similar clinical picture to proximal tracheobronchial injuries. Bilateral pneumothorax is rare, and the immediate problem is to begin management of the major air leak and confirm the presence of a major bronchial injury. The bronchial air leak often stops soon after an intercostal catheter is put in place. The definitive diagnosis may be delayed if the bronchus becomes obstructed by blood clot or mucus, and the air leak ceases. Under these conditions, the pulmonary lobe or segment becomes atelectatic and resists conservative methods to produce reexpansion. If infection does not occur, the injured bronchus may heal with significant distortion and obstruction, or the atelectasis may persist and lead subsequently to a correct diagnosis. Operative repair of the disrupted bronchus can be achieved even years after injury. If infection occurs at the site of the bronchial injury, the patient may develop pneumonia, distal bronchiectasis, and empyema. Resection of the bronchus and the involved pulmonary lobe is then required.

Pulmonary Injury

The lungs have a remarkable ability to tolerate penetrating injuries and blunt trauma without long-term residual effects. Civilian gunshot wounds of the chest penetrate a lung more frequently than any other structure, but the majority of patients with no other significant injury can be treated without a thoracotomy. Any penetrating object produces an air leak with a variable degree of pneumothorax. The disruption of tissue along the missile track causes bleeding, which usually ceases as the damaged parenchyma becomes swollen and filled with blood clot. With small-caliber and low- velocity bullet wounds that pass through the lung periphery, the amount of tissue damage produced may be sufficiently small that late follow-up chest x-rays fail to demonstrate the area of injury. With high-velocity bullets, the tissue destruction extends more widely; a peripheral bullet pathway may result in irreversible damage to a lobar or lung hilus.

The immediate management of the patient with a penetrating injury is the insertion of at least one intercostal catheter for evacuation of the associated hemopneumothorax. Serial arterial blood gas determinations and frequent evaluation of the patient's ventilatory ability allow an overall estimate of the effect of the injury on respiratory exchange. Low-velocity penetrating wounds rarely require ventilatory assistance or surgical intervention. However, high-velocity gunshot wounds frequently require thoracotomy to control bleeding or to resect an irreversibly injured lung.

Pulmonary Contusion

Pulmonary contusion is the consequence of blunt trauma to the lung. The frequent causes of contusion include rapid deceleration of the chest against a fixed object (e.g., an automobile steering wheel), falls from a height, and blast injuries. Particularly in young persons, severe pulmonary contusion can occur by transmission of force through the chest wall with minimal fractures of the ribs or sternum. In middle-aged or elderly persons, significant pulmonary contusion usually is accompanied by multiple fractures of the thoracic cage.

The contused lung is characterized by capillary disruption that results in intraalveolar and interstitial hemorrhage, edema, protein and fluid obstruction of small airways, and leukocyte infiltration. Serial chest x-rays begun right after injury show a fluffy infiltrate that progresses in extent and density over a period of 24 to 48 h. Although the maximum lung injury is directly related to that region of the chest wall that receives the trauma, a "contrecoup" effect may be responsible for a wider distribution of the pulmonary damage. Unless the contusion involves only a small region of one lung, it may result in serious loss of respiratory function. The associated injury to the chest wall is aggravated by the loss of pulmonary compliance, increasing the work of breathing. Small areas of atelectasis become confluent, and progressive hypoxia further diminishes the patient's ability to compensate for the loss of function.

Pulmonary contusion often is part of a major chest injury that includes one or more fractures of the thoracic cage, pneumothorax, and hemothorax. If not present initially, a pneumothorax may develop subsequently from actual disruption of the contused pulmonary parenchyma. Although it is infrequent in patients who survive to reach the hospital, a major pulmonary laceration may represent the maximum extent of pulmonary contusion. In some instances the tissue disruption is the result of extensive penetration by rib fragments, but in others the causative factor is probably a severe shearing force. The clinical and x-ray findings suggest a serious chest injury but do not differentiate the patient with a major lung laceration from the patient with pulmonary contusion and associated hemopneumothorax. Continued or uncontrolled hemorrhage and massive air leak generally mandate an early thoracotomy. A major pulmonary resection may be necessary, and the mortality rate is high.

Treatment of pulmonary contusion must include an accurate clinical assessment of the patient's respiratory exchange and careful monitoring by serial measurements of the arterial blood gases. Steroids have no role in the management of pulmonary contusion.

A high percentage of patients require temporary assisted ventilation, and it may be evident at the time of admission that endotracheal or nasotracheal intubation should be performed. Without question, aggressive respiratory therapy, including ventilatory support, should be initiated before cardiopulmonary decompensation requires treatment measures that add additional risks. Criteria for instituting assisted ventilation are shown in Table [16-3](#). For most patients, the need for assisted ventilation does not extend beyond a few days unless there is major injury to the chest wall or to other body regions.

Posttraumatic Pulmonary Insufficiency

The development of acute respiratory failure can be expected in a high percentage of patients who suffer major thoracic trauma. Preexisting pulmonary status influences the severity of respiratory insufficiency, and the extent of actual pulmonary damage determines whether the patient survives. An initial evaluation of respiratory exchange and ventilatory ability, confirmed by measurement of pulmonary mechanics and arterial blood gases, should be followed by serial reevaluations.

Especially in patients who have suffered multiple trauma, a respiratory distress syndrome may develop that is out of proportion to the extent of thoracic injury. Previous terms used to designate post-traumatic respiratory insufficiency included *wet lung*, *shock lung*, *congestive atelectasis*, and *adult respiratory distress syndrome*. There is certainly some overlap in the causation of the several forms of respiratory failure that follow major trauma, and it is important to determine the specific causes in individual patients. Blaisdell and Lewis have presented a thorough discussion of posttraumatic pulmonary insufficiency, choosing the term *respiratory distress syndrome of shock and trauma* for those cases not due to a specific cause. They suggest that eight different explanations for respiratory failure, other than the respiratory distress syndrome, occur with reasonable frequency in patients who suffer major injury. These include aspiration, simple atelectasis, lung contusion, fat embolism, pneumonia, pneumothorax, pulmonary edema, and pulmonary embolism.

On the basis of their experience with a large number of patients, Blaisdell and Lewis concluded that the respiratory distress syndrome (RDS) is one and the same as the fat embolism syndrome. Originally thought to result from fat embolism from fracture of long bones, the syndrome consists of pulmonary, neurologic, and systemic manifestations. The pulmonary manifestations appear first, generally within 24 to 36 h after injury, and consist of dyspnea, tachycardia, fever, and cyanosis. Documentation that much of the fat that appears in the blood after injury represents a mobilization of free fatty acids from body neutral fat as a result of shock and increased levels of catecholamines has helped in understanding the mechanism of this condition. Because some degree of intravascular coagulation can be demonstrated in all patients, this is almost certainly a factor in development of the syndrome.

For patients who suffer major chest injury, it may be impossible to define what part of their respiratory failure is a result of direct trauma and how much is a consequence of the RDS. Treatment must be based on correction of the direct results of injury and on the anticipation or early recognition of respiratory insufficiency. The radiologic changes of RDS, consisting of diffuse lung infiltrates that progress to become confluent, may be superimposed on the effects of pulmonary contusion and atelectasis. Changes observed on serial chest x-rays lag behind the changes in pulmonary function, and a patient may be in critical respiratory failure before the films suggest a progressive pulmonary lesion.

Management of RDS requires maintenance of good cardiovascular function and prompt institution of ventilatory support. An adequate volume replacement for external fluid and blood losses is complicated by the internal fluid losses because of increased capillary permeability in the lung, in all areas of direct tissue trauma, and to a varying degree throughout the body. Placement of a Swan–Ganz catheter to measure left atrial and pulmonary artery pressures is superior to monitoring central venous pressures for guidance of fluid and diuretic therapy. The need for inotropic myocardial support can be detected earlier by this access to left-sided heart pressures.

Current ventilatory support techniques allow a wide selection of methods of assisted respiration. These should be chosen to optimize oxygenation while minimizing airway pressures and adverse effects on cardiovascular function. There is no clear role for the use of steroids, which have not been shown to be effective in preventing or reversing RDS. Patient survival depends on meticulous supportive respiratory care and aggressive treatment of conditions (e.g., intra-abdominal sepsis, ongoing blood loss requiring transfusions) that may aggravate RDS.

CHEST WALL

Congenital Deformities

Pectus Excavatum

The most common congenital deformity of the chest wall is pectus excavatum, in which the body of the sternum is displaced posteriorly to produce a funnel-shaped depression (Fig. [16-19](#)). The etiology is not certain, but most authors believe that overgrowth of the lower costal cartilages and ribs is responsible. The severity of the defect varies widely. The depression is centered most often at the xiphisternal junction but may extend to the manubrium in rare patients. In lateral extent, the presentation varies from a narrow central cleft to a broad, dish-shaped defect extending from nipple to nipple. The depth of the depression is equally variable, with the sternum reaching or even overlapping the spine in extreme forms. Asymmetry is common and always involves greater depression of the right costal cartilages with rotation of the sternum to the right.

Pectus excavatum is present at birth and progresses at a variable and unpredictable rate through childhood. Infants and young children often have a protuberant abdomen that accentuates the deformity. Later in childhood a characteristic posture has been described, with rounded and forward-sloping shoulders, forward angulation of the head and neck, and dorsal kyphosis of the spine. Breast development in young women is frequently asymmetrical, with a smaller breast on the right.

Although most cases appear in isolation, a familial tendency has been noted, and the defect is seen frequently in more than one sibling. The anomaly is about three times more frequent in males. Pectus excavatum is frequently seen in Marfan syndrome and is one of a variety of chest wall deformities seen with increased frequency in patients with congenital heart disease.

A variety of classifications based on radiographic findings has been proposed. Haller and associates have used a “pectus index” based on the ratio of transverse to anteroposterior diameter measured on a single standard computed tomographic (CT) scan image. In their series, all patients undergoing operative correction had a pectus index greater than 3.25.

That pectus excavatum produces a cosmetic deformity is not a matter of debate. Between 30 and 70 percent of patients are reported to be symptomatic, with a broad range of presentations including exercise intolerance, atypical chest pain, dyspnea, bronchospasm, poor feeding, and arrhythmias. In all reported series, symptoms are almost always relieved by operative correction. Systolic ejection murmurs are reported frequently and are thought to reflect compression of the right ventricular outflow tract. ECG abnormalities are frequent and usually resolve after repair but are believed to reflect changes in axis due to rotation and displacement rather than any fundamental electrophysiologic disturbance.

Whether or not a physiologic defect is responsible for the characteristic symptomatology is controversial. Wynn and associates and Ghory and associates have reviewed previous reports spanning 70 years and have added their own observations. Ghory and associates studied 14 children with pectus excavatum and 14 normal controls and found a higher diastolic blood pressure and a decreased preejection period with exercise in older children (older than age 11) with pectus. No differences were seen in maximal work load, oxygen consumption, cardiac output, and stroke volume. Wynn and associates studied 12 patients with pectus excavatum. Eight were studied before and after operative correction, and four were tested twice but not corrected. No significant differences were observed in most parameters, such as maximal oxygen uptake, cardiac output, and stroke volume. The only difference that resulted in statistical significance was an 8 percent decrease in total lung capacity after operation; total lung capacity was unchanged in the nonoperated group ($p < 0.01$).

Pectus excavatum is primarily a cosmetic defect with a very broad anatomic and symptomatic spectrum. There is no conclusive evidence supporting the existence of a consistent functional defect. Operation is indicated to correct the cosmetic defect and can be expected to eliminate symptoms in the majority of patients, even though a physiologic basis for the symptomatology cannot be defined.

Operative Treatment

Most authors recommend operative correction during the preschool years (before 5 years of age) but not before 18 months. Operation at that age is thought to prevent the secondary postural, physiologic, and psychological consequences of the defect.

The technique most widely used is that described by Ravitch. All the deformed costal cartilages are excised, the xiphisternal joint is disarticulated, intercostal muscle bundles are separated from the sternum, and transverse posterior osteotomy of the sternum is performed above the point of depression. The osteotomy can be combined with a forward fracture of the sternum and insertion of a bone wedge in the osteotomy site to provide an overcorrection of the deformity.

In certain cases, the repair is reinforced with a metal strut that can be removed several months later. In recent large series, excellent short- and long-term results have been reported. Haller and associates achieved excellent long-term results in 95 percent of 664 patients followed for 1 to 40 years. Shamberger and Welch achieved satisfactory long-term results in 98 percent of 704 patients followed for 1 to 27 years.

Excellent results also have been reported with a very different operative approach. Sternal eversion involves transverse division of the sternum, division of the costal cartilages, 180° axial rotation of the sternum (sternal eversion), and suture reattachment. In essence, the concave deformity is made convex. The sternal plastron can be rotated on a mammary artery vascular pedicle or as a free graft. Although the series has not been updated recently, Wada and associates achieved satisfactory results with sternal eversion in 97 percent of 199 patients followed for 15 years.

Pectus Carinatum

The protrusion deformities of the sternum are much less common than pectus excavatum, accounting for less than 10 percent of patients presenting for repair. This group of “bird chest” deformities varies considerably in specific manifestations. The most common type is characterized by a deep depression of the costal cartilages along each side of the sternum, which accentuates the mild protrusion of the sternum by creating an illusion of greater anterior projection in relation to the ribs. The deformity usually is maximal below the nipple level; asymmetry is common, most often producing mild rotation of the sternum to the right.

Although symptoms reminiscent of pectus excavatum have been associated with the protrusion defects, it appears that most are asymptomatic, and the condition has seldom been studied physiologically. Operative correction is done through a curved submammary incision that allows broad exposure of the deformed cartilages and costochondral junctions (Fig. [16-20](#)). Subperichondral and subperiosteal resection of all deformed cartilages and ribs is performed throughout the length of their deformity. The excessive length of each perichondral bed is obliterated with reefing sutures, and the sternal contour is adjusted with a transverse osteotomy if necessary. In the only reported series of any size, Shamberger and Welch achieved 98 percent satisfactory results in 152 patients followed for 1 to 12 years.

Sternal Fissures

The sternum is formed when two lateral plates of mesoderm fuse in the midline during the tenth week of embryonic development. The clavicular heads also contribute primordia to the manubrium. Failure of fusion can be complete, or it can be confined to the superior end or the inferior end of the sternum.

Superior Sternal Cleft

In this type of defect the cleft is broad and U- or V-shaped, usually extending down to about the fourth costal cartilage (Fig. 16-21). The prominent pulsations of the heart, which is covered only by thoracic fascia and skin, create the illusion of cardiac displacement into the neck. In fact, the heart usually lies in approximately normal position, and the two separate halves of the sternum can be located at the periphery of the defect and reapproximated. Osteotomies in each half or distal transection of each half is usually necessary to bring them together. In some patients, especially those repaired after infancy, these techniques would not leave room for the heart, and coverage with prosthetic material is required.

Distal Sternal Cleft

A defect in the distal sternal is almost invariably part of a syndrome called *Cantrell's pentalogy*, which consists of the following five components: (1) a cleft distal sternum, (2) a ventral abdominal wall defect that may be a true omphalocele, (3) an anterior crescentic deficiency of the diaphragm, (4) communication between the parietal and peritoneal cavities through the diaphragm, and (5) congenital heart disease, usually with a ventricular septal defect and a left ventricular diverticulum.

Operative correction requires a staged approach taking into account the priorities of each defect. The omphalocele usually is repaired first. As with other forms of sternal cleft, early reconstruction offers the best chance for primary closure.

Complete Sternal Cleft

In this rarest form of sternal cleft, failure of midline fusion is complete, leaving the mediastinal contents bulging through a thin covering of skin and fascia. In the few cases described, an associated failure of midline abdominal fusion has been frequent and communication between the peritoneum and pericardium common. Repair in infancy is highly desirable and can be satisfactory.

Miscellaneous Anomalies of Rib and Costal Cartilage

The simplest anomalies consist of deformed, deficient, or enlarged cartilage or rib presenting as an isolated finding in an asymptomatic patient. More complex anomalies include absence or wide divergence of one or more lower ribs and are commonly associated with hemivertebrae, fused bony paravertebral bars, and progressive scoliosis. The chest wall defect can manifest obvious paradoxical respiratory motion and even true lung herniation, but the spinal anomalies usually are more functionally significant and demand more therapeutic attention.

Poland's syndrome consists of absence or hypoplasia of the pectoralis major and minor muscles, breast hypoplasia, and partial absence of the upper costal cartilages (Fig. 16-22). Brachysyndactyly, ectrodactyly, and ectromelia are frequently described associations. It is invariably unilateral. Depending on the extent of cartilage deficiency there may be an impressive lung hernia, paradoxical respiratory motion, or simple flattening of the anterolateral chest wall. When the anomaly is on the left side, the underlying heart and lung are significantly vulnerable because they are covered only by skin, fascia, and pleura. As the child grows, the concavity becomes more severe on either side.

Operative reconstruction is recommended for cosmetic reasons, to eliminate paradoxical motion, and to protect intrathoracic structures. Staged procedures involving split rib grafts from the contralateral side combined with Teflon felt or Marlex mesh have been advocated in the past. A logical outgrowth of the increasing popularity of pedicled myocutaneous flaps has been their application in the reconstruction of this anomaly. Urschel and associates described successful single-stage reconstruction in two patients using a latissimus dorsi flap and simultaneous augmentation mammoplasty.

Chest Wall Tumors

General Considerations

In reviewing this topic in 1949, Brian Blades observed, “. . . available statistical data concerning the exact incidence of thoracic wall tumors are incomplete and probably unimportant. Moreover accurate histological classification of the tumors is often confusing.” If he were writing today, he would probably add that the accurate histologic classification is unimportant. Now as then, most reports in the literature are limited case reports; the few larger series are reported from major tertiary referral hospitals with the highly selected patient population that characterizes those institutions. The available statistical data relating to the actual frequency of the various types probably remain inaccurate. Depending on the referral characteristics of the reporting institution, the incidence of primary malignancy of chest wall tumors ranges from 13 percent (Cavanaugh) to over 50 percent (Sabanathan et al.). It is likely that the true incidence of malignancy in the general population may be nearer the lower figure. Whatever the true frequency, malignancies are common, and clinical or laboratory findings are sufficiently uncertain that speculating about the probabilities of any given chest wall mass being benign or malignant usually is an unimportant exercise. It should be considered malignant until proved otherwise by detailed analysis by an experienced surgical pathologist.

The reported experience confirms that chest wall malignancies are often mismanaged. Inadequate biopsies are untrustworthy, fears about chest wall reconstruction difficulties may encourage inadequate local resection, and recurrence after inadequate initial resection is a common cause of failure. Most surgeons agree that an adequate biopsy should be undertaken early in the evaluation of the suspicious mass. Multimodality therapy increasingly has become a standard treatment for some adult soft tissue malignancies (e.g., Ewing's sarcoma). With advances in reconstruction techniques (e.g., myocutaneous flaps, improved synthetic materials), virtually any defect in the chest wall can be repaired. The current accepted approach is to obtain an accurate diagnosis in order to determine whether the patient should receive multimodality therapy or should undergo surgical resection as the primary treatment. Accurate diagnosis requires adequate tissue; fine needle aspiration should be avoided, and frozen sections should be performed on removed tissue before leaving the operating room, not to establish a diagnosis but to be sure suitable tissue for eventual diagnosis has been submitted. Metastatic tumors, especially to the ribs, and direct invasion of the chest wall from primary lung and breast carcinomas easily outnumber the tumors arising from the chest wall. While the biopsy is being processed, careful search for a primary neoplasm elsewhere should be made. Neoplasms originating in the thyroid, breast, and kidney are the most common to metastasize to the ribs.

The clinical manifestation of a chest wall neoplasm is most often either pain, a palpable mass, or an abnormality detected on a chest x-ray. With benign or malignant lesions, the discomfort is relatively mild, and patients often present with tumors that have been enlarging for months or years. Many patients attribute the tumor origin to some episode of localized trauma, or they state that they discovered the mass while rubbing their chest after a minor injury. A differential diagnosis includes the less frequent pulmonary infections that invade the chest wall, such as actinomycosis and nocardiosis, tuberculous chondritis, costochondral separation, and Tietze's syndrome (nonspecific chondritis). A suspicion of fluctuation in the mass and a corresponding pulmonary lesion may suggest that a diagnostic aspiration should be done anticipating probable infection. A true history of trauma and the ability to reproduce a clicking sensation with local pressure may reinforce the diagnosis of a suspected costochondral separation.

Work-up of the patient with a chest wall mass should focus on a search for other areas of neoplastic involvement by radioisotope studies and other imaging studies along with CT scans to map the local extent of the tumor invasion and to plan resection and reconstruction of the chest wall defect. It also must include evaluation of pulmonary function and assessment of the patient's ability to tolerate the physiologic deficit that might result from the procedure.

Benign Tumors

Among the more likely benign tumors are fibrous dysplasia, eosinophilic granuloma, osteochondroma, desmoid tumor, and chondroma.

Fibrous Dysplasia

The ribs are the most common site of solitary fibrous dysplasia (osteofibroma, bone cyst). Located most frequently in the posterior or lateral portion of a rib, it usually presents as a slowly enlarging, nonpainful mass. Diagnostic radiographs show expansion and thinning of the bony cortex, with a central trabeculated appearance. Fibrous dysplasia in ribs as well as other bones forms part of Albright's syndrome, a condition that includes skin pigmentation and precocious puberty in girls.

Eosinophilic Granuloma

The lesions of eosinophilic granuloma are sometimes part of a disease that includes pulmonary lesions called *histiocytosis X* or *eosinophilic granuloma of the lung*. When it occurs in a rib, the granuloma is a solitary destructive process, often associated with pain and localized tenderness. Radiographs reveal a punched-out osteolytic lesion, which, when subjected to excision and microscopic examination, is found to consist of a chronic granuloma. Healing may occur spontaneously, or a pathologic fracture may develop through the area of osteolysis.

Osteochondroma

These slow-growing tumors generally arise from the cortex of a rib. As with other neoplasms, the occurrence of pain may signal accelerated growth, which produces concern over the possibility of malignant change. The radiographic appearance is often that of a distorted rib cortex with an overlying mass that has a thin rim of calcification.

Chondroma

Chondromas occur at the costochondral junction, primarily in children and young people, and may be difficult to differentiate from chondritis or the sequela of traumatic costochondral separation. Chest and rib x-rays show an expansion of bone with thinned but intact cortex. Probably because of the abundance of cartilage in the chest wall, chondromas and chondrosarcomas are the most common benign and malignant tumors of the skeletal components of the thorax. Chondrosarcoma usually is a well-differentiated tumor easily misdiagnosed as a benign chondroma, resulting in inadequate local resection and consequent local recurrence.

Desmoid Tumors

These tumors are a form of low-grade fibrosarcoma. They have a high propensity to recur locally and should be resected with the same wide margins recommended for primary malignant tumors.

Malignant Tumors

Although the sarcomas arising in the adult chest wall usually are classified by the cell type of origin, prognosis is related to the histologic grade rather than cell classification. The grading system recommended by the American Joint Committee on Cancer no longer considers the cell of origin relevant and groups the wide variety of reported tumors under the general label of *adult soft tissue sarcomas*, although they may arise from any of the mesodermal tissues found in the chest wall. These tumors are potentially curable, usually by wide surgical resection.

Factors that influence prognosis include age, size of tumor, histologic grade, and the stage. If the patient is over 60 years of age and the tumor is more than 5 cm, has poorly differentiated cells, or has spread to the lymph nodes or distant sites, the prognosis is poor. If the tumor has a low mitotic index, is free of hemorrhage or necrosis, and is under 5 cm in largest diameter, it is likely to be curable by surgery alone. Higher-grade tumors are associated with a higher local treatment failure rate and a risk of distant metastases. The role of pre- or postoperative radiation or chemotherapy in preventing recurrence remains undefined.

These tumors occasionally can grow at extremely rapid rates. The case depicted in Fig. 16-23 and Fig. 16-24 was believed to be a walnut-sized breast mass 4 months before the CT scan was taken. Resection required complete removal of ribs 4 through 11 along with the lateral portion of the diaphragm and a generous wedge of lung.

Chest Wall Reconstruction

Because of the high rate of malignancy in chest wall neoplasms, there is need for an aggressive attitude toward management of any mass that likely represents a primary tumor. When malignancy is suspected, preliminary plans must be made for chest wall reconstruction that will allow resection of a generous margin of normal tissue around the neoplasm. The resection should include at least one normal adjacent rib above and below the tumor with all intervening intercostal muscles and pleura. In addition, it is often necessary to include an en bloc resection of overlying chest wall muscles such as the pectoralis minor or major, the serratus anterior, or the latissimus dorsi. When the periphery of the lung is involved with the neoplasm, it is appropriate to resect the adjacent part of the pulmonary lobe in continuity (see Fig. 16-24 C). Involvement of the sternum by a malignant tumor requires a total resection of the sternum with the adjacent cartilages. Techniques for postoperative respiratory support are sufficiently good that resection should not be compromised because of a concern about the patient's ability to ventilate adequately in the early postoperative period.

Reconstruction of a large defect in the chest wall requires the use of some type of material to prevent lung herniation and to provide stability for the chest wall (see Fig. 16-24 D). Mild degrees of paradoxical motion are often well tolerated if the area of instability is relatively small. Several authors, notably Pairolero and Arnold from the Mayo Clinic, have reported an extensive experience with chest wall reconstructions after removal of significant portions of the bony thorax. They emphasize that adequate resection and dependable reconstruction are essential ingredients to a successful operation and express the strong belief that a thoracic surgeon—plastic surgeon team is an important collaboration if these complicated problems are to be undertaken. Although historically a wide variety of materials has been used to reestablish chest wall stability, including rib autografts, steel struts, acrylic plates, and various synthetic meshes, current preference is either a 2-mm-thick polytetrafluoroethylene (Gore-Tex) patch or a double-layer polypropylene (Marlex) mesh sandwiched with methylmethacrylate. When soft tissue coverage is needed, this is provided by myocutaneous flap reconstruction.

DISEASES OF THE PLEURA AND PLEURAL SPACE

The inner surface of each hemithorax has a mesothelial lining, the parietal pleura, that is invaginated at each pulmonary hilum to form the visceral pleura. The two surfaces are normally in apposition, lubricated by a thin layer of serous fluid secreted by the mesothelium so that the steady motion of normal respiration is accomplished without friction. The pleural “space” is normally only a potential space lying between the visceral pleura investing the lung and the parietal pleura of the chest wall. The elastic recoil of the lung and the rapid continuous absorption of fluid from the pleural space create a balance of opposing forces that favors apposition of the visceral pleura to the parietal pleura. The introduction of fluid or air breaks this dynamic coupling and converts the potential space to a real space. Normal respiratory mechanics are impaired in proportion to the size of the space created and the pressure within it. Many of the processes affecting the pleural space are essentially mechanical, such as spontaneous pneumothorax or congestive heart failure, and are not associated with any pathologic alteration in either pleural surface. However, virtually any chronic form of pleural space disturbance is associated with pathologic changes that produce thickening and adherence of the visceral and parietal surfaces. The end results vary from a few filmy adhesions of no consequence to a dense fibrous and calcific obliteration of the pleural space with a permanent restrictive defect in pulmonary function.

Pleural Effusion

A pleural effusion is an accumulation of fluid in the pleural space. It is not a disease entity but signals the effect of pleural or systemic disease on the normal daily passage of fluid through the pleural space. Normally, the balance of hydrostatic and colloid osmotic forces favors movement of fluid from systemic capillaries in the parietal pleura to pulmonary capillaries. It is estimated that between 5 and 10 L of protein-free fluid traverses the pleural space in 24 h. Simultaneously, lymphatics drain smaller volumes of fluid containing protein, which would otherwise remain in the pleural space as a source of colloid osmotic pressure favoring retention of fluid. Alterations in systemic hydrostatic or colloid osmotic pressure that disturb the balance of forces across normal pleural surfaces produce an effusion consisting of a protein-poor ultrafiltrate of plasma classified as a *transudate*. Changes in capillary permeability caused by inflammation or infiltration of the pleura produce a protein-rich effusion classified as an *exudate*. Common causes of transudates and exudates are listed in Table 16-4 and Table 16-5. The distinction between transudate and exudate has diagnostic relevance, as noted in one series in which effusions were malignant in 42 percent of patients with an exudate and were caused by congestive heart failure in 83 percent of patients with a transudate.

Characteristics of fluid obtained by diagnostic thoracentesis that can help to make the distinction between transudative and exudative effusions are summarized in Table 16-6. Few findings are independently diagnostic, with the exception of positive cultures (empyema) and positive cytology (malignancy). Certain gross findings can be nearly diagnostic, such as the milky white fluid of chylothorax or the foul purulence of an empyema. Other findings can narrow the possibilities considerably. For example, grossly bloody fluid (red cell count > 100,000/mm³) is almost always caused by trauma, pulmonary infarction, or malignancy. A markedly elevated amylase level can be found in sympathetic effusions associated with pancreatitis, pancreatic pseudocyst, and esophageal perforation. Pleural fluid pH less than 7.20 (with an arterial pH > 7.35) strongly suggests bacterial infection and may appear before culture and Gram's stain are positive in some cases. Low pH also has been reported in some malignant effusions and in effusions associated with connective tissue disease.

There can be considerable overlap in the findings ostensibly separating exudate from transudate, and any chronic effusion tends to develop “exudative” characteristics. Too much can be made of laboratory distinctions, and it is rare for pleural effusion to be the sole manifestation of disease such that diagnosis hinges exclusively on pleural fluid analysis. The cause of most effusions is best recognized by simply looking carefully at the rest of the patient.

A concave meniscus in the costophrenic angle on an upright chest x-ray suggests the presence of at least 250 mL of pleural fluid. A lateral decubitus view can detect a smaller volume and confirms that the fluid is free in the pleural space if it is shown to layer out along a dependent surface. In some cases an effusion is completely contained between the base of the lung and the diaphragm (a subpulmonic effusion) and can be difficult to distinguish from an elevated hemidiaphragm or a subdiaphragmatic process. When this occurs on the left side, the position of the stomach bubble can provide a useful clue. On a supine film, a small to moderate effusion will be completely inapparent, and a large effusion only produces a uniform hazy appearance of the affected hemithorax that can be difficult to detect unless the process is unilateral. A very large effusion can produce complete opacification of one hemithorax that does not change in appearance with changes in position (Fig. 16-25). Adhesions can compartmentalize an effusion into loculations that assume a wide variety of radiographic configurations and that require CT scanning for definition. Presence of an air-fluid level has specific connotations because the air can only come from the tracheobronchial tree, from the esophagus, or directly through the chest wall.

Thoracentesis is the mainstay of diagnosis. Needle biopsy of the pleura can provide diagnostic tissue but has a high frequency of false-negative results because of sampling difficulties in diseases that do not involve the pleura uniformly. When clinical evaluation and thoracentesis fail to yield a diagnosis, videothoracoscopy with pleural and lung biopsies should be performed to determine the cause of the effusion.

Pleural effusions can produce dyspnea but also can be asymptomatic at rest. Therapeutic drainage is rarely indicated for transudative effusions because the fluid rapidly reaccumulates until the underlying condition is improved. Most exudative effusions warrant a more aggressive approach. The treatment of hemothorax was discussed earlier, and empyema, malignant effusion, and chylothorax are considered separately below. A number of nonmalignant, uninfected exudative effusions frequently are treated as if they were transudative; examples include collagen-vascular disease, pulmonary infarction, and sympathetic effusion secondary to abdominal pathology.

Malignant Pleural Effusion

More than half of all patients with malignancy will have a pleural effusion at some time in their course. The effusion is frequently massive and symptomatic. The pathophysiology is thought to be interference with venous and lymphatic drainage by direct tumor invasion. The fluid contains malignant cells in only about 60 percent of patients, and pleural biopsy, usually by videothoracoscopy, often is required to establish a diagnosis. Lung carcinoma is the most common primary, with breast and gastrointestinal malignancies close behind. The fluid is exudative in character and often bloody. Grossly bloody fluid (red cell count > 100,000/mL) has a 90 percent probability of being malignant, once trauma and pulmonary infarction are excluded. The presence of a malignant effusion is a poor prognostic sign, with mean survival after diagnosis of 3 to 11 months in most series.

Treatment

Treatment is palliative. Repeated thoracentesis has a high failure rate. Chest wall radiation, thoracotomy with decortication and pleurectomy, and even pleuropneumonectomy have been described but carry unacceptable mortality and morbidity to be considered standard treatment. At present, the standard therapy is tube thoracostomy or videothoracoscopy and pleurodesis.

Pleurodesis creates an inflammatory fusion between visceral and parietal pleura that eliminates the potential pleural space. An essential first step is complete evacuation of the fluid and reexpansion of the lung, accomplished by inserting a chest tube connected to a water-seal drainage system (Fig. 16-26). If the pleural effusion is large (i.e., greater than 1 L) or has been present for more than a few days, it should be drained slowly initially via the thoracostomy tube in aliquots of 200 to 300 mL/h in order to prevent re-expansion pulmonary edema. When loculations, inaccurate tube placement, or entrapment of the lung by a visceral pleural peel “trapped lung” prevent complete fluid removal and lung expansion, pleural symphysis will not occur uniformly, and pleurodesis will not succeed. This is probably more important than the choice of the chemical agent used in the next step. Innumerable agents have been used to induce the inflammation, including talc, nitrogen mustard, *Corynebacterium parvum*, doxorubicin, quinacrine, bleomycin, and tetracycline. Formerly, tetracycline was the agent most commonly used in the United States because of its low cost and easy availability. Because tetracycline is no longer commercially available, doxycycline, a newer antibiotic of the same type, is occasionally used. Talc, the preferred sclerosing agent in Europe for several decades, has become the most popular agent in the United States as well. Since talc is not soluble, it is administered as a suspension or “slurry” in saline via the tube thoracostomy or is insufflated as a dry powder “talc poudrage” across the pleural space by thoracoscopy. Following pleurodesis, the tube thoracostomy is left to underwater drainage until pleural fluid output is minimal, usually for approximately 72 h. The reported success rate of talc pleurodesis is 80 to 90 percent.

Use of an indwelling shunt connecting the pleural cavity to the peritoneum through a one-way valve has received attention recently. The system is analogous to the Denver shunt, except that the normal pressure gradient between the abdomen and chest is overcome with a subcutaneous squeeze- bulb pump. Although enthusiastically advocated by some authors, pleuroperitoneal shunts have the theoretical disadvantage of continuously circulating malignant cells. They also are easily clogged by proteinaceous debris and require the active participation of patients and families in compressing the bulb pump up to 100 times per day to maintain satisfactory function of the shunt. For the majority of patients with end- stage cancers, pleuroperitoneal shunts do not provide optimal palliation.

Empyema

Empyema is a suppurative infection of the pleural space. Empyema was studied carefully 2400 years ago by Hippocrates, who first described open drainage with rib resection. By the early 1900s empyema complicated pneumonia in 5 to 10 percent of cases. In the antibiotic era, empyema has become a less frequent complication of pneumonia, occurring in about 1 percent of cases, and the bacteriologic spectrum has shifted from *Pneumococcus* and *Streptococcus* to *Staphylococcus*, *Streptococcus*, gram-negative, and anaerobic organisms. Although empyemas are most frequently caused by pneumonia, they also can occur after trauma, pulmonary infarction, or pulmonary resection or can be caused by spread from an intraabdominal source.

Infection of the pleural space initially produces a large, exudative effusion with a high concentration of leukocytes. In hours to days, fibrinous adhesions succeed in limiting involvement to one or more loculated compartments. The ability of the lung to expand and obliterate potential space becomes very important in confining the infection and prevents formation of a fibrous “peel” over the visceral pleura that leads to a trapped lung. The pleura actually has remarkable ability to resolve infection when assisted by an expanded lung. A persistent air leak (bronchopleural fistula) potentiates infection by providing a route for constant inoculation of the pleural space and by promoting lung collapse. The difficulty of obliterating space after pulmonary resection, particularly pneumonectomy, accounts, in part, for the seriousness of postresection empyema.

Clinical Manifestations

Empyema should be suspected in a patient with a febrile illness and pleural effusion on chest x-ray. Thoracentesis with Gram stain and culture of the fluid obtained confirms the diagnosis and guides selection of antibiotics. The gross appearance of the fluid usually is unambiguous, although some seropurulent parapneumonic effusions are sterile. Pleural fluid with a pH below 7.20 and a glucose level below 40mg/dL strongly suggests empyema requiring drainage. Loculations or air-fluid levels on chest x-ray suggest empyema, but CT examination usually is required to distinguish simple empyema from an intraparenchymal process such as lung abscess, an infected congenital cyst, or an infected bulla (pyocyst).

Treatment

Successful treatment depends upon early recognition of the problem, selection of appropriate antibacterial therapy based on identification of the organism, and complete obliteration of the empyema space (Table 16-7). Thoracentesis alone provides adequate treatment in only about 10 percent of empyemas because the space rarely is obliterated. The first step should be insertion of a chest tube connected to a closed drainage system, applying suction as necessary to fully evacuate the space and promote lung expansion. Image guidance during tube or catheter insertion, with CT or ultrasonography, can be important when the empyema space is small or has complex loculations. Lung expansion is especially important in empyema associated with bronchopleural fistula. The success of streptokinase in lysing loculations to improve simple closed-tube drainage is controversial. In early empyemas, where simple debridement of the proteinaceous debris in the pleural space is needed to allow full expansion of the lung, videothoracoscopy is the fastest, most effective approach to treatment.

When tube thoracostomy successfully drains an empyema but there is a small residual space, the simplest option is to convert to open-tube drainage by cutting the existing tubes off near the skin, allowing chronic drainage into dressings. When this approach is successful, the cavity gradually will shrink and obliterate over weeks to months, slowly extruding the tubes, which are progressively shortened.

It is important that open drainage not be attempted before a secure pleural symphysis has occurred at the margins of the cavity, a process that requires at least 10 to 14 days. If this rule is violated, the resulting pneumothorax can cause a rapid spread of the infection throughout the pleural space. A simple test is to disconnect the tubes from the water seal and obtain a chest x-ray. If the space is ready for conversion to open drainage, the x-ray will be unchanged. Open drainage is most effective when the tubes are in a very dependent position, because suction cannot be used to augment gravity drainage. A combination of closed-tube and open-tube drainage is successful in at least 60 percent of patients.

Formal open drainage with rib resection (Fig. 16-27) was done more frequently in the preantibiotic era but is still useful today in the treatment of small, chronic, mature empyemas with a thick fibrous capsule (Fig. 16-28). Drainage is ensured by marsupialization of the empyema cavity. The same cautions important in conversion of closed-tube to open-tube drainage still apply—the cavity must be mature, the drainage site should be dependent, and production of a bronchopleurocutaneous fistula is best avoided. The larger wound allows easier access to the cavity, and drainage can be augmented with irrigation.

If drainage fails to expand the lung, a permanent restrictive defect in ventilation on the affected side is likely to result as the inflammatory membrane heals and contracts over the surface of the lung. Failure of expansion is frequent with a bronchopleural fistula. Between 4 and 7 days of high-suction drainage generally is considered an adequate trial, after which thoracotomy and decortication should be performed (Fig. 16-29). The empyema space is completely evacuated under direct vision, and drainage tubes are accurately placed in the most dependent position. The inflammatory “peel” is tightly adherent to the visceral pleura and should be entirely removed, a tedious process that must be done carefully to prevent development of new air leaks from tears in the lung. When an intraparenchymal abscess coexists with a large air leak, drainage or even pulmonary resection is necessary. Thoracoplasty and intrathoracic rotation of muscle flaps can be added to help obliterate the remaining space.

Decortication has been successful in 80 to 100 percent of patients in reported series and often shortens hospitalization. It has the disadvantage of requiring general anesthesia and a thoracotomy in patients who frequently have limited ventilatory reserve and are suffering the systemic consequences of chronic infection. Even so, in the long run, thoracotomy is often better tolerated than a chronic, draining infection.

Pleural Plaques and Calcification

Pleural plaques are thickenings of parietal pleura, usually smooth and white, and frequently calcified. They vary in size, are frequently bilateral and symmetrical, and do not occur at the apex or on the visceral pleura. They are often caused by exposure to asbestos but have no documented relationship to mesothelioma or other neoplasms. Localized inflammatory or traumatic events also can heal with production of calcified plaquelike lesions, but they are usually larger and unilateral.

Chronic pleuritis can result in diffuse, remarkably uniform thickening and calcification of parietal pleura. The original pleuritis may result from an unresolved hemothorax, from tuberculous or nontuberculous empyema, and from viral or bacterial pleuritis or pleuropneumonitis. When the wall of a chronic but active empyema becomes calcified, resolution by drainage alone never occurs, and resection, decortication, pleurectomy, and thoracoplasty are likely to be required.

Chylothorax

Leakage of lymphatic fluid (chyle) from the thoracic duct produces a characteristic milky effusion called a *chylothorax*. The most common cause is surgical trauma to the thoracic duct, most frequently seen after distal esophagectomy or procedures that involve dissection in the lower left cervical region at the confluence of the subclavian and internal jugular veins (Fig. 16-30). A mediastinal lymph node dissection is another less common cause. Traumatic chylothorax is almost always unilateral, usually on the left side. Nontraumatic chylothorax is less common, accounting for about one- third of patients in most series. Because venous hypertension in the brachiocephalic system is the most common cause of nontraumatic chylothorax, the process usually is bilateral. Underlying causes include superior vena cava thrombosis complicating central intravenous line placement and chronic elevation of central venous pressure after the Fontan procedure. The least common variety of nontraumatic chylothorax is that associated with malignancy, usually lymphoma. It is frequently accompanied by chylous ascites and is caused by neoplastic obstruction of lymphatic channels.

Aspiration of milky white, odorless fluid from the pleural space is virtually diagnostic. Pseudochoyle, which has a similar appearance, is a rare source of confusion seen in certain malignancies, infections, and connective tissue diseases. In comparison with chyle, pseudochoyle has a lower fat content and lymphocyte count, and its opalescent appearance is caused by the presence of lecithin-globulin complexes. If the patient is not eating, or if a coexisting problem could significantly dilute the chylous drainage, the gross appearance of the fluid may not be distinguishable from many other effusions. Table 16-8 summarizes characteristics of chyle that can be diagnostically helpful when gross appearance is ambiguous. The lymphocyte count and triglyceride level are most useful. Lymphangiography can define the site of the leak when it is not clinically obvious and is most useful in cases of nontraumatic chylothorax. It is rarely indicated in the traumatic variety.

Normal chyle flow ranges between 1.5 and 2.5 L/day but can vary much more widely depending on diet and on the fat content of the diet. During starvation or intravenous feeding, flow falls to about 250 mL/day of clear fluid. Chylothorax is frequently massive (Fig. 16-31) and symptomatic, and significant volume losses can occur through thoracentesis or chest tube drainage. Dehydration, nutritional losses, and a steady decline in circulating lymphocytes can produce significant disability and an increased susceptibility to infection.

Treatment

Until Lampson described successful treatment of chylothorax by ligation of the thoracic duct in 1948, mortality for the condition averaged 50 percent. Since that time, a better understanding of fluid and electrolyte management and the development of total parenteral nutrition have introduced additional options. Spontaneous resolution can occur, so a trial of nonoperative treatment usually is justified. Conservative treatment has two goals: One is to decrease chyle production, and the other is to keep the lung expanded against the mediastinum. Maximal reduction in chyle production is achieved by eliminating oral intake, while the patient is supported by total parenteral nutrition. A possible compromise is replacement of dietary fat with medium-chain triglycerides, which are not absorbed by lymphatics, but this often is unsuccessful. Fluid can be removed intermittently by thoracentesis, but continuous evacuation with a chest tube is much more effective. It is generally accepted that a 7- to 10-day trial of drainage and diet manipulation is justified. Experience in renal transplantation has shown that thoracic duct drainage produces measurable immunosuppression after 2 weeks. An occasional patient will have such massive drainage that persistence for more than a few days is unacceptably debilitating. A commonly accepted criterion is that drainage exceeding 500 mL/day in an adult or more than 100 mL/day/year of age in a child is an indication for abandonment of conservative therapy.

When conservative treatment fails, the operative approach is dictated by the cause and the location of the leak. For iatrogenic injury near the aortic arch or at the confluence of the subclavian and jugular veins, the left thorax or supraclavicular area is explored, the site identified, and the injury controlled with direct suture. If the site of injury cannot be clearly identified, the thoracic duct is ligated at the diaphragm where it enters the thorax, and a pleural abrasion is performed.

Ligation of the duct is most easily done through a right thoracotomy. If the problem is bilateral, the right side is approached first; ligation usually resolves the problem on both sides. Direct operative approaches through thoracotomy are successful in approximately 80 percent of patients. Treatment failures are most common in nontraumatic chylothorax. Based on early experience with thoracoscopy, thoracotomy may become a last resort for exploration.

The use of a pleuroperitoneal shunt has been added recently to the list of treatment options for chylothorax. Fluid is pumped from the pleural cavity to the peritoneal cavity with a small subcutaneous squeeze bulb on a one- way valve. Excellent results were obtained with Denver pleuroperitoneal shunts in 12 of 16 patients (75 percent) recently reported by Murphy and associates, and it was possible to remove the shunt in 10 of the 12 patients with good outcomes. However, pleuroperitoneal shunting should be considered only when other treatment options such as duct ligation are not possible or appropriate because they are potentially associated with risks of obstruction or infection.

Tumors

Mesothelioma

Mesothelioma is a neoplasm originating in the mesothelial lining of serosal cavities. Tumor presents in the pleura in 80 percent and in the peritoneum in 20 percent of patients. Although mesothelioma is a rare tumor (2.2 cases per million per year), an increased incidence has occurred since the association with asbestos exposure was recognized in the 1960s. In several large series, a history of asbestos exposure has been documented in 50 to 87 percent of patients. The disease is most common in geographic areas having local industries associated with high risk of exposure, such as shipbuilding or manufacturing processes using asbestos. Exposure to asbestos particles carried on the clothing of workers at risk has been implicated as a cause of increased incidence of mesothelioma in family members. The epidemiology is complicated by the fact that asbestos is nearly ubiquitous in any urban environment, and the characteristic refractile particles can be identified in many people without disease. Smoking is not an important etiologic cofactor, and the latency period from exposure to clinical disease is at least 20 years.

Mesothelioma previously was thought to occur in a benign and a malignant form. However, the term *benign mesothelioma* is now considered a misnomer. Such tumors, now more appropriately termed *benign fibrous tumors of the pleura*, are unrelated to asbestos exposure or to diffuse malignant pleural mesothelioma. They present as an intrathoracic mass on chest x-ray and usually are asymptomatic. Pleural effusion is rare. Pathologically, the tumor arises from visceral pleura as a discreet pedunculated mass (Fig. 16-32). Resection, performed via videothoracoscopy or thoracotomy, is curative and should be done to obtain an accurate diagnosis and to prevent further tumor growth.

In contrast, *malignant mesothelioma* is a locally aggressive neoplasm usually appearing after age 40, with a male predominance of more than 2:1. The tumor usually appears to be multicentric, with multiple pleural-based nodules ultimately coalescing to form sheets of confluent mass. There are four cell types: epithelial, mixed epithelial and fibrosarcomatous, fibrosarcomatous, and desmoplastic. Epithelial histology has the best prognosis, whereas desmoplastic and fibrosarcomatous have the worst and mixed tumors have an intermediate prognosis.

Although hematogenous and lymphatic spread occur in at least half of patients, the predominant clinical feature is local tumor growth. Involvement of lung, chest wall, diaphragm, and mediastinal structures is common. In one series of 69 patients, only 3 of 14 patients autopsied had disease limited to the thorax, and 9 patients developed tumor in a needle biopsy tract. Most untreated patients die of symptoms caused by the primary tumor rather than metastases. Thrombocytosis is frequently seen and is thought to be related to cytokine overexpression (interleukin 6). In one series the platelet count was more than 400,000/mm³ in 90 percent of patients and greater than 1,000,000/mm³ in 14 percent.

Clinical Manifestations

Chest pain, dyspnea, or both are present in virtually all patients with malignant mesothelioma. Pleural effusion is present at some time in 85 to 95 percent of patients, although as the disease progresses the pleural space tends to become obliterated with solid tumor. Radiographic findings other than effusion cover a wide spectrum including pleural thickening, lung nodules, chest wall masses, and mediastinal masses (Fig. 16-33). CT scan of the chest and upper abdomen is the most accurate means of non-invasive staging of the primary tumor (Fig. 16-34).

A tissue diagnosis is difficult to obtain without thoracoscopy or open pleural biopsy. In a series of 123 patients reported by Brenner and associates, thoracentesis in 60 patients revealed malignant cells in only 22, and a definitive diagnosis of mesothelioma could only be made in 7. Unless a representative specimen of adequate size is obtained, the epithelial cell type in particular can be easily confused with adenocarcinoma. Accurate pathologic diagnosis relies on immunohistochemistry. Malignant mesotheliomas rarely stain for carcinoembryonic antigen (CEA), whereas adenocarcinomas almost always do. When immunohistochemistry does not yield a definitive diagnosis, electron microscopy is required.

Treatment

The treatment of malignant pleural mesothelioma is controversial because the factors affecting prognosis are poorly understood. Prognosis is linked primarily to tumor stage and histology. The Butchart staging system (Table 16-9), developed in 1976, has been abandoned in favor of a new staging system proposed by the International Mesothelioma Interest Group (IMIG), which provides more accurate descriptors for tumor T and node N status (Table 16-10).

Recent information indicates that the natural history of malignant mesothelioma is more variable than previously thought. Survival in untreated patients ranges from 6 months for those with stage IV disease to 2 years or more for those with stage Ia tumors.

Patients who have very early stage tumors (stage Ia) are reported to have their survival prolonged by the intrapleural administration of gamma interferon, but most patients are not diagnosed until their tumors are more locally advanced. The role of surgical resection in the treatment of malignant pleural mesothelioma is not fully defined. Complete tumor resection by extrapleural pneumonectomy can now be performed with a mortality rate of 6 percent when patients are carefully selected and the operation is done by experienced surgeons. When coupled with adjuvant hemithoracic radiation, extrapleural pneumonectomy offers the best chance of local control. Patients who have stage I or II tumors of epithelial histology are the most likely to benefit from this approach. Although some centers routinely offer patients adjuvant chemotherapy, the benefit of chemotherapy is undefined, especially since no currently available drug has more than a 20 percent response rate.

Patients with stage I or II tumors who do not have the pulmonary function to tolerate an extrapleural pneumonectomy can be palliated by thoracoscopic talc pleurodesis or by pleurectomy/decortication. At Memorial Sloan-Kettering Cancer Center, experience with 105 patients managed by pleurectomy/decortication and adjuvant radiation (external-beam hemithoracic radiation with or without intraoperative brachytherapy) suggests that about one-quarter of operable patients may have their life expectancy prolonged by this treatment approach. Most patients develop locally recurrent disease within 2 years, and the risk of radiation-induced pneumonitis or pericarditis is significant.

Patients who have more locally advanced or metastatic mesothelioma have few treatment options. Radiation cannot be delivered in therapeutically effective doses in patients who have not had a pneumonectomy because of the sensitivity of the lung to high doses of radiation. The low response rates to currently available chemotherapy make it unattractive to treat patients routinely with standard agents. Other approaches such as immunotherapy or gene therapy are experimental.

Metastatic Pleural Tumors

Over 90 percent of pleural tumors are metastatic. Lung and breast carcinoma are the most common primaries. In more than half of all cases, gross tumor is not visible but produces a malignant pleural effusion. When multiple nodules or diffuse oblitative spread occur, differentiation from mesothelioma is impossible without biopsy.

Spontaneous Pneumothorax

Nontraumatic pneumothorax most commonly results from rupture of a pulmonary bleb or bulla. Negative intrathoracic pressure throughout the respiratory cycle favors movement of air into the pleural space with egress prevented by the ball-valve effect of collapsing tissue during expiration. The pneumothorax continues to progress until the leak seals with fibrin at a rate directly related to the size of the bleb. Large leaks can produce life-threatening tension pneumothorax. Spontaneous resolution can occur once the leak stops, but the gas in the space is mostly nitrogen and is very slowly reabsorbed by the pleural surfaces.

Up to 80 percent of patients with spontaneous pneumothorax are young adults, usually male, without clinically significant pulmonary disease. A tall, asthenic habitus is common. In 85 percent of patients, blebs or bullae of varying sizes are found in the lung apices (Fig. 16-35), and it is not known whether their origin is congenital or acquired. After the first episode, the chance of ipsilateral recurrence is 50 percent, and the risk rises with each recurrence to 62 percent after a second episode and 80 percent after a third episode. The risk of a contralateral pneumothorax after the first episode is about 10 percent.

In patients over age 40, significant pulmonary disease usually is present, most frequently emphysema in a tobacco addict. *Catamenial pneumothorax* is a rare condition in which pneumothorax occurs predictably within a few days of menses, usually in women over 30, and almost always on the right side. The mechanism is not known. The two most frequently cited possibilities are pleural endometriosis and small perforations of the diaphragm.

Clinical Manifestations

Chest pain is the most common presenting symptom, followed by dyspnea. If the lung is more than about 25 percent collapsed, a decrease in breath sounds is evident on auscultation, and the affected side is hyperresonant to percussion. Young patients without underlying lung disease can be asymptomatic at rest with nearly complete collapse of one lung, and arterial blood gases are nearly normal. A more dramatic presentation, including tachypnea, cyanosis, and hypoxia, is seen in patients with underlying lung disease and limited ventilatory reserve. An occasional patient with extensive lung disease and a pleural space obliterated with adhesions will present with massive subcutaneous emphysema and pneumomediastinum because air escaping from the ruptured bleb follows the path of least resistance retrograde through the peribronchial soft tissue.

The characteristic radiographic finding is absence of lung markings and a faintly visible line defining the edge of the lung. When the lung collapses almost completely, it is visible as an irregular density attached to the hilus (Fig. 16-36). Presence of a small amount of fluid with an air-fluid level is common. The fluid usually is serosanguineous and insignificant. On occasion, bleeding from a torn pleural adhesion produces a large and increasing hemothorax that can require urgent exploration. The lung fields must be examined closely for evidence of gross abnormalities, such as apical blebs or bullae. Although blebs and bullae frequently are obvious at thoracotomy, only about 15 percent are visible radiographically.

Treatment

An asymptomatic or mildly symptomatic pneumothorax with less than 30 percent collapse that is shown not to increase in size over 6 to 8 h can safely be observed. Simple needle aspiration of the airspace can nearly eliminate the space in a stable pneumothorax and greatly reduces the amount of time required for spontaneous resolution. In a report by Delius and associates, aspiration through an 8F Teflon catheter was successful in 69 percent of 114 patients and in 87 percent of patients with simple pneumothorax. The average cost was comparatively low. Needle aspiration of a tension pneumothorax can be a lifesaving temporizing maneuver.

Thoracostomy tube drainage is the most common treatment. The tube is inserted either anteriorly (second interspace, midclavicular line) or laterally in a lower interspace (middle to anterior axillary line), with the tip directed toward the apex. The tube is connected to water seal, to which suction can be added to increase the gradient favoring removal of air from the pleural space. Water seal alone suffices in many cases. As the lung reexpands, the patient feels pain as the visceral and parietal surfaces reoppose. The pain gradually subsides but usually is much more acute and severe when suction is applied initially. Rapid reexpansion occasionally leads to reexpansion pulmonary edema involving the ipsilateral lung, which was seen in 21 of 146 patients (14 percent) reported by Matsuura and associates. Those at greatest risk were patients 20 to 39 years of age with large pneumothoraces. Although the cause is not understood, the entity needs to be kept in mind because mortality has been as high as 20 percent. Some authors favor attaching the tube to a one-way flutter valve (Heimlich valve) that permits outpatient treatment of a pneumothorax in a reliable patient with a small leak.

Serial check x-rays are followed to assess reexpansion, and the size of the air leak is monitored by observing the rate of bubbling in the water-seal chamber. Air crosses the water seal only with cough or Valsalva in a pneumothorax caused by a leak that has already sealed, and the bubbling usually ceases altogether within 24 h. At the opposite extreme,

continuous bubbling occurring through both phases of respiration reflects a large, active leak that may take days to seal, if it will seal at all. With large air leaks, a single tube may be inadequate. If two tubes connected to suction still fail to expand the lung, thoracotomy is required. Even a large leak can seal if the lung can be expanded fully, which promotes adhesion formation between parietal pleura and the site of the leak in the visceral pleura. Intrapleural instillation of sclerosing agents such as talc appears effective in reducing the recurrence rate.

Operation is indicated for a massive air leak with failure of lung reexpansion or for a smaller leak that has persisted for more than a week. Because of the frequency of recurrence after one episode, operation is recommended after any second episode and in any patient with a previous contralateral pneumothorax. Operation might be recommended after a first episode to anyone with large apical bullae visible on chest x-ray, to persons likely to be exposed to dangerous changes in atmospheric pressure (airline pilots, scuba divers), or to persons living in remote areas. Complications such as empyema or hemothorax occasionally develop and mandate operation. Conservative treatment is continued as long as possible in older patients with underlying lung disease because of their limited ventilatory reserve.

At thoracotomy or thoracoscopy, the site of the leak can be identified and resected, oversewn, closed with staples, or ablated with laser energy. Pleural abrasion also should be performed to promote formation of adhesions between visceral and parietal pleura, an especially important maneuver if no leak site can be identified. Pleurectomy, accomplished by stripping all the parietal pleura off the underlying ribs and intercostal muscles, is undeniably effective but has substantially greater morbidity and is reserved for extreme cases. Either method is 90 to 95 percent effective.

LUNG

Development and Anatomy

In the 4-mm (3-week) embryo, an outpouching from the primitive foregut appears caudad to the paired pharyngeal pouches and bifurcates into the right and left primitive bronchial buds. Over the next 2 weeks, further branching occurs with 10 segmental tubes on the right and 8 on the left, providing an early indication of the lobar development that continues in each lung. Progressive branching of epithelial tubes results in a rich arborization of bronchioles and alveolar ducts and sacs. It is estimated that 300 million alveolar sacs eventually develop. As the structural maturation is taking place, histologic differentiation progresses from the cuboidal epithelium that lines the terminal buds during the first 4 fetal months to the flattened epithelium present at birth. It is likely that the basic architecture of the lungs is completely developed at birth. As the lungs grow, they bulge into the lateral pleural cavities, leaving a dorsal mesentery to encase the developing mediastinal structures. The most caudal pair of aortic arches (the sixth) gives rise to the pulmonary arteries, with the remnant of the left sixth arch persisting as the ductus arteriosus. Vascular sprouts from the unilocular atrium fuse with the developing capillary vasculature in the lung mesenchyma to become the pulmonary veins.

Although the number of respiratory units may not increase after birth, it does seem apparent that the newborn's lung is structurally immature. In place of alveoli, the lungs are made up of primitive air sacs that differentiate into alveolar ducts and sacs. Alveoli develop by outpouching and compartmentalization, and maturation continues throughout the first 8 years of life. The fully developed alveoli give a surface area of 70 to 80 m² at three-fourths maximal inflation of the adult lung.

Segmental Anatomy

The segmental anatomy of the lungs and bronchial tree is illustrated in Fig. 16-37. Although there is continuity of the pulmonary parenchyma between adjacent segments of each lobe, the separation of the bronchial and vascular stalks allows subsegmental and segmental resections whenever the clinical situation requires or allows preserving lung tissue. This may be particularly important in patients with impaired pulmonary function or in those with disease processes that are apt to become multifocal, requiring multiple resective procedures. Less than lobar resections are desirable when dealing with localized inflammatory diseases such as tuberculosis and bronchiectasis that characteristically involve segmental units of the upper and lower lobes, respectively, but often do so in a way that leaves one or more segments of the same lobe unaffected. Both these diseases, as well as metastatic pulmonary neoplasms, may involve more than one pulmonary lobe, synchronously or metachronously. Many surgeons doubt the necessity of extending the resection even of primary lung neoplasms beyond the field necessary for adequate margins around the tumor. The advantages of a segmental concept of surgical treatment are important in all these circumstances.

Lymphatic Drainage

Abundant lymphatic vessels are located beneath the visceral pleura of each lung, in the interlobular septa, in the submucosa of the bronchi, and in the perivascular and peribronchial connective tissue. The lymph nodes that drain the lungs are divided into two large groups, the pulmonary lymph nodes and the mediastinal nodes, referred to as N1 and N2 nodes, respectively, in the TNM system of staging of lung cancer (Fig. 16-38). In turn, the pulmonary (N1) lymph nodes consist of (1) intrapulmonary or segmental nodes that lie at points of division of segmental bronchi or in the bifurcations of the pulmonary artery, (2) lobar nodes that lie along the upper-, middle-, and lower-lobe bronchi, (3) interlobar nodes situated in the angles formed by the bifurcation of the main bronchi into lobar bronchi, and (4) hilar nodes located along the main bronchi.

The interlobar lymph nodes lie in the depths of the interlobar fissure on each side and have special surgical significance because they constitute a lymphatic sump for each lung, referred to as the *lymphatic sump of Borrie* (Fig. 16-39). This designation derives from the fact that all the pulmonary lobes of the corresponding lung drain into that group of nodes. On the right side, the nodes of the lymphatic sump lie around the bronchus intermedius bounded above by the right upper lobe bronchus and below by the middle lobe and superior segmental bronchi. The lymphatic sump on the left side is confined to the interlobar fissure, with the lymph nodes disposed in the angle between the lingular and lower lobe bronchi and in apposition to the pulmonary artery branches.

The mediastinal (N2) lymph nodes consist of four principal groups: (1) anterior mediastinal, (2) posterior mediastinal, (3) tracheobronchial, and (4) paratracheal. The anterior mediastinal nodes are located in association with the upper surface of the pericardium, the phrenic nerves, the ligamentum arteriosum, and the left innominate vein. Within the inferior pulmonary ligament on each side are found the paraesophageal lymph nodes that constitute a major part of the posterior mediastinal group. Additional paraesophageal nodes may be located more superiorly between the esophagus and trachea in the region of the arch of the azygos vein.

The tracheobronchial lymph nodes are made up of three subgroups that are located about the bifurcation of the trachea. Included are the subcarinal nodes, the lymph nodes lying in the obtuse angle between the trachea and each mainstem bronchus, and a few nodes that lie anterior to the lower end of the trachea. The paratracheal lymph nodes are located in proximity to the trachea in the superior mediastinum. Those on the right side form a chain with the tracheobronchial nodes inferiorly and with some of the deep cervical nodes above. A few of the latter are referred to as the *scalene lymph nodes* because they lie on the anterior scalene muscle. Lymphatic drainage of the right lung is ipsilateral except for an occasional incidence in which drainage to the superior mediastinum is bilateral. Drainage from the left lung to the superior mediastinum is as frequently ipsilateral as it is to the opposite side.

Diagnostic Tests

Although the importance of an accurate history and physical examination cannot be overemphasized, diagnostic tests play an integral part in the investigation and management of pulmonary disorders. Because the lung is an air-containing organ, relatively simple radiologic examinations can be used to great advantage. With increasing frequency, both chest x-rays and CT scans are used to investigate most pulmonary disorders. The communication between oropharynx and respiratory system allows secretions to be collected from the lung and airways for bacteriologic and cytologic examination and direct endoscopic examination of the tracheobronchial tree.

Sputum examination

In many of the acquired pulmonary disorders, sputum collection and examination are indicated as an initial diagnostic procedure. In inflammatory conditions, a specific etiologic agent can be identified by bacteriologic examination. Unfortunately, at times, the patient cannot expectorate sputum but provides saliva contaminated with oral flora, making a bacteriologic diagnosis difficult. When necessary, the oropharynx can be bypassed to obtain appropriate bacteriologic specimens. A percutaneous transtracheal or cricothyroid insertion of a sterile needle connected to an appropriate syringe can obtain secretions directly from the trachea. This is facilitated by injecting 5 to 10 mL of a sterile saline solution, which promotes coughing and allows aspiration of these secretions.

In investigating malignancies, sputum cytology is most appropriate when the tumor is in the proximal portion of the tracheobronchial tree (inner one-third of the lung) or in major airways (e.g., distal trachea and main bronchi).

The diagnostic yield from sputum samples can be increased by obtaining the specimen early in the morning when the patient is most likely to produce a large volume pooled from the peripheral airways. Serial specimens obtained on separate days also may improve diagnostic yield.

Imaging Studies

The standard posteroanterior chest film together with a companion lateral view remains the most frequent initial imaging study. When correlated with symptoms, physical findings, and, where available, previous radiographs, it can provide much of the information needed for the diagnosis and management of most lung disorders.

CT scanning has been used extensively in order to further define the abnormalities seen on chest x-ray. As well, CT scanning does allow investigation of the various mediastinal structures that cannot be isolated on a routine chest x-ray. This is especially important in investigating hilar abnormalities and assessing spread of malignancies to the lymph nodes in the mediastinum. Most patients now undergo CT scanning when surgical therapy is being considered in the management of their pulmonary disease.

Magnetic resonance imaging (MRI) does have value in investigating thoracic disorders, but in most instances of lung disease, this examination is unnecessary. MRI has its greatest value in assessing invasion of contiguous structures, including the vertebral bodies, spinal canal, brachial plexus, subclavian vessels, aorta, and chest wall.

Isotope studies are used less frequently. In assessing lung function, quantitative ventilation-perfusion scanning is of prime importance, especially when surgical resection of functioning lung in a pulmonary- compromised individual is anticipated. The percentage of functioning lung to be removed can be calculated and used in conjunction with pulmonary function tests to estimate the residual lung function after resection. In addition, perfusion scanning continues to be an important means of diagnosing suspected pulmonary embolism.

Other scans are used in investigating sites of metastases prior to the treatment of cancer patients. Routine organ scanning is not cost-effective before treatment, but when symptoms suggest abnormalities in distant organs, bone scans and CT or MRI scans of the brain are useful in detecting metastatic disease in these sites.

Positron emission tomography (PET) has been used to diagnose lung cancers and to detect metastases. This form of radionuclide scanning using fluorodeoxyglucose (FDG) is based on the principle that tumors take up and metabolize more glucose than do normal tissues and therefore appear as "hot spots" on FDG-PET scans. The proper application of this new imaging modality is being investigated. It has significant value in diagnosing cancer in an abnormality seen on chest x-ray and in identifying occult sites of metastases in

regional lymph nodes and other organs. In the future, it is possible that a single whole-body PET scan could replace doing multiple other scans in staging patients with thoracic malignancies (Fig. 16-40).

More invasive imaging plays less of a role today than it did previously. Although pulmonary angiography is still valuable, especially in confirming the diagnosis of pulmonary embolus and congenital vascular abnormalities, it has been supplanted largely by improved techniques of MRI and high- resolution CT. Similarly, imaging of the PP265">airway (contrast bronchography) has been replaced by thin section helical CT that can be spacially reconstructed to delineate airways.

Occasionally, bronchial arteriography has value in identifying and treating sites of bleeding from the airway. The bronchial arteries arising from the proximal descending aorta can be selectively catheterized, and abnormal bronchial arteries or bleeding bronchial arteries can be identified using routine percutaneous arteriographic techniques. This is valuable in diagnosis but also has become an integral part of the management of massive hemoptysis by embolizing these vessels to arrest the bleeding.

Bronchoscopy

Bronchoscopy is the most valuable invasive modality used in the investigation of pulmonary disease. With the advent of fiberoptic instruments 30 years ago, intubation of the tracheobronchial tree has become simple and expedient. Prior to this, rigid open tubes were required to visualize the tracheobronchial tree. These required general anesthesia and, because of their rigidity, limited the scope of investigation to the major airways (Fig. 16-41). The flexible fiberoptic equipment measures only a few millimeters in diameter and allows for intubation of the tracheobronchial tree using local anesthesia and mild sedation. These qualities extend the scope of bronchoscopy immensely (Fig. 16-42). Tiny channels within the bronchoscope allows for direct acquisition of secretions, cytologic brushings of abnormalities, and direct biopsy. This is especially important when investigating malignancies of the tracheobronchial tree and identifying the sites of bleeding when investigating hemoptysis.

Peripheral pulmonary nodules and diffuse lung disease can be diagnosed using fiberoptic bronchoscopic equipment combined with fluoroscopic imaging. Using fluoroscopic control, fine biopsy forceps and needles can be passed into the periphery of the lung to obtain biopsies of peripheral airways and subjacent lung tissue. As well, nodules of larger size (usually >3 cm) can be biopsied with this type of fluoroscopic control.

The bronchoscope not only plays an important role in diagnosis, it is also frequently required for therapeutic intervention. It is in these instances that the rigid "open" bronchoscope frequently is required, e.g., for removing foreign bodies or obstructing tumors or inserting or managing hemoptysis. In these situations, the fiberoptic bronchoscope is often used in conjunction with the rigid scope to remove secretions, inspect peripheral airways, or assist in laser therapy. The procedure is performed under general anesthesia with the patient ventilated via the rigid scope, and the flexible bronchoscope is passed through the rigid scope.

Transthoracic Biopsies

When investigating undiagnosed pulmonary lesions, percutaneous transthoracic biopsies using fine needles have proved to be exceptionally effective in obtaining material for bacteriologic and cytologic examination. In most instances, this is performed under fluoroscopic or CT-guided imaging by specially trained interventional radiologists. Diagnostic accuracy for malignancy approaches 95 percent when adequate specimens can be obtained (Fig. 16-43). Because this approach requires penetration of the lung, pneumothorax is a common complication (up to 30 percent) but usually is asymptomatic and requires no treatment.

Scalene Node Biopsy

Because of the predilection of lung cancer to spread to mediastinal lymph nodes and ultimately supraclavicular (scalene) lymph nodes (Fig. 16-44), in the past, before the newer imaging techniques previously discussed were available, biopsies of scalene lymph nodes (Daniel's procedure) often were done to rule out metastases that would contraindicate a surgical operation in dealing with patients with lung cancer. These "blind" scalene lymph node biopsies no longer are used. When physical examination or imaging reveals supraclavicular adenopathy, percutaneous needle aspiration biopsy or open scalene lymph node biopsy is necessary to identify metastatic disease in the palpable nodes.

Mediastinoscopy

When investigating diseases where the CT scan identifies enlarged mediastinal lymph nodes, mediastinoscopy has become the procedure of choice. This examination requires a suprasternal incision to expose the trachea and then insertion of an open tube instrument (mediastinoscope) into the mediastinum anterior to the trachea where the superior mediastinal lymph nodes draining the lung are located (Fig. 16-45). Because the results of surgery alone for patients with involvement of superior mediastinal lymph nodes are poor, most surgeons prefer to identify this before consideration of surgical resection of a lung cancer. Most physicians consider a 1-cm transverse diameter on CT scan as the highest limits of normal sized lymph nodes in this location.

This technique (mediastinoscopy) also is extremely useful in identifying other causes of mediastinal lymphadenopathy, e.g., sarcoidosis, lymphoma, granulomas, and other inflammations.

Parasternal Mediastinotomy

Mediastinoscopy cannot reach those lymph nodes which drain to the anterior mediastinum. This is extremely important in diseases of the left upper lobe, where the usual lymph node drainage is to the aortopulmonary window, and in diagnosing anterior mediastinal tumors. In this instance, parasternal mediastinotomy (Chamberlain procedure) can be used for diagnosis. A short transverse incision is made, usually in the second intercostal space or through the bed of the second costal cartilage (after removing it), in order to reach the paramediastinal area. The mediastinoscope can be introduced through this approach to biopsy lymph nodes along the aortic arch and in the aortopulmonary window just beneath the arch. Alternatively, the second costal cartilage can be excised, the pleura and lung retracted laterally, and the aortopulmonary window exposed directly. A similar approach can be used on the right side to biopsy tumors or lymph nodes anterior to the superior vena cava.

Thoracoscopy

Almost 100 years ago, thoracoscopy was performed using a cystoscope to examine the pleural space in order to diagnose the various causes of pleurisy. This technique was adapted to perform "collapse" therapy for pulmonary tuberculosis. Since that time, surgeons have used open-tube instruments (e.g., mediastinoscope) or cystoscope-like instruments to examine and biopsy the pleural space. With improved imaging, optical, and surgical equipment and the ability to unilaterally collapse the lung, video- assisted thoracoscopy has become an extremely valuable technique not only for investigating the pleural space but also for inspecting the mediastinum and obtaining lung biopsies. With advanced instrumentation, one, two, or three small intercostal incisions can be made, the pleural space and surface of the lung can be inspected, and using appropriate grasping and stapling equipment, biopsies can be obtained not only of the pleura but also of the lung and mediastinum. This has become especially valuable in investigating diffuse lung disease (e.g., sarcoidosis, interstitial pneumonia) and for removing peripherally placed pulmonary nodules for diagnosis.

Open Lung Biopsy

Before the advent of video-assisted thoracoscopic techniques, open lung biopsy was used extensively in obtaining lung biopsies for diagnosing interstitial lung disease. On occasion, especially in those patients in whom lung collapse is not appropriate (e.g., ventilated patients with acute disorders), open lung biopsy is the approach of choice. This can be performed very simply using a short intercostal incision directed to the area of interest in the lung and inserting stapling instruments to perform the biopsy.

Pulmonary Resections

In the surgical management of lung disease, diagnostic or therapeutic, portions of pulmonary parenchyma frequently are excised. Depending on the type of disease and the reason for removal of lung tissue, large or small portions of lung may be removed up to and including a complete ipsilateral lung (pneumonectomy).

After pulmonary resections, the remaining lung hyperexpands, and the diaphragm and mediastinum shifts in order to decrease the size of the hemithorax, thereby accommodating the loss of pulmonary tissue. Chest tubes attached to underwater drainage systems and, frequently, suctioning facilitate this and allow any parenchymal leaking areas to heal.

In the case of pneumonectomy, serum begins to fill the empty hemithorax and, over a few weeks, completely replaces the empty space created by the pneumonectomy. As with other resections, the affected hemithorax decreases in size by shifts in the mediastinum and elevation of the ipsilateral diaphragm.

Wedge Resection

(Fig. 16-46) This type of resection removes a small portion of peripheral parenchyma. It is performed most frequently for diagnosing diffuse lung disease or solitary pulmonary nodules. In many instances, video-assisted thoracoscopic techniques can be used. To maximize hemostasis and closure of air leaks, staplers are used to perform the biopsies. Larger portions of lung parenchyma also can be removed nonanatomically using this stapling technique. This is frequently used in the surgical treatment of emphysema (bullectomy and lung volume–reduction surgery).

Segmental Resection

(Fig. 16-47) Segments or subsegments of lobes can be removed in a more anatomic fashion by isolating and dividing the segmental bronchus and artery and then removing the lung in an intersegmental plane. This removal may be facilitated by using mechanical stapling equipment or dividing lung tissue and sewing the opened parenchyma. The intersegmental plane is defined by the pulmonary veins, and the segment to be removed may be stripped from the remaining lung along this intersegmental plane. Segmentectomy is best used in treating localized inflammatory diseases or small peripheral tumors limited to one segment where the patient cannot tolerate a larger resection.

Lobectomy

Any of the five lobes of the lung can be removed in an anatomic fashion. In all such resections, the lobar bronchus is divided, and the pulmonary arteries and veins to the lobe are individually ligated and divided. Frequently, the fissures between the lobes are incomplete and have to be divided before completing the resection. This is the most frequent type of resection used in treating localized lung tumors.

Pneumonectomy

Removing a total ipsilateral lung is reserved for those instances where lesser resections cannot be accomplished, e.g., tumors involving the hilum of the lung and inflammatory diseases where total destruction of the lung has occurred. In these resections, care must be taken to ensure that preoperatively the patient has been found to have sufficient lung function to tolerate such a resection. At the time of surgery, the main pulmonary artery must be divided and ligated as well as both pulmonary veins. In most instances, the final division is the main bronchus, which is divided and closed by suture or stapling devices. Because of the risk of bronchopleural fistula, this bronchial closure is sometimes reinforced with adjacent vascularized tissue.

Developmental Anomalies

Many developmental anomalies of the lung present at birth, resulting in stillborn infants or infants with severe respiratory distress. These include lobar emphysema, cystadenomatoid malformation, sequestration, pulmonary agenesis or severe hypoplasia, and giant bronchogenic cysts. These may require urgent or emergent treatment in the newborn period. Less severe anomalies are certainly compatible with life and may not present until later, even as late as adulthood. Many are mild anatomic variants and may never require treatment. An interesting and exciting development is the emergence of intrauterine diagnosis with ultrasound. Many of these life-threatening lesions may, in the future, be diagnosed and treated during intrauterine growth (fetal surgery).

Anatomic Variants

Many anatomic variants can occur because of abnormal intrauterine development. Most of these are interesting but inconsequential. The most common of these, the *azygous lobe*, occurs in 0.5 percent of individuals. The azygous vein lies in the substance of the right upper lobe on a pleural mesentery separating this “azygous lobe” from the remainder of the lung. This presents radiologically as a separate apical lobe appearing as an “inverted comma” in the medial apex of the right upper lobe. Occasionally, repeated infections can occur in this isolated area. *Situs inversus* is a rare entity in which the thoracic viscera (*situs inversus thoracis*) or the thoracic and abdominal viscera (*situs inversus totalis*) undergo complete mirror- image reversal in position during development. Although an important entity to recognize, this total reversal of organs rarely causes a problem unless associated with other conditions. *Kartagener's syndrome*, a familial association of situs inversus with chronic sinusitis and bronchiectasis, has been identified as being associated with abnormal ciliary anatomy and function causing repeated infections of sinuses and bronchi.

Occasionally, *aberrant bronchial origins* are present. Most frequently the right upper lobe bronchus originates from the lateral wall of the trachea. This is an important anatomic abnormality to identify at bronchoscopy before operative intervention.

Agenesis and Hypoplasia

Arrest in embryonic development can result in abnormalities of the lung varying from bilateral total pulmonary agenesis to hypoplasia of a portion of one lung. In about 50 percent of cases, other congenital anomalies coexist. Complete bilateral pulmonary agenesis is incompatible with life. Unilateral agenesis, when not associated with other life-threatening anomalies, is compatible with a relatively normal life. It usually occurs on the left side, with the right lung filling both hemithoraces.

Severe hypoplasia usually results in stillbirth or rapid death. In those patients presenting with severe respiratory distress and unassociated with other life-threatening congenital anomalies, neonatal pulmonary transplantation may offer hope for long-term survival. Unilateral hypoplasia often is associated with other space-filling congenital defects, e.g., diaphragmatic hernia that restricts intrauterine lung growth. An interesting variant of hypoplasia occurs in the right lung; the *scimitar syndrome* includes hypoplasia of the right lung, anomalous pulmonary venous drainage that enters the inferior vena cava instead of the left atrium, and an aberrant pulmonary arterial origin from the aorta. The venous drainage abnormality produces the “scimitar sign” outlining the venous drainage into the inferior pulmonary vein.

Cystic Adenomatoid Malformation

This anomaly is characterized by overgrowth of bronchioles resulting in cystic overgrowth of terminal ventilating units and usually presents as respiratory distress in the newborn. However, other congenital abnormalities, prematurity, and stillbirth are commonly associated. In the neonate presenting with respiratory distress, the chest x-ray shows a multicystic “swiss cheese” configuration with over-expansion of the involved area, shifting the mediastinum and compressing the normal lung (Fig. 16-48). The lesion can be diagnosed during intrauterine life and has been with increasing frequency using prenatal ultrasound examinations. Attempts have been made to treat this with prenatal intervention, either aspirating the cystic area or resecting it. There has been some success using this approach.

The postnatal management in surviving patients is resection of the diseased lobe. On occasion, these patients present beyond the neonatal period with slowly increasing respiratory problems, and the diagnosis and treatment are delayed. With a successful early operation, the prognosis is excellent, barring other life-threatening congenital anomalies.

Thoracotomy reveals a dense, meaty mass studded with cysts that may be partially aerated but have no ventilatory function. Anomalous vessels are very rare. Histopathology reveals multiple components of respiratory tissue, including a maze of irregular tubules resembling fetal bronchioles lined with disorganized respiratory epithelium that can resemble an adenoma. Lobectomy is the treatment of choice, although pneumonectomy has been necessary in some cases. The prognosis is excellent with a successful early operation.

Pulmonary Sequestration

During development, a portion of lung may be isolated from the remainder of the lung and receive its blood supply from an aberrant branch of the aorta instead of the pulmonary artery.

Intralobar sequestrations rest within a lobe and do not have their own visceral pleural envelope but usually have a communication with the tracheobronchial tree producing a cystic appearance. They occur invariably in the posterobasal portion of the lower lobes, most frequently on the left side. The sequestered lung is supplied by an anomalous branch of the aorta that usually is tortuous and disproportionately large, arising from the thoracic aorta in 70 percent of patients and from the abdominal aorta or its branches in the remainder. The vessel reaches the sequestered lung by passing through the inferior pulmonary ligament. Venous drainage is through the pulmonary veins. The radiographic appearance is that of a dense mass usually containing cysts with air-fluid levels, indistinguishable from cystic adenomatoid malformation or other congenital cystic abnormalities without an arteriogram (Fig. 16-49). Intralobar sequestrations frequently present as asymptomatic lesions or recurrent localized infections, often in the young adult. When these occur in the posterior basal areas of the lung, the surgeon should be suspicious. Confirmation can be made by CT scan or aortography to identify the anomalous artery (Fig. 16-50 and Fig. 16-51). When symptomatic, surgical resection is the treatment of choice, usually by lobectomy. Care must be taken to identify and ligate the anomalous arterial supply (Fig. 16-52).

Extralobar sequestration is a much less common entity in which the sequestered lung is enclosed by a separate pleural envelope sitting in the inferoposterior mediastinum adjacent to lung and esophagus, usually on the left side (Fig. 16-53). Association with other congenital anomalies, especially diaphragmatic hernia, occurs in over 50 percent of patients. Although tracheobronchial communication is said not to occur, an esophageal communication is seen occasionally. Extralobar sequestration is more likely to be detected early in childhood, appearing as an unexplained dense triangular mass in the posterior lower lung field (Fig. 16-54). Venous drainage is to the azygous or portal venous system. When this anomaly is symptomatic, resection of the abnormal lobe is indicated.

An *accessory lobe* is distinguished from sequestration by the fact that it communicates with the trachea or a major bronchial branch, most commonly on the left side. The accessory lobe can be supplied by the pulmonary or systemic circulation.

Congenital Cysts

These are a diverse group of abnormalities that can be single or multiple and vary greatly in size. They usually are confined to a segment or lobe and almost invariably present with infection. Their cause is poorly understood and may be equally diverse. Unilateral absence of a pulmonary artery branch may produce hypoplastic development of the lung, leading to cystic transformation. The lesions may begin as intrapulmonary bronchogenic cysts, which are thought to represent remnants of developing bronchial buds pinched off in the lung periphery. A region of cystic lung can develop distal to congenital or acquired bronchial obstruction.

The cysts are lined typically with respiratory epithelium and are filled with a viscid opaque fluid until they develop communication with the airway, after which they become partially air-filled and infected. The presence of respiratory epithelium is thought to indicate true congenital origin, but once chronic infection destroys the epithelium, it becomes impossible to separate a congenital cyst from a chronic pulmonary abscess or bronchiectasis, grossly or histopathologically.

Resection is indicated for large or chronically infected cysts. Preoperative evaluation includes bronchoscopy and occasionally arteriography to exclude sequestration. On occasion, massively enlarged cysts present as neonatal respiratory distress requiring urgent excision (Fig. 16-55).

These bronchogenic cysts also can present as mediastinal masses in the paratracheal or subcarinal regions. The diagnosis often is delayed until adult life. When they enlarge, they can produce airways obstruction. There is controversy as to whether or not such cysts should always be removed to avoid acute airway obstruction in the future. These asymptomatic cysts can now be approached through minimal-access surgery, e.g., video- assisted thoracoscopy or mediastinoscopy (see Fig. 16-50).

Arteriovenous Malformation

Pulmonary arteriovenous malformation is a fistula between pulmonary arteries and pulmonary veins. One or more thin-walled saccular channels with an endothelial lining are present without reaction in the surrounding lung tissue. The lesions are somewhat more frequent in the lower lobes. Multiple small (<1cm) lesions associated with capillary abnormalities elsewhere, a feature of hereditary hemorrhagic telangiectasia (Osler- Weber-Rendu syndrome), account for half of all reported cases. In the other half, the lesions are singular or few in number and larger (1 to 5 cm) in size. Although the vascular pattern is variable, one afferent pulmonary arterial branch with two or more efferent venous branches is the most common arrangement (Fig. 16-56). The lesions frequently are very superficial and vulnerable to erosion, and they can present with spontaneous hemothorax.

Diagnosis is easiest in the 20 percent of patients presenting with cyanosis, polycythemia, and clubbing. Cyanosis is said to develop when the shunt fraction exceeds 25 percent of the total blood flow and usually is not seen until adolescence or early adulthood.

There is an interesting difference between the right-to-left shunt seen in congenital heart disease and that seen with a pulmonary arteriovenous malformation. In the latter condition, vascular resistance in both the pulmonary capillary bed and the fistula is negligible, so the total volume returned to the left atrium is not increased—it is simply desaturated in proportion to the amount of flow going through the fistula. In contrast to the patient with an intracardiac shunt, the patient with the fistula has normal cardiac output, pulse, blood pressure, venous pressure, ECG, and heart size. No murmur is audible over the heart, but in over half the cyanotic patients a continuous murmur can be heard peripherally over the fistula. Patients with either condition have arterial desaturation and polycythemia.

Diagnosis is more difficult in the asymptomatic, acyanotic patient. Small, single malformations can be indistinguishable from all other solitary pulmonary nodules. Larger lesions may have a more characteristic lobulated appearance, and the afferent and efferent vessels often can be demonstrated on plane tomography or CT scan. Pulmonary angiography confirms the diagnosis (Fig. 16-57). Both lungs should be examined carefully for multiple lesions.

Significant complications occur in at least 25 percent of all patients and in a higher percentage of patients with hereditary telangiectasia and multiple fistulas. Complications can be local effects such as hemothorax, but consequences of polycythemia, such as cerebral thrombosis, are more common. Resection is indicated in all patients with solitary nodules and in selected patients with multiple nodules when adequate lung tissue can be preserved.

Although resection was the treatment of choice in patients with solitary nodules and in selected patients with multiple nodules when adequate lung tissue could be preserved, most

recently, invasive angiographic approaches, embolizing the fistulae with coils, balloons, and gel foam, are effective in most patients. Cyanotic patients with widespread, multiple, small fistulas are difficult to treat, although improvement with embolization has been reported in a few patients.

Lobar Emphysema

Lobar emphysema presents with massive distention of a lobe or segment that shifts the mediastinum and compresses the contralateral lung. It is the most common of the four structural lesions usually considered in the differential diagnosis of respiratory distress in the newborn. The other three are sequestration, cystic adenomatoid malformation, and bronchogenic cyst. Etiology has been a matter of continuing debate, and the clinical syndrome may represent the final common pathway of several distinct processes. Dysplasia of bronchial cartilage has been recognized most frequently, occurring in about 25 percent of patients. Acute bronchiolitis, extrinsic compression by lymph nodes or anomalous vessels, bronchial atresia, and several other possible causes of bronchial obstruction have been cited. No specific cause can be identified in more than 50 percent of patients. Associated malformations, mostly cardiovascular, are present in about 40 percent of patients, a fact that some authors interpret as support for a congenital etiology.

Respiratory distress typically appears from 4 days to 6 months postpartum. Half of all reported cases have occurred within the first 4 weeks of life. Physical findings of hyperresonance and decreased breath sounds mimic pneumothorax, which is a dangerous misapprehension if it leads to impulsive chest tube insertion. The chest x-ray also can mimic pneumothorax, because the distended lung is very hyperlucent. Another common misinterpretation occurs when the compressed normal lung is thought to be atelectatic and the distended lung compensatory. An important radiographic clue is that the diaphragm usually is depressed on both sides with lobar emphysema but is normal or elevated on the side of primary atelectasis (Fig. [16-58](#)).

In an infant with florid, progressive respiratory distress and a characteristic chest x-ray, emergency thoracotomy and lobectomy are indicated without further study. In such circumstances, the mortality without operation approaches 50 percent. Involvement of an upper lobe or the right middle lobe is the most frequent finding. Lower lobe involvement is rare. Resection is ordinarily straightforward anatomically and completely curative. When the clinical presentation is less fulminant, the decision is more difficult. There is no question that varying degrees of lobar emphysema can be produced by aspiration of mucus or amniotic fluid or by acute bronchiolitis. In such cases the emphysema almost always resolves in a few days with appropriate medical therapy.

Emphysema

Emphysema is characterized by enlarged air-spaces produced by a complex process of elastic tissue destruction resulting in alveolar wall breakdown and coalescence of damaged alveoli, resulting in impaired alveolar ventilation and gas exchange. The disease represents one manifestation of advanced *COPD, overlapping with other manifestations including chronic bronchitis and reactive bronchial asthma. Pathogenetic differences promoting a predominant pattern of alveolar breakdown (versus bronchitis) are difficult to isolate, and the response to a common mechanism of injury such as cigarette smoking is not predictable. An exception is the diffuse emphysematous pattern produced by the loss of normal restraints on tissue destruction as seen in α_1 -antitrypsin deficiency* (Fig. [16-59](#)).

Only a few aspects of this disorder are appropriately managed by surgical resection. Emphysema is often not diffuse, and involved areas may be localized into collections of small cysts (blebs) or very large ones (bullae). Blebs usually are subpleural, do not extend deeply into more central parenchyma, and may have little effect on gas exchange or overall pulmonary function, even when multiple, discrete collections of considerable size are present. Their main significance relates to their potential for rupture with production of pneumothorax. The localized collection of blebs frequently encountered in the apices of otherwise normal lungs in healthy young individuals with spontaneous pneumothorax bears an uncertain pathogenic relationship to emphysema and may be congenital. Although most blebs are apical, the blebs encountered in typical COPD are more likely to be multiple and in other parts of the lung.

Bullae result from a much larger coalescence of destroyed alveolar septa and develop deep within the lung parenchyma, compressing and distorting adjacent normal lung. They may assume the form of a single large cyst, or remnants of interstitium may remain to form a multiloculated space. Many bullae remain stable or increase in size slowly over many years, but they have the potential for rapid expansion, producing acute respiratory distress. An enormous single bulla may be very difficult to distinguish from a tension pneumothorax clinically or radiographically, completely compressing and obliterating the remaining normal lung.

Diffuse emphysema is characterized by a more uniform destructive process producing profound effects on pulmonary function and gas exchange. The full-blown clinical presentation includes extreme dyspnea, with a barrel chest and attenuation of intercostal and diaphragmatic musculature. Lungs with diffuse emphysema frequently contain areas of bleb or bullous disease.

Surgical Considerations

The surgeon confronts emphysema in two principal situations: in operations performed on patients with emphysema and in operations performed for emphysema. In the former category, patients with emphysema undergoing procedures with general anesthesia have an increased operative risk due to their abnormal gas exchange and are at higher risk for barotrauma during mechanical ventilation. Postoperative pulmonary toilet can be a major challenge in a patient with greatly reduced expiratory forces and chronic bronchitis whose ability to cough is compromised further by postoperative pain. Every surgeon has seen patients with marginally compensated emphysema become ventilator-dependent, with a tracheostomy and bilateral chest tubes, after routine abdominal surgery. Elective operation in such patients requires cautious assessment of risks and careful attention to preoperative pulmonary physiotherapy.

Thoracic operations designed to correct specific manifestations of emphysema carry similar risks but also offer the expectation of improvement. The history of operations for emphysema illustrates the danger of allowing surgical intuition to precede an understanding of pathophysiology. Some of the earliest procedures were designed to “make room” for the hyperinflated lung by further enlarging the barrel chest with sternotomy or chondrectomy. On observation that such procedures were ineffective at best, attempts were made to decrease lung volume by phrenic nerve destruction or thoracoplasty, with counterproductive results that would be considered predictable today.

Over the past 30 years, it has become evident that certain patients can be improved by surgical extirpation of a functionless portion of lung. Initially, surgery was only directed at areas of bullous disease compressing other relatively normal areas of lung. Resection of such bullae in selected patients appears to improve these patients. Spurred by the work of Brantigan almost 40 years ago and, more recently, by Cooper and associates, it has been identified that the hyperinflated lungs produce a mechanical dysfunction of the chest wall. *Lung-reduction surgery* for diffuse emphysema has resulted in maximizing chest wall function by decreasing the size of the lungs, thereby improving chest wall mechanics and diaphragmatic function. Short-term follow-up of such patients has resulted in a salutary effect.

In all patients, every effort is made preoperatively to maximize pulmonary function and eliminate chronic bronchial infection. Pulmonary rehabilitation programs have a decidedly beneficial effect. Unilateral disease can be approached through a posterolateral or anterolateral thoracotomy, and with increasing experience, video-assisted techniques can be used. For bilateral bullous or diffuse disease, median sternotomy or bilateral anterior thoracotomies can be used. Even in this disease, surgeons adept at video-assisted techniques have used bilateral *VATS to minimize postoperative pain*.

In performing surgery for emphysema, mechanical stapling devices are extremely helpful, often adding pledgets of bovine pericardium or artificial fibers to reinforce staple lines, because planes of division through emphysematous lungs are friable and prone to air leak. All attempts are made to eliminate any residual pleural space by careful placement of chest tubes and judicious use of volume-reducing maneuvers such as “pleural tents” or thoracoplasty.

Surgery for Bullae

Selection for operation is directed toward identifying patients in whom resection of localized bullous disease is likely to improve pulmonary function (Fig. [16-60](#)).

Ventilation-perfusion scans and fine-cut contrast-enhanced CT scans or angiograms can help define areas of “normal” lung that are compressed and non-ventilated but normally perfused (Fig. [16-61](#)) and are adjacent to large, compressing bullae. The most favorable patients are young (under 55 years of age), with unilateral disease and marked asymmetry of function, recently progressive symptoms, well-defined bullae, and evidence of crowded vessels in adjacent “normal” parenchyma. Large bullae displacing more than half the hemithorax are more likely to produce symptomatic improvement after resection than smaller bullae.

The resection is carefully tailored to preserve all adjacent vascularized “normal” lung parenchyma (Fig. [16-62](#), Fig. [16-63](#)). In carefully selected patients, resection of such large, compressing bullae can be extremely worthwhile.

Percutaneous drainage of expanding emphysematous bullae has been used effectively in patients for whom surgery is contraindicated. The emphysematous bulla is intubated with a large catheter, usually under direct vision using the techniques that have already been described in the open drainage treatment of lung abscess. A small rib resection is performed under local anesthesia, and the bulla is identified and intubated after a purse-string suture has been used to secure the catheter. Although a bronchopleural fistula results, the lung will expand and, ultimately, the fistula will close. This approach should be reserved for situations in which the risk of surgical resection is prohibitive.

Surgery for Diffuse Emphysema

Lung volume-reduction surgery (LVRS) is a relatively new approach to the treatment of severely incapacitating diffuse emphysema. The objective of this operation is to reestablish normal chest wall and diaphragmatic mechanics (Fig. [16-64](#)). Patients must be selected carefully based on their pulmonary function tests, arterial blood gases, exercise tolerance (6-min walk test), and the type and extent of bullous disease. Patients with irregular areas of bullous disease do better than those with very diffuse disease. Because most of the poorly ventilated, poorly perfused lung occurs in the apices and anterior portions of the upper lobe, these are the portions of lung that resected. Careful preoperative localization of the least perfused areas of lung is required using ventilation-perfusion and high-resolution CT scanning (Fig. [16-65](#) through [16-68](#)). Bilateral and unilateral resections have been performed. Because the procedure is new, long-term results are unknown. The improvement in pulmonary function and reduction in dyspnea last for at least 2 years. To determine the ultimate role of LVRS, a randomized national trial is under way in the United States comparing non-surgical pulmonary rehabilitation with LVRS in patients with incapacitating diffuse emphysema.

Lung Transplantation

During the past 20 years, lung transplantation has become an accepted *clinical procedure. In the management of COPD, it is reserved for patients with end-stage disease who are unsuitable for LVRS because of severe hypercarbia or other reasons. End-stage emphysema is the most common indication for unilateral or bilateral lung transplantation.*

Other forms of end-stage lung disease are indications for unilateral or bilateral lung transplantation, including primary or secondary pulmonary hypertension (often combined with septal defect corrections present), idiopathic and secondary pulmonary fibrotic diseases, and end-stage bronchiectasis (usually cystic fibrosis).

In well-selected patients, usually under age 60, the results of unilateral and bilateral transplantation are encouraging. Up to 60 percent of such individuals are alive and functioning normally after 5 years. Bilateral lung transplantation provides better pulmonary function, but unilateral lung transplantation is performed more frequently because of the lack of suitable donors. In those diseases marked by chronic pulmonary infection, however, bilateral lung transplantation is required. To improve donor availability, in certain circumstances, especially in pediatric patients, in whom a single adult lobe can replace a lung, living-related donors have been used to provide the needed organ.

Pulmonary Infections

As recently as the 1960s, thousands of patients each year in the United States required pulmonary resection for lung abscess, bronchiectasis, and chronic granulomatous disease. Since that time, effective antibiotics, aggressive methods for accurate early diagnosis, an increased standard of living, and public health programs are among the factors that have diminished the surgeon's role dramatically. Fifteen or twenty years ago it appeared that suppurative pulmonary infections would be a common surgical problem only in areas of the world with limited access to care and early treatment. Such problems are once again becoming increasingly frequent in developed countries; for example, patients immunosuppressed after transplantation or as part of cancer chemotherapy and patients with the acquired immunodeficiency syndrome (AIDS) commonly develop serious pulmonary infections.

The pathologic spectrum of "surgical" pulmonary infections is very broad, ranging from the indolent bronchiolar and peribronchial suppuration of bronchiectasis, to the contained parenchymal necrosis of lung abscess, to the pleural space infection of empyema. The clinical expression of pulmonary infection is determined by the route of inoculation, the competence of host defenses, and the specific organism(s) involved, which can include aerobic and anaerobic bacteria, viruses, and fungi, often in synergistic combinations. This broad spectrum of pathology has an equally broad spectrum of treatment in which surgical management is important.

Lung Abscess

Lung abscess is a focus of infection within the lung associated with parenchymal necrosis, usually with cavitation. Distinction between a lung abscess and a consolidated pneumonia is made as area(s) of cavitation appear on the chest x-ray and as the peripheral margins of the infection develop sharper definition. Causes of lung abscess are outlined in Table [16-11](#).

It is most commonly a complication of necrotizing pneumonia, often as a result of aspiration. Aspiration of gastric contents or infected saliva produces an infectious focus due to mixed aerobic and anaerobic bacterial contamination. Frequently concomitant aspirated particulate matter helps promote the abscess formation. Aspiration has a well-recognized association with altered states of consciousness (Fig. [16-69](#)), including alcoholic stupor, drug overdosage, head trauma, cardiopulmonary resuscitation, and after general anesthesia. Because most episodes of aspiration occur with the person supine, the abscess is characteristically located in the lung segments that are dependent in the supine position, i.e., the posterior segments of the upper lobes and the superior segments of both lower lobes. Bacteriologically, the infection usually is mixed, with anaerobic mouth organisms such as *Bacteroides* species frequently predominating.

The tissue necrosis that is the hallmark of "necrotizing" pneumonia is a function of the specific organism involved and is most prominent with *Klebsiella*, *Pseudomonas*, and other gram-negative organisms. Tissue necrosis is rare with Group B streptococcal pneumonia (*Pneumococcus*), whereas necrosis and abscess formation are frequent with *Staphylococcus aureus* and Group A streptococci.

Staphylococcal lung abscess is most common in the first year of life and has characteristic pathology, most frequently with pyopneumothorax and pneumatoceles. The latter are large cystic spaces, typically not containing true pus, that are thought to result from air trapping distal to bronchiolar obstruction. Staphylococcal lung abscess in infancy also is characterized clinically by a remarkable tendency to resolve completely with antibiotics alone, even when temporary drainage of the pleural space has been required.

Establishment of a gram-negative pneumonia begins with major alteration in the bacteriologic composition of upper respiratory flora. The defense mechanisms that reduce gram-negative organisms to transient visitors in normal individuals are seriously impaired in many hospitalized patients (nosocomial infections). For example, experiments in mice suggest that an *Escherichia coli* peritonitis interferes with recruitment of polymorphonuclear leukocytes in the lung, increasing susceptibility to gram-negative (*Pseudomonas*) but not to gram-positive infections. Over half the pneumonias seen in seriously ill hospitalized patients are gram-negative, and a significant proportion of such patients manifest a necrotizing infection leading to lung abscess formation.

Systemic sepsis can produce multiple bilateral foci of parenchymal infection that are radiographically discrete and are most frequently caused by *Staphylococcus* and other gram-positive organisms. One or more of the foci can become an abscess, although most resolve without a trace. Unlike staphylococcal pneumonia of tracheobronchial origin, hematogenous infection does not form pneumatoceles and can be seen in septic patients of any age. Lung abscess developing in a pulmonary infarction after pulmonary embolus is most frequently a special case of hematogenous infection seeding an area of devitalized or injured tissue.

A cavitory necrotizing infection also can form distal to an obstructive lung carcinoma or intrabronchial foreign body (Fig. [16-70](#)), a reminder that bronchoscopy is an important diagnostic maneuver. Not infrequently, a carcinoma becomes visible as the distal infection responds to antibiotic treatment.

In certain parts of the world, parasitic infection is a common cause of lung abscess. *Entamoeba histolytica* (amebiasis) can produce a lung abscess by hematogenous spread or by direct extension from the liver, in which case it is almost always associated with an empyema. Metronidazole usually is effective treatment, and operative intervention rarely is required. Hydatid infection (*Echinococcus* species) is associated with lung abscess in some cases when cysts are secondarily infected (Fig. [16-71](#)). In this disease, intraabdominal pathology (liver cysts) is much more prominent. Treatment with oral mebendazole or albendazole has been moderately successful, but resection frequently has been required.

Clinical Manifestations

The clinical presentation of lung abscess is relatively uniform. The patient appears ill, is likely to be febrile, and often will describe recent onset of copious foul sputum production, reflecting decompression of the abscess into the airway. Whether the initial infection originated from the airway or the bloodstream, the necrotizing process finds its way into the tracheobronchial tree, a development heralded radiographically by the appearance of cavitation on chest x-ray. Because of the necrotizing, erosive nature of the communication with the airways, hemoptysis can occur and can be massive, or the abscess can burst into the pleural cavity (pyopneumothorax). In contrast to pneumonia, dyspnea is not a prominent symptom. Auscultatory findings, if any, are more likely to be attributable to coexistence of pneumonia, a pleural effusion, or empyema. In the acute phase of abscess development, constitutional symptoms overlap with those of acute pneumonia. As the process becomes more chronic, symptoms frequently ameliorate as the abscess becomes walled off.

In a febrile patient with copious production of foul sputum, the differential diagnosis can be reduced to three entities: lung abscess, bronchiectasis, and cavitating carcinoma. Chronic copious sputum production is most characteristic of bronchiectasis; the other two entities have acute or episodic sputum production. A dramatic febrile illness is most consistent with lung abscess, reflecting its origins in a necrotizing pneumonia, although bronchiectasis and carcinoma can have febrile episodes associated with exacerbations of the inflammatory process surrounding the primary pathology. A chest x-ray showing a well-delineated cavity with an air-fluid level confirms the diagnosis but usually is of little help in distinguishing a cavitating carcinoma from lung abscess, for which bronchoscopy and biopsy are essential.

Treatment

The options for treatment of lung abscess are outlined in Table [16-12](#). Primary treatment consists of antibiotics and drainage. Antibiotics are administered intravenously in high doses based on the sensitivities of the infecting organism. In the most fortunate patients, spontaneous drainage by expectoration is adequate. Frequently, drainage must be achieved by other means, the least invasive of which is bronchoscopic aspiration. Responding patients require at least 8 weeks of antibiotics. Such methods are successful in more than 75 percent of all patients with lung abscess (see Fig. [16-69](#)), although convalescence can be prolonged. In one representative nonoperative series, the average duration of therapy was 4 months. Numerous surgical series have demonstrated that convalescence can be shortened dramatically in properly selected patients.

When antibiotics and internal drainage are ineffective or in very large abscesses (>5 cm), surgical intervention is indicated. There are three options for operative treatment: (1) tube pneumonostomy (catheter drainage), (2) pneumonotomy, and (3) pulmonary resection. Before antibiotic treatment, tube pneumonostomy and pneumonotomy depended on the development of a secure pleural symphysis between the parietal pleura and the visceral pleural surface closest to the abscess cavity. Both techniques established external drainage through the area of symphysis to avoid spilling a contained infection into the free pleural space.

Percutaneous catheter drainage usually is adequate and is having a resurgence in popularity, as recent reports indicate (Fig. [16-72](#)). The technique involves insertion of a relatively small percutaneous drainage tube that usually is connected to a water seal with or without suction. Tube drainage is theoretically limited by the viscosity and particulate content of the pus and by the presence of a foreign body. As reports from vanSonnenberg and associates and Yang and associates illustrate, fluoroscopic, CT, and ultrasound guidance can be used to localize the abscess and accomplish effective drainage in a single stage. With any form of external drainage, the possibility of erosive injury to large-caliber bronchioles and vascular structures that lie close to the soft inner surface of the abscess cavity imposes a theoretical risk of bronchopleurocutaneous fistula or hemorrhage. Most abscesses are more peripheral than central, which reduces this risk, and peripheral location helps promote the kind of aggressive pleural symphysis that favors external drainage when necessary.

Pneumonotomy is almost never used today but historically avoided the risk associated with an indwelling foreign body and was favored for the most organized cavities containing large volumes of especially viscid pus mixed with large amounts of necrotic debris. Drainage was achieved through a generous incision with rib resection. If inspection of the parietal pleura suggests that adequate symphysis had not taken place, the wound was packed open without opening the pleura. After several days, this resulted in an exuberant pleural reaction and symphysis, allowing safe access to the abscess cavity. This technique is now reserved for the rare patient who fails catheter drainage and cannot tolerate a resection.

The most definitive operative treatment for lung abscess is pulmonary resection. Standard indications for resection are failure of simpler approaches, serious hemorrhage, and suspicion of associated carcinoma. Lobectomy usually is required. Resection has the advantage of removing the entire infection promptly and is less hazardous than external drainage when the abscess is very large or centrally located. On the other hand, resection does not eliminate the risk of pleural space contamination. It can be technically difficult to remove a thin-walled abscess presenting close to the visceral pleura without spillage, and empyema after lobectomy is far more serious than primary empyema. Anesthetic technique and patient positioning are critical to prevent spillage of pus through the tracheobronchial tree across to the dependent lung. In the past, the patient was often positioned prone or supine. Currently, double-lumen endotracheal tubes, which can effectively isolate the two sides, generally allow use of the lateral decubitus position. Safety can be further augmented by frequent intraoperative bronchoscopy performed through the endotracheal tube, providing accurate irrigation and aspiration of the airways.

Life-threatening hemorrhage requires prompt resection when the bleeding site is unequivocally localized by bronchoscopy. This complication most frequently arises in patients least able to tolerate thoracotomy and who often are bleeding because of coagulopathy secondary to sepsis and multiple organ failure. Acceptable temporizing measures designed primarily to protect the uninvolved lung include insertion of a double-lumen endotracheal tube, placement of a bronchial blocker on the affected side, and aggressive toilet of the unaffected side with rigid or flexible bronchoscopy. Bronchial artery embolization is worth consideration but is limited by the rich collateral circulation of the lung.

Opportunistic Infections

Cancer chemotherapy and organ transplantation are creating a steadily increasing population of immunologically compromised individuals who would not have been alive 20 years

ago. Patients nursed through major trauma or complications of surgery exhibit a characteristic spectrum of immunologic compromise. Recently, AIDS has become epidemic. All categories of immunocompromised patients share a predilection for pulmonary infections caused not only by familiar agents but also by pathogens rarely seen in healthy individuals. Pulmonary infections are the most common infections seen in immunocompromised patients, perhaps reflecting the fact that the lung is the only organ capable of presenting pathogens to a delicate nonsquamous epithelium many times each minute.

AIDS is the most relevant model of pulmonary infection in the immunocompromised host. AIDS was first described in 1981 in a group of homosexual men with *Pneumocystis carinii* pneumonia and mucosal candidiasis. The causative agent, a retrovirus named the *human immunodeficiency virus* (HIV), was identified in 1983. Although improved medical treatment has lengthened the average survival time after diagnosis, mortality is high. The disease has escaped the confines of the originally recognized risk groups (homosexual males and intravenous drug abusers) to become a venereal scourge threatening the general population.

Numerous examples of transmission through blood transfusion during the 1980s accelerated development of techniques for blood conservation, although blood donors screened using available serologic testing for the presence of HIV infection pose a very small risk (about 1 in 100,000 per unit of blood).

Opportunistic infections and Kaposi's sarcoma are hallmarks of active disease. *P. carinii*, the most common pathogen, is present in 50 to 60 percent of all patients. This is a protozoan originally thought to be a trypanosome, and it was not recognized in human beings until 1938. Characteristic interstitial pneumonitis thought to be caused by *Pneumocystis* was recognized in epidemics among undernourished infants and among the very elderly. In the United States, *Pneumocystis* emerged as an infection confined to iatrogenically immunosuppressed patients on chemotherapy or after transplantation until its dominant role as a pulmonary pathogen in AIDS was recognized in 1981. The characteristic clinical presentation consists of diffuse interstitial infiltrates on chest x-ray, dyspnea, and an increased a-aO₂ gradient.

Trimethoprim-sulfasoxazole (TMP-SFX) is effective treatment in most cases and also is used as prophylaxis. Pentamidine is equally effective but usually is reserved for resistant cases or patients allergic to TMP-SFX. Medical treatment and medical prophylaxis with TMP-SFX or aerosolized pentamidine are effective in transplant recipients in part because the level of immunosuppression can be modulated. Medical treatment is considerably less successful in AIDS patients, with mortality approaching 50 percent.

The surgeon's role in the management of immunocompromised patients with pulmonary infection can be diagnostic and therapeutic. Transbronchial biopsy combined with bronchoalveolar lavage can establish a specific infectious diagnosis, which has markedly reduced the need for open lung biopsy. In a report by Bonfils-Roberts and associates, the need for open biopsy in AIDS patients dropped from 28 percent during 1983 to 0.7 percent in 1987, and a successful change in therapy based on the results of open biopsy was possible in only 1 of 66 patients biopsied (1.5 percent). Open lung biopsy and thoracoscopic biopsy are reserved for confusing mixed infections with inadequate response to empirical treatment.

In patients with AIDS, the most frequent therapeutic procedure is insertion of a chest tube for control of a pneumothorax, which frequently complicates this infection and may be difficult to control. Several tubes usually are required (Fig. 16-73), and on occasion, thoracotomy or thoracoscopy for direct control of the leak and pleural abrasion or pleurectomy are justified (Fig. 16-74). Gerein and associates illustrate that satisfactory palliation can be achieved with operation, but the overall prognosis is very poor and is dictated by the activity of the underlying lung disease rather than by the surgical treatment.

Operative therapy in immunocompromised patients without AIDS can be applied with greater optimism. The spectrum of infections includes *Pneumocystis* and the other organisms seen in AIDS but also includes a greater number of infections with more common pyogenic bacteria and *Aspergillus*. An open lung biopsy is required if other diagnostic attempts fail. Empyemas and intraparenchymal infections should be approached aggressively using indications for operation that are the same as those applied to normal hosts with similar infections. Although mortality and morbidity generally have been higher in immunosuppressed patients, it has been possible to perform major pulmonary resection, thoracoplasty, and decortication with excellent long-term results (Fig. 16-75).

Bronchiectasis

Bronchiectasis is characterized by dilatation of peripheral bronchi, recurring pulmonary infections, and a chronic course. Second- to fourth-order segmental bronchi in the basal segments of the lower lobes, the right middle lobe, and the lingula are involved most frequently. Isolated upper lobe involvement is very rare and usually is associated with tuberculosis or bronchial obstruction. Approximately one-third of bronchiectasis is unilobar, one-third is unilateral bilobar, and one-third is bilateral. Although the bronchial mucosa usually remains intact and is lined with pseudostratified columnar epithelium, the bronchi are filled with mucus, pus, and an occasional broncholith resulting in destruction and dilatation of the underlying bronchial wall. The changes vary in degree from mild tubular dilatation to cystic or saccular changes with almost unrecognizable gross architecture (Fig. 16-76). Collateral ventilation is only partially effective in maintaining expansion of alveoli distal to chronically obstructed segments. Hypertrophy of bronchial arteries occurs as part of the inflammatory process, producing a locally extensive precapillary left-to-right shunt into the pulmonary venous system and laying the substrate for erosive hemorrhage and hemoptysis.

In most cases, the disease has to be considered idiopathic, although, in the past, it was considered often to be due to suppurative diseases in infancy (e.g., whooping cough). It is occasionally associated with chronic bronchial obstruction by tumor, foreign body, or bronchostenosis. Immunodeficiency diseases have been implicated in certain instances. Kartagener's syndrome (situs inversus, pansinusitis, bronchiectasis) is a rare congenital disorder possibly related to a defect in ciliary structure and function. In many patients afflicted during childhood, a history of recurrent bronchitis and bronchopneumonia, presumed to be viral, is present. Progressive bronchiectasis is a hallmark of cystic fibrosis with progressive respiratory failure leading to death.

Clinical Manifestations

The clinical picture is dominated by cough and production of mucopurulent sputum, varying in volume from scant to as much as 500 to 1000 mL/day. Fever usually is low-grade with acute exacerbations. The systemic effects of chronic infectious illness can dominate the picture to produce a broad spectrum of constitutional symptoms, weight loss, and retarded development. The disease can occur at any age and is seen equally in both sexes. In the United States, an unusually high incidence has been identified in Alaskan native children, many of whom have required aggressive surgical treatment (Fig. 16-77). Dyspnea is not common except in diffuse disease or in progressive disease with cor pulmonale. Hemoptysis occurs in about 50 percent of patients, usually late in the disease, and is only major in about 10 percent. Serious hemoptysis is more frequent in association with lung abscess than with bronchiectasis.

Physical findings are dominated by stigmata of chronic disease and can include digital clubbing and pulmonary osteoarthropathy, even though cyanosis is rare. Auscultatory findings are primarily related to the presence or absence of associated pneumonitis and to the effectiveness of pulmonary toilet. A history of chronic profuse sputum production strongly suggests the diagnosis, and in children, associations such as cystic fibrosis, immune deficiency, and a α_1 -antitrypsin deficiency should be ruled out. Chest x-rays are nonspecific but may show linear streaking and atelectasis in the affected areas. Bronchiectasis is one possible explanation for the "middle-lobe syndrome," which is isolated middle-lobe atelectasis (Fig. 16-78).

Diagnosis

Bronchoscopy is done to exclude the rare case of correctable bronchial obstruction or carcinoma, to obtain accurate cultures, and to aspirate the tracheobronchial tree. Careful bronchoscopic pulmonary toilet can achieve surprisingly durable symptomatic benefit. Contrast bronchography used to be the definitive test in defining the anatomy when resection was contemplated but now has been replaced by CT scanning. A report by Munro and associates compared thin-section (3-mm) CT scanning with bronchography and found equivalent accuracy for the two techniques. With modern computer-generated reconstruction techniques, the CT scan can simulate a bronchogram.

Treatment

The majority of patients with bronchiectasis do not require operative treatment. Postural drainage and chest physical therapy (see Fig. 16-79) minimize retention of purulent sputum, and antibiotic treatment of all episodes of pneumonitis should be pursued indefinitely. When debilitating effects of chronic infection become prominent and surgery is contemplated, the anatomy should be defined carefully with CT scan before resection is planned. When extensive saccular disease is confined to one lobe or segment in a sufficiently symptomatic patient, resection is a clear choice. In children, interference with growth is often the sign of debility. Frequent hemoptysis associated with localized disease deserves operation. Patients with diffuse bilateral disease, of which cystic fibrosis is the best example, rarely benefit from resection and should be considered for lung transplantation as they approach end-stage disease.

Before resection, all patients should receive a maximal effort to reduce sputum volume and infection. Care must be taken during anesthesia to prevent spillage of infected secretions into uninvolved segments. A double-lumen endotracheal tube is used to protect the contralateral lung. As with operation for lung abscess, the risk of postoperative empyema is higher than for "clean" surgery, and considerable effort must be expended in postoperative pulmonary toilet to ensure that residual infected secretions are expectorated. The operative strategy is to remove as little normal lung as possible without entering the central focus of infection. This usually requires segmentectomy or lobectomy and depends on an accurate preoperative assessment of the diseased segments. Pneumonectomy is rarely indicated in this disease.

As with many clinical situations in which choice of therapy is based largely on quality-of-life decisions, there are no prospective, controlled series to compare relative benefits of medical versus surgical treatment. In general, isolated areas of bronchiectasis do well after resection. In a Turkish series of 487 patients treated over a 12-year period, including 190 treated with pneumonectomy, mortality was 3.5 percent, and 71 percent were asymptomatic (follow-up 4 months to 10 years). Surprisingly good quality of life has been described by Laros and associates from the Netherlands in 30 patients treated with unusually extensive bilateral resections (11 to 13 segments) and followed for 30 years. As a comparison, other authors have estimated that 70 percent of patients treated medically develop persistent or progressive symptoms.

Tuberculosis

Sanskrit written in 6000 b.c. refers to tuberculosis as the king of diseases. Tuberculosis, then called *phthisis*, was well known to Hippocrates. A generation beginning to face an uncertain battle with the AIDS complex would do well to recall that pulmonary tuberculosis was epidemic in Europe during the eighteenth and nineteenth centuries and took an extraordinary toll on young adults in the prime of life, much like the devastation of AIDS today. In the United States in the 1940s, pulmonary resection for tuberculosis carried a mortality rate of about 25 percent, and effective chemotherapy did not exist until streptomycin was discovered in 1944. Tuberculosis remained epidemic in parts of the United States (Alaska) as recently as the early 1960s. Until recently, the modern treatment of tuberculosis was largely reduced to a straightforward recipe of medical therapy using combinations of antibiotics.

The past decade has seen a recrudescence of tuberculosis as a significant public health concern. Noncompliance with drug therapy among those populations at particular risk, patients in low socioeconomic strata and drug addicts, has given rise to a more virulent multiple-drug-resistant (MDR) organism. The expanded population of immunocompromised patients, whether from AIDS or from posttransplantation drug therapy, has increased the pool of the highly susceptible. Many authorities are anticipating a return to the need for relatively long-term isolation of the MDR infected patient with active disease and a resurrection of the challenging, if nearly forgotten, surgical techniques characteristic of the

pre-drug treatment era.

Pathophysiology and Clinical Manifestations

A pulmonary infection with *Mycobacterium tuberculosis* behaves like a lung abscess, with notable differences based primarily on the peculiar growth characteristics of the organism and the nature of the host response. The primary infection usually is asymptomatic and “heals” spontaneously. The disease usually becomes clinically apparent when a previously acquired and quiescent infection is reactivated, usually in the apical or posterior segments of an upper lobe or in the superior segment of a lower lobe. In an immunocompetent host, the characteristic cycle of caseous necrosis and scar formation (healed primary infection) eventually produces what amounts to a chronic tuberculous lung abscess. Just as with pyogenic lung abscess, the smoldering central focus of infection tends to find communication with the tracheobronchial tree, providing a route for drainage and expectoration of purulent sputum loaded with tubercle bacilli and allowing ingress of air to produce cavitation. The ultimate extent of infection is determined by the size of the inoculum, the immune competence of the host, and the success of antituberculous drugs. Rapid progression can produce a tuberculous empyema surrounding a destroyed lung. If growth of the bacillus is controlled, the cavity may collapse and obliterate or may remain open indefinitely. Such cavities (“open negative”) are no longer sites of tuberculous infection but remain potential sites for secondary infection, the classic example of which is an *Aspergillus* “fungus ball” (mycetoma).

As with lung abscess and bronchiectasis, the intense inflammatory process in the periphery of a cavity tends to promote hypertrophy of bronchial arterial and pulmonary arterial branches. These may be eroded by the necrotizing process in the center of the cavity to produce hemoptysis, which can be life-threatening. The vessel responsible is most frequently a dilated bronchial arterial branch, referred to as a *Rasmussen aneurysm*. The management of massive tracheobronchial hemorrhage has been discussed in connection with lung abscess. Because a tuberculous cavity is likely to be chronic, when major hemoptysis occurs, the vessel involved is likely to be larger than that seen in a pyogenic lung abscess, and emergency pulmonary resection often is necessary.

Occasionally, the most intense inflammatory process is confined to regional lymph nodes, which can enlarge enough to produce bronchial stenosis and distal atelectasis, one cause of the middle-lobe syndrome. If this occurs distal to an evolving cavity, rapid expansion can occur as a result of air trapping to produce a “tension cavity.” The responsible nodes are choked with caseating granulomata. Discrete bronchial stenosis also can be seen without evidence of extrinsic compression, in which case it usually is ascribed to an intense tuberculous bronchitis occurring in a segment draining an active parenchymal infection.

In a typical case of pulmonary tuberculosis, the diagnosis is easily made on the basis of characteristic cavitary changes in an upper lobe on chest x-ray occurring in a patient with a positive purified protein derivative (PPD) test whose sputum has grown *M. tuberculosis*. A suggestive x-ray without positive cultures does not confirm the diagnosis, and a negative PPD test does not rule out active tuberculosis. Definitive diagnosis rests on growth of the organism in culture (Fig. [16-80](#)). A positive acid-fast stain provides a highly suggestive provisional diagnosis but cannot discriminate completely between *Nocardia* and *M. tuberculosis*.

Culture also allows identification of “atypical” mycobacteria, which deserve special mention. Clinically and pathologically, infection with an atypical *Mycobacterium* can be indistinguishable from infection with *M. tuberculosis*. Resistance to multiple antituberculous drugs is common among the atypical mycobacteria, and chances for a nonoperative cure depend on accurate assessment of appropriate medical therapy. Culture becomes important for identification of a specific species and for characterization of its drug sensitivities. Even with successful medical treatment, the use of three or four drugs for 2 to 4 years can be anticipated. Because of drug resistance, a higher proportion of atypical than of typical mycobacterial pulmonary infections require operation. *M. kansasii* and *M. avium-intracellulare* are the species most often associated with pulmonary disease (Table [16-13](#)).

Treatment

Operative treatment of tuberculosis is ordinarily an elective procedure performed after a period of treatment with antituberculous drugs. Emergency operation is only required for life-threatening hemorrhage or massive air leak from a tension pneumothorax. In these cases, the procedure of choice is almost always lobectomy or segmentectomy. The technical precautions necessary to minimize the risk of tracheobronchial or pleural spread of infection are the same as those discussed with regard to resection for lung abscess. Other indications for resection include (1) extensive pulmonary destruction with bronchopleural fistula and empyema, (2) persistently active disease with drug-resistant organisms, (3) inability to rule out coexisting bronchogenic carcinoma, and (4) posttuberculous bronchostenosis with recurrent nontuberculous infection. There has always been some controversy regarding resection of (5) an “open negative” cavity. In these patients with negative sputum, the goal of treatment is to ensure that the cavity itself is made truly negative. There is considerable evidence that well-managed long-term medical therapy can produce results at least as good as those following resection. Partial adherence to a drug regimen can encourage emergence of resistant strains, so noncompliant patients are poor candidates for medical treatment of large residual cavities and should have a resection. This discussion applies equally to typical and atypical Mycobacterial infections.

An occasional patient will have persistent active infection but such limited ventilatory reserve that resection would not be tolerated. Under these circumstances, lung-collapse therapy (e.g., thoracoplasty) is a valid alternative. Thoracoplasty is designed to collapse the affected portion of lung to promote healing, which can be remarkably effective. Collapse is achieved without entering the pleural space by performing an extrapleural resection of the first five ribs, followed in 10 to 14 days by resection of the sixth and seventh ribs. Although this procedure has its origins in the treatment of tuberculosis, the principle of chest wall collapse occasionally is used to obliterate the pleural space after a pulmonary resection as well, especially in managing tuberculosis.

Chronic Granulomatous Infections

By about 1900 all the major fungal pathogens had been isolated and named, but recognition of their role in disease underwent a very characteristic evolution. They were initially thought to be rare and fatal infections, but as diagnostic acumen increased, it became clear that mild asymptomatic infection was far more common. Now the pendulum is beginning to swing back somewhat, as fungal infections find their way into our enlarging reservoir of iatrogenically immunocompromised patients. Features common to all fungal infections include protean manifestations in compromised hosts, sensitivity to amphotericin B, and mimicry of carcinoma and tuberculosis. Amphotericin B is a very important drug that has rendered most fungal infections medically treatable, but it is also a highly toxic drug that cannot be administered as casually as antituberculous drugs are given. Certain bacilli (e.g., *Actinomyces* and *Nocardia*) also produce granulomatous infections.

Actinomycosis

For many years the *Actinomyces* were misclassified as fungi because they form branching hyphae and spores. Only in the past 20 years has it been recognized that the *Actinomyces* are bacteria. This taxonomic distinction has therapeutic relevance because the pathogens in this group are highly sensitive to penicillin and sulfonamides but not to amphotericin B.

Actinomycosis is caused by *A. israelii*, an anaerobic filamentous bacillus that is not found in nature but is a normal commensal inhabitant of the oral cavity and tonsillar crypts. It is not known what causes this organism to become an invasive pathogen, but in about three-fourths of cases some kind of predisposing factor can be identified, such as immunosuppression or breakdown of local tissue barriers (i.e., tooth extraction). About 60 percent of cases are cervicofacial, and only 15 percent are thoracic. Thoracic infection is presumed to result from aspiration of infected secretions. Classically, the disease is characterized by suppuration, abscess and sinus tract formation, and relentless invasion with complete disregard for tissue planes. Multiple sinus tracts are observed today in only one-third of cases, and the lesion is seen more commonly as a parenchymal process mimicking bronchogenic carcinoma. When involvement of ribs or extension into mediastinal structures is seen, without a cancer diagnosis, actinomycosis must be high on the list of possible causes (Fig. [16-81](#)).

Expectorated sputum, material from a sinus tract, and biopsy material can demonstrate sulfur granules, which are yellow-brown clusters of microcolonies. This finding is highly suggestive, but since *Nocardia asteroides*, certain fungi, and *S. aureus* are also capable of producing clumps of material resembling sulfur granules, diagnostic confirmation rests on identification of the bacillus within the granules, for which special stains are required. Cultures are positive in only about one-fourth of patients.

The organism is sensitive to penicillin, although large doses are required to penetrate the dense colonies, and medical treatment is most often quite successful. The surgical strategy is to make an accurate diagnosis at an early stage of disease. Because the disease can have gross resemblance to fungal infections and to carcinoma, and because tissue stains are essential to the diagnosis, operation frequently is required to obtain adequate biopsy material. Successful diagnosis is followed by high-dose intravenous penicillin and a long subsequent course of oral administration. Resection rarely should be necessary except in unusually advanced presentations with an inadequate response to penicillin.

Nocardiosis

Nocardiosis is caused by *N. asteroides*, an aerobic acid-fast filamentous bacillus widely distributed in nature as a saprophyte in soil and domestic animals. It is a rare pathogen except in an immunocompromised host and is most often manifested as a thoracic infection, beginning as a pneumonic process difficult to distinguish grossly from tuberculosis, fungal infections, and carcinoma. It also can closely mimic actinomycosis, with chest wall involvement, sinus tract formation, and production of sulfur granules. The acute infection is often much more aggressive than actinomycosis, with extensive pulmonic necrosis and abscess formation and metastatic dissemination to the central nervous system and elsewhere.

Nocardia is relatively easy to culture and to identify with standard stains, so the diagnosis frequently can be made by brush or needle biopsy and even from expectorated sputum. The organism is sensitive to sulfonamides, which usually provide successful therapy. Other drugs can be added in poorly responsive cases (trimethoprim-sulfamethoxazole, minocycline) with good results, and surgery is purely adjunctive in the majority of patients. Pulmonary resection, drainage of empyema, and similar procedures can be performed safely when necessary.

Fungal Infections Histoplasmosis

Histoplasmosis is the most common systemic fungal infection in the United States. *Histoplasma capsulatum* is a dimorphic fungus common in the great river valleys of the Midwest (e.g., Ohio and Mississippi), where it lives in mycelial form in soil, decaying organic material, and guano. It assumes yeast form in the cytoplasm of pulmonary alveoli after inhalation. It is extremely common in endemic areas as an asymptomatic infection; the severity of disease is determined by the size of the inoculum and the immune competence of the host. Release of a large inoculum can produce outbreaks of acute pneumonic illness in normal hosts and usually occurs after an environmental disruption such as excavation or demolition. Such infections ordinarily resolve without specific treatment but not before widespread lymphatic and hematogenous dissemination has occurred, apparent later as scattered calcific nodules in the lungs, mediastinum, spleen, and liver. In symptomatic patients, the disease can take many forms and often is distinguishable from tuberculosis only by culture. Skin testing reagents are available but are not as reliable as purified protein derivative (PPD). Serologic diagnosis is also available but is no more reliable and can be misleading if obtained following skin testing. As with tuberculosis, definitive diagnosis requires growth of the organism from pathologic specimens.

Amphotericin B is effective treatment in the majority of patients and is always the treatment of choice in a serious illness once the diagnosis is made. Most infections are asymptomatic or moderately symptomatic and self-limited, and chemotherapy is not recommended for skin test conversion as it is in tuberculosis.

Operation is applied much as it is in tuberculosis. Cavitory disease is common, and in a recent large series from an endemic area (Tennessee), this was the most frequent indication for resection. Large, thick-walled cavities that have failed to improve after a course of amphotericin B are likely to progress and can be resected with low morbidity and mortality. Another frequent indication for operative intervention is an inability to establish a definitive diagnosis, especially when the lesion presents as a solitary pulmonary nodule grossly consistent with carcinoma (Fig. [16-82](#)). As in tuberculosis, hemoptysis can require operation, and bronchostenosis produced by extrinsic nodal compression can require resection.

The lymphogenous phase of *Histoplasma* dissemination leads to remarkable nodal enlargement in some patients, producing symptoms related to compression of mediastinal structures and a radiographic appearance resembling mediastinal malignancies. Mediastinal involvement also can produce a sclerosing mediastinitis with obstruction of the superior vena cava, pulmonary arteries or veins, esophagus, or tracheobronchial tree. These may require surgical correction with autogenous venous bypass grafting to relieve symptoms of superior vena cava syndrome. The pathophysiology of this desmoplastic response to infection is not completely understood but is an idiosyncratic reaction.

Coccidiomycosis

Coccidiomycosis is the second most common fungal infection encountered in the United States. *Coccidioides immitis* is a dimorphic fungus found in mycelial form as a saprophyte in the arid soil of the American Southwest. Arthrospores released by the hyphae are inhaled and initiate the parasitic phase by becoming spherules that release infective endospores. In the normal host, most infections are asymptomatic, but some will manifest “valley fever,” essentially a mild pneumonic form of the illness. The organism is not difficult to recover from sputum or pathologic specimens. The skin test and serologic titers are almost always positive in active disease but are more ambiguous in the more chronic and indolent forms of infection. Except for a propensity to form thin-walled cavities, the spectrum of gross and microscopic pathology is similar to that seen in histoplasmosis, tuberculosis, and other fungal infections. In most instances, spontaneous resolution occurs.

For patients with symptomatic illness requiring treatment, amphotericin B is the primary therapy. As with histoplasmosis, the lungs provide an effective barrier against serious systemic illness in most patients, and specific treatment is frequently not required. Aggressive necrotizing pulmonary infection and disseminated disease usually are seen in immunocompromised hosts and require early and aggressive medical treatment.

Indications for operation are virtually identical to those applied to histoplasmosis. Resection of cavitory disease and resection for definitive diagnosis of a solitary pulmonary nodule are performed most frequently. Specific indications in cavitory disease include progressive enlargement, hemoptysis, rupture, and secondary infection.

Blastomycosis

Blastomycosis is caused by *Blastomyces dermatitidis*, a round, single budding yeast endemic in the southeastern United States and other scattered areas. Although there is a common cutaneous form, the disease always is acquired through aspiration of spores into the lungs, where it can assume a variety of appearances—pneumonic infiltrates, cavitation, solitary granulomatous nodules, and disseminated disease. The cutaneous form is characterized by crusty, ulcerative lesions from the margins of which the organism can be cultured readily. Cutaneous and pulmonary infection can occur together, and the cutaneous form has a better prognosis. Diagnosis rests on identification of the organism, which can be done with a sputum Papanicolaou stain.

Although it can mimic tuberculosis and other fungal infections, in endemic areas it most frequently mimics bronchogenic carcinoma, and resection frequently is required if a definitive diagnosis cannot be established (Fig. [16-83](#)). Aggressive infection is treated with amphotericin B. Cutaneous infection and mild pulmonary infection also respond well to 2-hydroxystilbamadine. As with other fungal infections, treatment with amphotericin B often can be avoided in mild presentations of illness, especially in normal hosts during outbreaks in endemic areas.

Aspergillosis

Aspergillus is a filamentous fungus with septate hyphae that is ubiquitous in nature. Inhalation of spores from *A. fumigatus*, *A. niger*, and other species initiates infection in susceptible individuals. Aspergillosis presents in three forms: allergic bronchopulmonary, saprophytic, and invasive. The first is characterized by asthmatic symptoms resulting from host response to fungus in the airways and is of little surgical importance other than diagnosis. The invasive form usually is seen in the immunocompromised host, can involve any organ system, and is almost always fatal. Robinson and associates and others have advocated early resection if disease is limited to one lobe with some encouraging results. Surgical attention focuses mainly on the saprophytic form, produced by colonization of a preexisting pulmonary cavity (an aspergilloma, mycetoma, or “fungus ball”). Hemorrhage can occur, requiring surgical intervention. On chest x-ray the aspergilloma appears as a solid, rounded mass within a cavity, surrounded by a crescent of air between the fungus and the cavity wall (Fig. [16-84](#)). *Aspergillus* precipitins are almost always detectable in patients with aspergilloma. Skin testing is available but is positive in only 30 to 75 percent of patients. The value of sputum cultures has been debated, but recent evidence suggests that two or more positive cultures carry excellent specificity and sensitivity.

For disseminated disease, amphotericin B is the mainstay of therapy. Penetration of the drug into a cavity containing an aspergilloma is very poor, so resection is considered the treatment of choice for a significant aspergilloma. Excellent results have been obtained with repeated intracavitary instillation of amphotericin B using fluoroscopic guidance. Operative treatment most frequently requires lobectomy, segmentectomy, or pneumonectomy. Percutaneous drainage or cavernostomy (open drainage through the chest wall) is occasionally performed in patients with poor ventilatory reserve and can be augmented by intracavitary instillation of antifungal agents. Repeated percutaneous instillations of amphotericin have been successful definitive therapy. Operation is most often justified as prevention for hemoptysis, which occurs in 50 to 83 percent of patients and can be life-threatening in a fraction of that total. Even so, operation remains somewhat controversial because it is associated with considerable mortality and morbidity. This is related to the poor health of most susceptible hosts and the technical difficulty of resection through dense inflammatory tissue. In a recent series from the Mayo Clinic, underlying lung disease or immunologic risk factors were present in 92 percent of patients, and complications occurred in 78 percent of patients with complex aspergillomas. In this series and others, operative mortality has been 5 to 10 percent, and complications in “simple” aspergilloma resection have ranged from 25 to 34 percent. Nonetheless, in the Mayo Clinic series the late results were excellent in about 75 percent of patients. It is probably prudent to observe small, asymptomatic aspergillomas, but in most cases, if intracavitary treatment fails, resection should be performed, accepting the increased risk in favor of potential benefits.

Cryptococcosis

Cryptococcosis is caused by *Cryptococcus neoformans*, a round, budding yeast found in soil and pigeon droppings. Infection occurs through inhalation of the organism and in most individuals produces a comparatively benign bronchopulmonary illness. The chief radiologic finding is a granulomatous complex with hilar node involvement, indistinguishable from the Ghon complex of tuberculosis. It is rarely of surgical significance except in the compromised host, when the entire spectrum of fungal pulmonary pathology seen in more inherently virulent infections can be seen on occasion (see Fig. [16-84](#)). The best known disseminated manifestation is meningitis. Infection can be controlled in many cases with amphotericin B and 5-fluorocytosine, even with meningeal involvement.

Innumerable other fungi can be associated with pulmonary disease in human beings, but as the list diverges further and further from the recognized pathogens, it becomes increasingly confined to immunocompromised hosts. *Candida*, mucormycosis, sporotrichosis, monospirosis, *Torulopsis*, even *Penicillium*—all have been described. Surgical treatment is rarely indicated and seldom definitive.

Lung Tumors

Primary Carcinoma of the Lung

Epidemiology

Cigarette smoking is the predominant factor in the cause of lung cancer. Each year in the United States there are 450,000 excess and preventable smoking-related deaths. Approximately 170,000 of these deaths are a result of lung cancer. Many industrial chemicals (e.g., nickel, chromium, uranium, and asbestos) are co-carcinogens.

From a trivial health problem at the beginning of this century and a minor one by 1930 (a death rate of 5 per 100,000), lung cancer has now become the main cancer killer in both men (70 per 100,000) and women (40 per 100,000). Reflecting the changing smoking habits in women that have occurred since the early 1950s and the long exposure required for the development of this malignancy, the rate in women is rising at an alarming pace (Fig. [16-85 A, B](#)). The histologic changes in the bronchial mucosa occur gradually over many years, evolving from metaplasia through dysplasia to carcinoma. Those who stop smoking reduce, but never eliminate, their risk. It is vital that primary care physicians educate patients and their families about the importance of smoking cessation and a smoke-free environment because even “passive” smoking has been implicated as an etiologic agent.

Pathology

Primary bronchogenic lung cancer is broadly divided into small cell and non- small cell variants. The World Health Organization Classification of primary lung tumors is shown in Table [16-14](#).

Non-small cell lung cancer accounts for approximately three-quarters of all tumors. In many countries, including the United States, adenocarcinoma rather than squamous cell has become the predominant cell type. It is hypothesized that the routine use of cigarette filters leads to deposition of carcinogens more peripherally in the lung, where adenocarcinomas tend to occur. Bronchioloalveolar carcinoma (BAC) is an uncommon form of adenocarcinoma. It develops in the very distal airways and lines the bronchioles and alveoli without invading through the basement membrane. Most frequently, it presents as an indolent solitary peripheral nodule. It also can present as a multifocal carcinoma (synchronous or metachronous), which is appropriately managed by surgical resection. Rarely, it causes diffuse lobar or pneumonic consolidation. In this situation, it is a more aggressive tumor best treated by chemotherapy.

Small cell lung cancer, accounting for approximately 20 percent of all lung cancer cases, arises from the neuroendocrine (Kultchitsky) cells lining the deeper layers of epithelium and has a more aggressive biologic behavior, frequently presenting with distant metastatic disease. It is rare for this variant to be considered for surgical treatment.

With increasing frequency, pathologists are identifying neuroendocrine cell tumors as part of the spectrum of non-small cell lung cancer. These large cell neuroendocrine carcinomas appear to have a prognosis worse than non-small cell lung cancer, approaching that of small cell lung cancer. Many of these were identified previously as “large cell undifferentiated tumors.” With immunohistochemical staining, neuroendocrine features can be identified and worsen the prognosis of this type of non-small cell cancer.

Stages of Lung Cancer

The prognosis of lung cancer is stage-dependent, and is influenced by the size and location of the tumor (T status), the presence or absence of lymph node metastases (N status), and the presence or absence of distant metastases (M status). For this reason, the initial clinical staging of the tumor is important not only for prognosis but also for decisions about treatment.

The most recent AJCC/UICC TNM Staging Classification is indicated in Table [16-15](#). As can be seen, with increasing stage, the 5-year survival rate after surgical resection

decreases.

Primary lung tumors arise for the most part in proximal or distal airways, resulting in the term *bronchogenic* carcinoma. It has been found that patients with tumors less than 3 cm (T1) fare better than those with tumors greater than 3 cm (T2). Tumors invading adjacent non-vital structures (T3) fare worse, but worst of all are tumors invading adjacent vital structures (T4), such as the mediastinal organs (heart, great vessels, etc.).

When tumors spread by lymphatic extension to intrapulmonary and hilar lymph nodes (N1), the prognosis worsens. Spread to the mediastinum is an even poorer prognostic sign and often contraindicates a surgical approach. Patients with ipsilateral mediastinal lymph node spread (N2), however, fare better than those with tumors with contralateral or supraclavicular lymph node spread (N3).

In the TNM staging, M0 indicates no evidence of metastases, whereas M1 indicates evidence of metastases whether they be single or multiple. When considering surgical treatment, on occasion, patients who have totally resectable primary tumors and solitary sites of metastases may be considered for a surgical approach to both lesions. Patients with metastatic disease usually are treated by systemic therapy (chemotherapy), with palliative radiotherapy used to control the primary locoregional tumor. Patients with metastatic disease, on the whole, fare poorly, with less than 50 percent surviving 1 year.

Surgical therapy is directed at those tumors for which a complete excision can be accomplished. These generally include patients who clinically have been staged I, II, or IIIa. For all the preceding reasons, it is important in assessing patients with lung cancer to attempt as accurate a clinical staging as possible before deciding definitive treatment.

Clinical Manifestations

Bronchogenic carcinoma is seen with increasing frequency in patients over age 50, with a peak incidence between the ages of 60 to 70. At one time, men predominated (10:1), but with increasing smoking among women, the incidence of carcinoma is now only 1.5:1 in favor of men.

Patients may be asymptomatic at diagnosis with a lung mass detected on a routine chest x-ray. These patients are the most likely to be cured because they have early-stage lung cancer. However, attempts at routine screening of smokers by annual chest x-rays have failed to improve the survival of the mass screened population because chest x-ray does not detect tumors at an earlier, more curable stage of the disease.

Clinical manifestations of lung cancer can be broadly divided into symptoms and signs related to local growth of the tumor within the airway and lung (chronic cough, hemoptysis, symptoms of obstructive pneumonia) and symptoms related to local extension of disease to surrounding structures (hoarseness from recurrent laryngeal nerve palsy, chest pain, symptoms of superior vena caval obstruction and dyspnea from pleural effusion).

Other symptoms include those resulting from metastatic spread, which occurs most frequently to the brain and bones or generalized manifestations of metastatic malignancy, including general debility, weight loss, and anorexia. Because carcinoma of the lung can metastasize to any organ, the symptoms and signs of metastatic disease vary.

A small percentage of patients with lung cancer present with extrapulmonary non-metastatic manifestations that are considered to be a consequence of the elaboration of hormone-like substances by the neoplastic cells. The occurrence of these signs and symptoms does not necessarily imply systemic spread of the tumor, and often resection of the primary tumor is associated with regression of the symptoms. The most common of these is the syndrome of clubbing and hypertrophic osteoarthropathy (Fig. [16-86](#)). Extrapulmonary nonmetastatic syndromes are listed in Table [16-16](#). Frequently, these are associated with small cell lung cancer, and it is felt that the neurosecretory type granules seen in these tumors have an important role in producing these non-metastatic manifestations. Examples of this are Cushing's syndrome and inappropriate antidiuretic hormone (ADH) secretion. Hypercalcemia, caused by a parathormone-like polypeptide, is seen most frequently in squamous cell carcinoma and has an unusually poor prognosis.

Diagnosis and Staging

When first seen, approximately two-thirds of patients have disease beyond that treatable by surgical excision. For this reason, diagnostic evaluation must include an effort to determine whether the disease is localized (stages I and II), locally advanced (stages IIIa and IIb), or metastatic (stage IV). The key to this effort is a meticulous history and physical examination to elicit symptoms and signs of local extension and distant spread. Frequently, symptoms alone suggest a diagnosis. The minimum imaging studies required for diagnosis and staging include a chest x-ray and a CT scan of the chest and upper abdomen—the latter to detect occult metastases in the liver and adrenals. In most instances, a diagnosis should be confirmed by non-invasive or invasive testing already outlined. This can include sputum cytology, needle aspiration biopsies, bronchoscopy, mediastinoscopy or invasive exploration of the hemithorax by video- assisted thoracoscopy or thoracotomy. Once a diagnosis has been established, a metastatic work-up is valuable. In addition to an accurate history and physical examination, imaging studies designed to detect that area of metastatic disease (e.g., CT or MRI of brain and bone scan) should be performed.

In order to determine whether or not a patient benefits from surgical removal of their tumor, accurate clinical staging is important. This often is accomplished by physical examination, chest x-ray, CT scan, and bronchoscopic examination. There is considerable debate as to how intensive the search for involved mediastinal lymph nodes (a very poor prognostic sign) should be. Any mediastinal lymph nodes that are enlarged on CT scan (>1 cm) are potentially malignant and should be biopsied by mediastinoscopy (Fig. [16-87](#)). However, because approximately 10 percent of normal-sized nodes contain microscopic metastatic disease and despite "normal" nodes on CT, some surgeons perform mediastinoscopy in all patients except those with very small peripheral tumors (clinical T1 N0). Because of the poor prognosis of clinical N2 disease, primary surgical resection is rarely indicated thus the intensive search for this preoperatively.

Local factors suggesting total inoperability include a malignant pleural effusion (T4) and contralateral or supraclavicular mediastinal lymph node involvement. When patients present with these suspected findings, histologic confirmation by invasive testing (e.g., mediastinoscopy, scalene lymph node biopsy, thoracentesis, or thoracoscopy) should be performed.

Treatment

Patients with clinical stages I and II non-small cell lung cancer should be considered for surgical treatment. Before surgical resection, patients must be assessed for their ability to withstand the proposed resection. A detailed preoperative assessment including history and physical examination, pulmonary function studies, and cardiac evaluation as indicated must be carried out. Lobectomy or pneumonectomy generally requires resection of the tumor. If local extension has occurred, en bloc resections of chest wall or adjacent airways may be required for complete excision.

Patients with small cell lung cancer rarely may be candidates for surgical resection. This applies only to those patients with clinical stage I or II disease (a rare event, <10 percent) in those patients presenting with small cell lung cancer (Fig. [16-88](#)).

It is worthwhile to discuss the management of non-small cell lung cancer according to the various stages of disease. The aim of surgical resection is to completely remove the primary tumor and all involved lymph nodes (Fig. [16-89](#)). There has been debate by surgeons on whether mediastinal lymph nodes simply should be sampled or should be completely removed at every surgical operation. When involved, surgical removal is indicated.

Stage I Disease

This early stage of lung cancer is afforded the best opportunity of long- term cure. In stage Ia disease, less than 20 percent of patients recur after surgical resection (Table [16-17](#)). Most frequently, a lobectomy is required for treatment. Less commonly, more proximal tumors require a pneumonectomy. Lesser resections (segmental resection, wedge resection) are reserved for those patients with extremely poor pulmonary function who cannot tolerate a lobectomy. It has been shown that these lesser resections carry a higher rate of local recurrence and decreased long-term survival, probably because of failure to remove occult intrapulmonary lymphatic spread.

Stage II Disease

Once the tumor has spread to intrapulmonary and hilar lymph nodes, the prognosis after surgical resection is worse. In these patients, pneumonectomy is more frequently required to extirpate all tumor. Since most tumors, when they recur, do so at distant sites, one would hope that adjuvant therapy, either chemotherapy or chemoradiotherapy, would have a beneficial effect. To date, several randomized trials have failed to show that adjuvant chemotherapy or radiotherapy improves survival.

When tumors directly involve adjacent structures (T3), en bloc excisions can be performed, removing the lung and the adjacent involved structure. In these cases, especially with apical (superior sulcus) tumors that involve the upper ribs and lower brachial plexus (Pancoast tumors: chest pain, T1 nerve root pain, and Horner's syndrome), the role of perioperative radiotherapy appears to be beneficial. Other T3 tumors, especially those involving proximal airways, can be treated with various plastic reconstructive techniques (sleeve resection of bronchi) such that ipsilateral lobes can be preserved.

Stage IIIa Disease

Management of tumors that have spread to ipsilateral mediastinal lymph nodes remains controversial. Historically, when surgeons identified preoperatively the presence of involvement of mediastinal lymph nodes, surgical therapy was not offered to the patient. These patients were treated primarily with radiotherapy and had a 5 to 15 percent chance of being cured using this treatment. During the past 15 years, multiple chemical trials have been performed to determine if induction chemotherapy or chemoradiotherapy improves resectability and survival. Results of these trials indicate that complete surgical resection is frequently possible, with up to 30 percent of patients being cured of their tumor. Similarly, a combination of chemotherapy and primary radiotherapy has been used in treating such tumors without the use of surgery. Patients treated in this manner also are provided a better chance of cure than with radiotherapy alone. Since both treatment strategies are effective in about 20 to 30 percent of patients, there are two randomized trials now being conducted, one in North America and one in Europe, in an attempt to determine which is the best approach in managing such patients. Both approaches presently appear to be valid.

Stage IIb Disease

Once lymph node involvement includes contralateral mediastinal areas (N3), it is highly unlikely that patients will be cured of their tumor by a surgical approach. In such instances, patients usually are offered chemotherapy and radiation to control their disease with a small chance of complete cure. Occasionally, T4 tumors that minimally invade adjacent vital organs of the mediastinum can be removed together with portions of that organ (e.g., left atrium, superior vena cava, aorta, esophagus). The results of such surgery can, on occasion, be gratifying when only minimal invasion of such structures has occurred and a complete resection is possible. Although not proven, it is routine that combined-modality therapy (e.g., radiotherapy and chemotherapy) is added in an attempt to improve the results of such treatment.

Metastatic Disease (Stage IV)

On occasion, patients with totally curable primary tumors present with a solitary metastasis (e.g., brain). In such instances, it is possible to effect a cure in 10 to 15 percent of patients by removing the primary tumor and the solitary site of metastasis.

When considering a surgical approach for a patient suffering from lung cancer, the surgeon must be aware of the results of such surgery—morbidity and mortality as well as long-term survival (Table 16- 18). The cost-benefit ratio of the procedure should be weighed before offering the treatment to the patient. The more advanced diseases with the more extensive resections result in increasing morbidity and perioperative mortality with a lesser chance of ultimate cure. Surgeons must be very selective in determining whether or not an individual patient should be offered a surgical procedure for treatment and must base that decision on the preoperative assessment of the patient and the clinical staging of the tumor.

Solitary Pulmonary Nodule

Among patients with lung cancer, the best survival can be expected from that group of patients in whom the cancer is first found as an asymptomatic solitary peripheral pulmonary nodule on the chest x-ray. Because of this relatively favorable outlook if the nodule is a cancer, and because of the variety of nonmalignant lesions that also present in this fashion, there is considerable interest in the differential diagnosis of these “coin lesions.” The *solitary pulmonary nodule (SPN)* has been defined by general agreement to be an abnormal density up to 4 cm in diameter, rounded or ovoid in appearance (Fig. 16-90 and Fig. 16-91), surrounded by a zone of lung tissue by x-ray, and free of cavitation or associated lung infiltrates. Eccentric flecks of calcium may be present, but lesions that are largely calcified or that have concentric calcium rings are not considered.

Earlier reports suggested a malignancy rate of about 40 percent in the SPNs; in more recent analyses, approximately 80 percent of all coin lesions were malignant in patients over 50 years of age. Only when the nodule is known to have been present for a long period with absence of growth and with a pattern of calcification characteristic of the several benign lesions that can occur (e.g., granuloma, hamartoma) should histologic diagnosis of these nodules be delayed.

Proof that the lesion has not developed recently requires inspection of previous chest x-rays because faint, small lesions are frequently overlooked in screening films and a “negative” official report often has been in error. Even when previous x-rays document radiologic stability of a solitary nodule for up to 2 years, malignancy must be suspected.

The differential diagnosis of an SPN includes many entities, among which are pulmonary hamartoma, granuloma, pulmonary arteriovenous fistula, pulmonary infarction, and several benign and malignant tumors. Surgery is rarely indicated for the nonmalignant lesions under consideration, but many clinics have recommended extensive diagnostic efforts to ascertain the correct diagnosis before thoracotomy. There is a significant false-negative rate in most of the studies short of excisional biopsy. With improving imaging techniques, the odds of a correct preoperative diagnosis are improving, but with the consequences of error so costly and with the reasonably good results after operative resection, many physicians feel that it is unwise to gamble on a determination of benignity unless the lesion is in a young (under age 35), nonsmoking patient with a known history of radiographic stability and advise surgical excision for diagnosis in all other patients.

Where calcification is present, CT is useful in selecting the rare patient who can be observed safely. CT scans have the capacity to measure absorption coefficients, which indicate tissue density. Siegelman and associates used CT with thin sections to assess tissue density in 91 apparently noncalcified pulmonary nodules in 88 patients. They established a separation between benign and malignant lesions on the basis of high attenuation values in the benign *SPNs*. The high values were presumably due to diffusely distributed calcium deposits within the lesions not visible on standard radiographs. In the absence of calcium, benign SPNs may have attenuation values in the same range as malignant lesions. Similarly, in a small proportion of patients, malignant lesions may have sufficient calcium to result in the high values characteristic of benign nodules.

For many years there has been a difference of opinion regarding the management of patients with SPNs, with some groups advocating early thoracotomy with resection of the lesion for all patients above 35 years of age and others urging a more conservative approach with greater emphasis on diagnostic studies and observation. Although the resection policy did result in a 50 percent frequency of removal of benign lesions, excellent cure rates could be expected if primary malignancies were found. With the sharply falling frequency of tuberculous granulomata and the continuing rise in lung cancer rates, current odds favor early resection unless there is strong evidence of a benign process. Sputum cytology, bronchoscopic washings and brushings, and percutaneous needle biopsies each can provide clear-cut positive information of both malignant and benign disorders. In many instances no diagnosis is made, and in these patients, malignancy cannot be excluded.

If previous films are available for comparison, or if a course of observation is elected for other reasons, calculation of the time necessary for doubling of the tumor volume is a useful indicator of the nature of the lesion. Serial radiographs provide the data for calculation of growth rate, and it usually is possible to detect that a lesion is growing within a few weeks. The doubling time of malignant nodules is usually between 37 and 465 days. If the lesion is growing more slowly or more rapidly than this, the evidence is in favor of benignity. Advocates of a conservative approach insist that present knowledge supports the concept that a pulmonary nodule can be watched safely for a period of time to determine whether it is growing.

A logical approach to the *SPN* is outlined in (Fig. 16-92). With the advent of video-assisted thoracoscopic techniques, resection of nodules in the periphery of the lung result in minimal morbidity and virtually no mortality. In patients known to be at high risk, including those with age over 50, smoking history, absence of certain knowledge, similar lesions on chest x- ray for more than 2 years, and failure to make a diagnosis by other means, resection of the nodule is indicated for diagnosis. Conversely, if the patient is under age 35 and a nonsmoker and the chance of malignancy is small, watchful waiting in close observation for growth or change may be indicated. This is especially true if the nodule is not easily accessible by video-assisted techniques.

Other Lung Tumors

Carcinoid Tumors

This group of tumors includes all neuroendocrine neoplasms arising from Kulchitsky cells. At the benign end of the spectrum, the “typical” carcinoid histologically resembles the carcinoid tumors of the small intestine. Along with cylindroma and mucoepidermoid tumors, these previously were termed *bronchial adenomas*. All of these tumors are in reality malignant, and the term *adenoma* has been abandoned. Even with “typical” carcinoid tumors, metastases to regional lymph nodes do occur on occasion. A more ominous variant is the “atypical” carcinoid, which demonstrates more mitotic figures and a higher potential for metastases and behaves much more like lung cancer. These tumors have a much poorer prognosis than typical carcinoids, which are almost always cured by surgical resection.

The term *bronchopulmonary neuroendocrine tumor* has been used to include the spectrum of tumors with neuroendocrine features from typical carcinoid tumor to the highly malignant small cell lung cancer.

Over 80 percent of carcinoids arise in proximal bronchi, but peripheral origin beyond cartilage-containing bronchi does occur. The tumors grow slowly and protrude into the bronchial lumen, causing signs and symptoms of bronchial obstruction as the principal clinical presentation. Unusual vascularity may cause hemoptysis as a presenting complaint (see Fig. 16- 91). The vascularity gives the tumor a deep pink or “cherry” red color when visualized through a bronchoscope.

The extent of bronchial wall involvement is variable, but there is usually invasion of the underlying cartilages. Rarely, direct extension through the bronchial wall results in invasion of mediastinal structures. In typical carcinoids, regional lymph node deposits are found in approximately 10 percent of patients. Distant metastases are rare. In keeping with the neuroendocrine origin of these tumors, a few patients with bronchial carcinoid have Cushing-like syndromes that seem attributable to the tumor.

Although the average age of patients with a carcinoid tumor is approximately 40 years, the neoplasm does occur in children. The clinical presentation usually is a result of bronchial obstruction with infection and pulmonary atelectasis. Sputum cytology is negative, but more than 80 percent of the lesions can be visualized by bronchoscopy. The carcinoid syndrome is seen rarely and in very large tumors can occur without extrathoracic metastases.

The treatment for bronchial carcinoid tumor is surgical resection. Neither the primary neoplasm nor lymph node metastases are sensitive to radiation therapy. Lobectomy is the accepted operation in most patients. In view of its low potential for spread, lung-conserving procedures, such as sleeve resection or local bronchial excision with bronchoplasty, are used whenever feasible (Fig. 16-93). The expected long-term survival rate is over 90 percent in typical carcinoid tumors after complete resection.

Tumors of Bronchial Gland Origin

Adenocystic carcinoma (cylindroma) and mucoepidermoid tumors are the most common neoplasms arising from the bronchial glands. Their location is predominantly central, and they are said to take origin only from bronchi containing cartilage and glands. Both neoplasms may show a spectrum of behavior from benign to malignant with regional and distant metastases. The treatment is surgical resection, including en bloc removal of regional lymph nodes when possible. Although the long-term cure rate usually is higher than that of primary carcinoma of the lung, it does not equal the results in typical bronchial carcinoid or benign adenomas. Other rare tumors of bronchial gland origin are occasionally reported; the majority seem to be forms of adenocarcinoma.

Carcinosarcoma

Making up less than 1 percent of lung cancers, carcinosarcoma consists of epithelial and mesenchymal types of tissue, and electron microscopy has confirmed that the sarcomatous elements are not simply transformed components of epithelial origin. The term *blastoma* recently has been used for some tumors that show histologic evidence of association with embryonal tissue.

Carcinosarcomas may be located in the lung periphery or in proximal bronchi, and they have been reported in a wide age range, including children. They are found most commonly in patients with tobacco addiction in the sixth decade. The overall survival of resected patients is worse than primary non-small cell lung cancer, with only about 20 percent of patients surviving 5 years.

Other Tumors

A variety of mesenchymal sarcomas and tumors of reticuloendothelial origin may occur in the lungs. As a group, these tumors represent approximately 1 percent of all primary neoplasms removed at operation. The age range of presentation is considerably wider than that for bronchogenic carcinoma, and the tumors may arise anywhere in the lung or bronchial tree. Precise histologic identification of these neoplasms can be difficult, and they may be mistaken for highly undifferentiated carcinomas or metastatic neoplasms.

The symptoms are the same as those of primary carcinomas, but there is no association with cigarette smoking. When the tumors develop as intrabronchial polypoid neoplasms, the symptoms of bronchial obstruction lead to earlier diagnosis and, therefore, a relatively higher cure rate after resection. Leiomyosarcomas, for example, had a 5-year cure rate of approximately 40percent in McNamara's report.

Lymphomas rarely may develop in the lung without evidence of tumor elsewhere. Routine chest x-rays discover an asymptomatic pulmonary lesion in a number of patients; other lesions become symptomatic because of pressure of the growing tumor or lymph nodes on adjacent structures. Although there is no pathognomic radiographic appearance, primary

lymphomas may present as an ill-defined pulmonary mass with central consolidation and air bronchograms. The diagnosis rarely is suspected from sputum cytology or bronchoscopy. Percutaneous needle biopsy may provide the diagnosis, but often this is made after resection. A sufficient number of lymphomas are localized to make the prognosis good after pulmonary resection. Often adjuvant chemotherapy is advised. A 5-year survival exceeding 50 percent may be anticipated.

Hodgkin's disease frequently involves the lung, and a rare patient is seen in whom a solitary pulmonary lesion is unassociated with other evidence of tumor. If resection has been performed, the patient should have complete staging of the disease so that decisions regarding additional therapy can be made.

Fibrosarcoma, rhabdomyosarcoma, neurofibrosarcoma, and other tumors of mesodermal origin may occur rarely in the lung but without specific clinical presentation. The treatment is operative resection, and the prognosis depends on the stage at which the neoplasm was discovered. High-grade sarcomas fare much worse than low-grade ones.

Benign Tumors

Primary or metastatic cancers make up 99 percent of all pulmonary tumors resected, and benign tumors are therefore a relatively small fraction of all lung tumors. Among that group of rare tumors, the hamartoma (chondroadenoma) is the most common. Although occasionally seen in children, they usually appear in men in their fifth to sixth decades, produce no symptoms, and are found in the periphery of dependent portions of the lungs (Fig. 16-94). The characteristic marble-like feel of these cartilaginous tumors makes it usually possible to simply enucleate them; wedge resection may be preferred if the physical findings leave the surgeon uncertain about the diagnosis. Epithelial elements are generally present, and there may be fat, muscular, or fibrous tissue interspersed.

Because benign tumors theoretically can occur wherever the cells from which they might arise are present, a wide variety of other exceedingly rare tumors occasionally is reported. They can be epithelial (tumorlet and papilloma), mesenchymal (fibroma, leiomyoma, lipoma, hemangioma, lymphangioma, neuroma, and rhabdomyoma), or lymphoid (plasmacytoma, lymphocytoma, plasma cell granuloma). Neurofibromas may occur, particularly in patients with neurofibromatosis. The significance of these tumors is almost exclusively related to the differential diagnosis from malignancies. When these neoplasms develop in a major bronchus, they may obstruct and present with the effects of chronic infection.

Metastatic Tumors

The lung is a common site of metastases from solid tumors arising in extrathoracic sites. Cells are shed into the vascular system and are trapped in the first capillary bed encountered; for primary sites not in contact with the venous portal system, this is the lung. Surgical manipulation of primary tumors might contribute to the shedding of cells, but metastases can appear at any time, even long after the primary seems controlled. Five-year survival rates after resection of one or more pulmonary metastases are 25 to 45 percent for patients with several types of carcinoma or sarcoma.

For the past 30 years, the treatment of solitary and even multiple but resectable pulmonary metastases in patients with controlled primaries has been operative resection. Randomized, prospective clinical trials comparing operative resection with medical management have not been done and might be considered unethical. Because of the relatively good results of resection and the absence of evidence that other therapeutic modalities are curative, surgical removal of isolated pulmonary metastases is the accepted treatment for many tumors.

There is some controversy about which tumor characteristics correlate with long-term survival. While it seems logical that long tumor doubling times (>40 days) should portend a better outcome, this has not been a consistent finding in reported surgical series, and shorter doubling times do not contraindicate surgery if other criteria are met. Five factors generally are considered important in selecting patients for pulmonary metastasectomy: The primary tumor has been controlled or is controllable; there is an absence of extrathoracic metastases; the lung metastases can be removed completely; the patient can tolerate the required pulmonary resection; and no better treatment (i.e., chemotherapy) for the pulmonary metastases exists. A recent report from an International Registry of Lung Metastases also emphasizes that the disease-free interval (greater or less than 36 months since control of the primary tumor) and single versus multiple metastases are important prognostic factors. The cell type of the original tumor is linked to prognosis (Fig. 16-95) when one considers that the four groups can be identified with different prognoses (Fig. 16-96).

A problem to be considered when a cancer patient presents with a solitary pulmonary nodule, synchronous or metachronous, is whether this nodule is a metastasis, a primary pulmonary tumor, or a non-neoplastic lesion. When lesions are picked up on routine chest radiographs after previous resection of a primary cancer, almost 90 percent are malignant. CT scans, which can locate lesions 2 mm in size, have a much lower specificity for tumor; only 45 percent of subcentimeter lesions picked up on CT scan are neoplastic. If a solitary nodule appears, 50 percent of patients with previous breast or colon cancers will have a primary lung cancer and not a metastasis. Conversely, patients with melanoma or sarcoma have an 80 percent risk of the lesion being metastatic. Because the type of surgical resection required depends on whether the lesion is primary or metastatic, this distinction is important but may not be available until the time of surgery.

Metastases usually can be removed adequately by wedge or segmental resection. Lobectomies or pneumonectomies are rarely required. A more extensive resection should be considered if the lesion is near the hilum and cannot be removed safely by a lesser operation. It is important to conserve as much lung tissue as possible because a certain percentage of these patients will undergo further resections in the future.

Some surgeons advocate a bilateral simultaneous approach for pulmonary metastatic disease in order to identify occult lesions on the contralateral side. Most surgeons use the preoperative CT scan to decide whether unilateral or bilateral exploration is indicated. Median sternotomy or bilateral anterior thoracotomy allows palpation of both lungs and provides adequate exposure for most resections. Proponents of bilateral approaches argue that in a certain percentage of patients, more metastatic disease will be found on the contralateral side and can be resected at the time of surgery. This is especially true in the case of metastatic sarcomas in which more than one lesion is found unilaterally. At the time of any thoracotomy or metastatic disease, systematic palpation of the collapsed lung is important; in 20 percent of patients, more lesions will be found than identified on CT scan.

With the increased effectiveness of chemotherapy for some malignancies, the role of surgery for pulmonary metastases is changing. With nonseminomatous germ cell tumors, chemotherapy has been so successful that the surgical resection of metastases is used primarily to establish the diagnosis, to resect metastatic disease unresponsive to chemotherapy, and to determine whether lesions that do not disappear totally after chemotherapy are malignant. Similarly, osteogenic sarcomas and Wilms' tumors usually are pretreated with chemotherapy, surgery being used to resect residual disease. Other tumors such as colon and renal cell cancers have yet to be treated preoperatively with chemotherapy in an effective manner.

Although resection is the standard treatment in a group of highly selected patients with isolated pulmonary metastases, this may change in the future as chemotherapy for many of these tumors improves.

TRACHEA

Anatomy

The trachea is a centrally located unpaired organ that conducts air from the oropharynx to the lung. It follows an oblique course from a vulnerable superficial position in the neck deep into the well-protected middle of the mediastinum. The adult trachea has an average length of 11 cm (range 10 to 13cm) segmented by 18 to 22 semicircular cartilaginous rings and has elliptical internal dimension averaging 2.3 cm in lateral diameter and 1.8 cm in anteroposterior diameter. The cricoid cartilage of the larynx, which has membranes attachments to the first tracheal ring, is the only complete cartilaginous ring in the upper airway. The membranous trachea is a flexible sheet of fibroelastic tissue forming the posterior wall of the trachea between the ends of the rings and directly abuts the posteriorly placed esophagus. The rigid rings of the anterior two-thirds of the trachea, combined with the flexible posterior third, impart great flexibility without collapse over a broad range of flexion, extension, and torsion of the neck and maintain patency of the lumen through the extremes of coughing and forced respiration. The loss of cartilaginous support that occurs in tracheomalacia allows dynamic collapse resulting in airway obstruction during the expiratory phase of breathing.

The important anterior relationships of the trachea are the thyroid isthmus, lying across the second to third rings; the innominate artery, crossing obliquely several more rings distally; and the aortic arch, crossing just above the carina. Laterally, the recurrent nerves lie close to the trachea in the tracheoesophageal groove, with the left nerve following a longer course and joining the trachea just above the carina after passing around the aorta and ligamentum arteriosum. Because its blood supply enters posterolaterally, dissection along the trachea is safest when confined to the anterolateral planes. The major arterial supply for the cervical trachea comes from the inferior thyroid artery. Lower portions of the trachea are supplied by branches of the bronchial arteries. Small branches from other mediastinal arteries supply the proximal bronchi and can assume importance after tracheal division, such as, for example, the coronary arterial branches that have been shown to provide blood supply to the tracheal anastomosis after heart-lung transplantation.

Congenital Lesions

Tracheoesophageal Fistulas

The most common congenital lesion involving the trachea is a tracheoesophageal fistula, which presents with many variations. It is discussed in detail in Chapter 37.

Tracheal Stenoses

Congenital tracheal stenosis presents as several variants, all of which are uncommon. Simple mucosal weblike diaphragms can occur, usually at the subcricoid level. Segments of functional stenosis as a result of tracheomalacia (cartilaginous softening) can be seen at sites of compression by a congenital vascular ring or an anomalous left pulmonary artery (pulmonary artery sling). Often associated with these anomalies is another variant characterized by absence of the membranous trachea with fusion of the cartilaginous rings posteriorly over a variable distance, thereby narrowing the trachea and presenting in three principal forms: (1) segmental stenosis, (2) funnel stenosis, in which the distal trachea tapers to a tight stenosis just above the carina, and (3) diffuse hypoplasia of the entire trachea. Frequently, a trifurcated main carina occurs in association with this; the right upper lobe bronchus arises at this level together with the two main bronchi.

Diagnosis

Congenital stenosis should be suspected in any infant with noisy breathing, wheezing, and inspiratory retractions occurring shortly after birth. In less severe forms symptoms may not occur until later in development. The necessary diagnostic evaluation can be exhaustive, reflecting the broad differential diagnosis and the association with other anomalies. Radiographic studies include chest films with magnification focused on the trachea, in inspiration and expiration, barium swallow, xeroradiography, CT scan of the neck and mediastinum, and angiography to identify associated vascular anomalies. Inspiratory and expiratory flow-volume curves, echocardiography, bronchoscopy, and rarely, bronchography also can be helpful. Great care is taken during bronchoscopy to prevent mucosal irritation and edema from converting partial to total obstruction.

Treatment

Therapy is individualized to suit the anatomy and the age of the child. Simple congenital webs often can be removed bronchoscopically. Occasionally, temporizing measures such as tracheostomy or repeated dilation allow resolution with growth. Operative treatment is indicated if these measures fail to allow such resolution, but every attempt is made to postpone reconstruction during early infancy. When possible, the stenotic segment is resected and the trachea reconstructed with an end-to-end anastomosis. Diffuse involvement presents a greater technical challenge in which, until recently, successful results have been rare. Satisfactory reconstructions can be performed using splints constructed from rib or costal cartilage to patch the length of the stenotic segment (Fig. 16-97) reinforced with a vascularized omentum pedicled flap. Innovative plastic procedure (slide tracheoplasty) have been used with success (Fig. 16-98). Perioperative airway management and the maintenance of luminal patency during healing and remodeling pose major challenges.

Relief of tracheal stenosis related to vascular anomalies requires more than simple correction of the vascular anomaly in about half of patients, the defect in the trachea also requiring correction.

Trauma

Blunt and penetrating trauma produce a spectrum of tracheal injuries ranging from simple laceration or contusion to complete transection. Much more common are injuries to the larynx. Hemoptysis, stridor, wheezing, or the presence of mediastinal and subcutaneous air (surgical emphysema) after trauma requires that the possibility of tracheal or proximal brachial injury be evaluated by careful bronchoscopy or exploration. Often a primary reconstruction is indicated, but the more conservative approach of inserting a tracheostomy tube at the site of injury may suffice if the injury is anterior and in the neck.

Post-intubation Injuries

The most common tracheal injury requiring treatment is that occurring as a complication of tracheal intubation for mechanical ventilation (Fig. 16-99). In the past, rigid rubber tracheostomy tubes and stiff tracheostomy cuffs frequently caused tracheal stenosis. Modern endotracheal and tracheostomy tubes have soft, low-pressure cuffs that have largely eliminated this problem. At the site of a tracheal stoma, exuberant granulations can form a bulky obstruction, or loss of cartilage combined with cicatricial healing may form an anterolateral stricture. Ischemic necrosis at the site of the tube cuff or the tube tip can produce a segment of ischemic stricture, a segment of tracheomalacia with functional obstruction during exhalation, or erosion and fistula formation with the esophagus or the innominate artery. Patients often present with airway symptoms long after the tracheal intubation (months or years).

Areas of stricture should be defined carefully with radiographic studies that include magnified air-contrast examination of the trachea, xeroradiography, and helical CT scans of the cervical region and upper mediastinum. Particular attention is paid to definition of laryngeal function. Bronchoscopy is essential to define associated laryngeal injury and preoperative assessment of the stricture.

Treatment

The operative approach for reconstruction of tracheal strictures is similar to that described below for the resection of tracheal neoplasms (Fig. 16-100). Endoscopic laser ablation can be useful for removal of granulation tissue but rarely has a place in the definitive treatment of stricture. A dreaded complication of tracheostomy is the occurrence of a tracheoinnominate artery fistula resulting from erosion by the tracheostomy tube through the anterior wall of the trachea creating a fistula with the overlying innominate artery. This produces sudden and often exsanguinating hemorrhage. A sudden hemorrhage through a tracheostomy tube should always make one think of this complication. Emergency surgery is required to divide and ligate the damaged innominate artery. Despite this division, postoperative neurologic sequelae are rare. Well-vascularized adjacent tissue (e.g., strap muscles) should be used to protect the ends of the divided innominate artery from the adjacent tracheostomy.

A tracheoesophageal fistula occurring during intubation and ventilation usually occurs because of pressure erosion of the cuff posteriorly through the membranous trachea creating a fistula with the subjacent esophagus. Whenever possible, repair of this problem should be delayed until the patient no longer requires mechanical ventilation. Definitive repair requires closure of the esophageal fistula and resection of the damaged trachea with reanastomosis of healthy tracheal ends and interposition of a pedicle of vascularized tissue between the injured sites.

Neoplasms

Primary tracheal neoplasms are uncommon and are vastly outnumbered by tumors that involve the trachea by direct extension from a bronchial, laryngeal, esophageal, or thyroid primary tumor. More than 80 percent of primary tracheal neoplasms are malignant, with squamous cell carcinoma and adenoid cystic carcinoma accounting for the vast majority. Adenoid cystic carcinomas usually are slow-growing, even when metastatic, so resection, if indicated, combined with radiotherapy can offer excellent long-term results even if considered non-curative. Squamous papillomas and fibromas are the most common benign tumors. A number of rare benign and malignant neoplasms have been identified in the trachea, including carcinoid tumors, chondromas, adenocarcinoma, mucoepidermoid carcinoma, and many others.

Diagnosis

Symptomatic patients with tracheal tumors present with some combination of dyspnea, cough, wheezing, inspiratory stridor, hemoptysis, and less frequently, recurrent respiratory infections. Chest x-rays can be misleading because the lung fields are clear, but an alert physician will obtain a CT scan of the cervical and upper mediastinal region to make the diagnosis. Standard pulmonary function testing may be normal, but flow-volume loops can detect upper airway obstruction. Bronchoscopy is an essential part of the evaluation and is best approached cautiously after maximum information has been obtained from noninvasive methods because of the potential for precipitation of acute airway obstruction. When approached carefully, most tumors can be biopsied endoscopically and a tissue diagnosis confirmed before proceeding with further treatment.

Treatment

Resection, airway stenting, and endoscopic ablation are the surgical options in the treatment of tracheal neoplasms.

Palliative Maneuvers

Airway stenting is palliative in patients with inoperable disease. Endoscopic ablation has been gaining more favor because of advances in laser technology. Developed in 1960, lasers produce coherent, low-divergence, high-intensity light capable of destroying tissue. As early as 1964, lasers were used experimentally to destroy tumor cells by thermal energy. The CO₂ laser was the first laser to be used for resection of airway neoplasms. This laser is very effective for tissue cutting and vaporization, but use in the trachea requires a rigid bronchoscope because the beam cannot be passed through fiberoptic systems. The neodymium-yttrium aluminum garnet (Nd-YAG) laser is favored because it is as effective as the CO₂ laser, can be used through fiberoptic systems, and allows cauterization of vascular tumors. In several large series, excellent results have been obtained in up to 92 percent of patients treated for unresectable obstructing tumors of the trachea (Fig. 16-101). Laser resection, however, is not always better than more traditional methods of palliative endoscopic resection using forceps and cautery. Use of laser ablation as the sole treatment for benign tracheal neoplasms is controversial. Once the airway is opened by stenting or endoscopic resection, radiation therapy usually is used as adjunctive therapy for long-term control.

Tracheal Resection

The ease of ventilation during tracheal reconstruction has been greatly facilitated by a variety of anesthetic techniques including the development of high-frequency “jet” ventilation, which is delivered to the distal airway through a small catheter passed through the endotracheal tube (Fig. 16-102). A small tidal volume is delivered at high frequency (60 to 150 breaths/min), maintaining lung expansion, alveolar ventilation, and oxygenation in the normal range. The catheter is small enough to pass through most stenoses and interferes little with surgical exposure.

The choice of incision for tracheal reconstruction depends on the level of tracheal involvement. In a young patient, hyperextension of the neck brings more than half the trachea above the suprasternal notch, accessible through a cervical incision. In older patients, it can be difficult to bring more than the first few tracheal rings above the notch. Lesions involving the upper half of the trachea usually are approached through a cervical collar incision, augmented as necessary with a midline upper sternal extension. Lesions involving the lower half of the trachea can be approached through a midline sternotomy incision or, when near the carina, through a posterolateral thoracotomy.

Surprising lengths of trachea (up to one-half its total length) can be resected and reconstructed with an end-to-end anastomosis. During the operative procedure, frozen-section analysis identifies clear resection margins. Minimizing tension on the anastomosis can be facilitated by releasing the larynx above the hyoid bone (suprahyoid release) and mobilizing one or both hila of the lung (hilar release). Both these maneuvers allow additional lengths of trachea to be resected. In the postoperative period, tension on the anastomosis is prevented by maintaining the neck in flexion, loosely suturing the chin to the anterior chest. For proximal tumors, a variety of complex reconstructions can involve the subglottic area and larynx and, in distal lesions, the carina and both main stem bronchi. The use of prosthetic materials and allografts for tracheal reconstruction generally has been unsuccessful.

Radiotherapy

Both adenoid cystic carcinoma and squamous cell carcinoma are highly radiosensitive. In instances where tracheal resection cannot be accomplished, primary radiotherapy can be curative (squamous cell carcinoma) and palliative (adenoid cystic).

MEDIASTINUM

The mediastinum is the central cavity of the thorax, bounded on either side by the pleural cavities, bounded inferiorly by the diaphragm, and merging superiorly with the thoracic inlet. No compartment of the body carries more physiologic traffic. Many liters of blood pass through the mediastinum each minute, as is true of liters of air, all ingested material and saliva, most autonomic nervous activity, and all the body's lymphatic fluid. Much of the embryologic development of the circulatory, respiratory, and digestive systems takes place within the mediastinum. Congenital, traumatic, inflammatory, and neoplastic processes all find frequent expression in this complex compartment and produce a broad spectrum of pathology in which anatomic relationships assume paramount importance.

The mediastinum is conveniently divisible along rough anatomic boundaries into subcompartments that contain characteristic lesions. The most traditional classification recognizing four spaces has largely given way to a system recognizing three spaces, which divides the highly overlapping contents of the superior compartment between the more surgically relevant anterior and posterior compartments (Fig. 16-103). In this system, the anterior mediastinum lies anterior to the heart and extends cephalad into the anterior half of the thoracic inlet, where it meets the posterior mediastinum. The posterior mediastinum lies behind the heart, extending cephalad into the thoracic inlet where the anterior borders of the upper thoracic vertebrae form its boundary with the anterior mediastinum. The middle mediastinum is the wedge in between with its base lying on the diaphragm and its apex at the top of the aortic arch.

The anterior mediastinum contains the thymus, along with a variable amount of adipose, areolar, and lymphatic tissue. The middle mediastinum contains the heart and pericardium, aorta, trachea and main stem bronchi, and associated lymph nodes. The posterior mediastinum contains the descending aorta, the esophagus, autonomic nerve trunks, and the thoracic duct.

Most mediastinal lesions appear as mass lesions radiographically, and most are neoplasms or cysts. A small number of mediastinal mass lesions are inflammatory or infectious. Vascular lesions, such as aneurysms, are considered elsewhere.

Tumors and Cysts

Mediastinal tumors and cysts in adults are distributed by type with similar frequencies in most large series. Among 400 patients with mediastinal masses reported on by Davis and associates, 25 percent had primary cystic lesions; thymic neoplasms were the most common primary tumors (17 percent), followed closely by lymphoma (16 percent), neurogenic tumors (14 percent), and germ cell tumors (11 percent). Malignant neoplasms have increased to 42 percent of the total over the 56 years encompassed by the series. Among the 62 percent of asymptomatic patients, the fraction of benign neoplasms has decreased from 93 percent before 1967 to 76 percent. This trend toward more frequent detection of occult malignancies probably reflects improved sensitivity of diagnostic techniques. In childhood series, the distribution of neoplasms is skewed toward malignancy, with nearly 50 percent having Hodgkin's or non- Hodgkin's lymphoma, while neurogenic tumors are a distant second. Lymphoma is the most common malignant neoplasm in all age groups.

Clinical Manifestations and Diagnosis

Mediastinal masses produce a wide variety of signs and symptoms, and one-half to one-third of patients are asymptomatic. The most common symptoms are nonspecific (chest pain, cough, dyspnea), and most can be ascribed to compression of adjacent structures, the trachea and esophagus in particular. Superior vena caval obstruction, recurrent nerve palsy, and Horner's syndrome are less common examples, but their presence focuses diagnostic attention on the mediastinum. Certain mediastinal tumors are associated with symptomatic endocrine syndromes, such as hypertension (pheochromocytoma), hypercalcemia (parathyroid tumor), thyrotoxicosis (intrathoracic goiter), and gynecomastia (choriocarcinoma). In such cases, symptoms have nothing to do with the mediastinal location but are systemic consequences of the disease. Pel-Ebstein fevers associated with Hodgkin's disease are a similar example.

The presence of symptoms correlates with malignancy. Ninety-five percent of mediastinal masses that are discovered as incidental radiographic findings are benign, whereas symptomatic lesions are about half benign and half malignant. This correlation is less meaningful in children, whose airways are more vulnerable to compression. In a large series (188 children) from the Mayo Clinic, 78 percent of patients with benign mediastinal masses under age 2 had symptoms and signs of tracheal compression. Signs and symptoms of nerve compression, such as Horner's syndrome, vocal cord paralysis, or hemiplegia, usually reflect aggressive direct invasion and carry a poor prognosis.

Diagnostic evaluation begins with chest radiography. Localizing the mass to one of the three subcompartments of the mediastinum narrows the possibilities (Fig. 16-104) and guides selection of further studies. In most patients the next step is CT, which can sort out the uniform radiographic densities of the mediastinum, identifying normal vascular and soft tissue structures with great cross-sectional clarity. CT of the mediastinum is most diagnostic of benign pathology, such as a cystic mass with an attenuation coefficient close to that of water. The CT appearance of solid malignancies is less definitive, but malignant characteristics such as extension, compression, and invasion are often readily demonstrated. The diagnostic power of CT is enhanced by oral and intravenous contrast material. In one series of children with mediastinal abnormalities, CT provided additional diagnostic information in 82 percent of patients, and in 65 percent the CT findings contributed to a change in clinical management.

MRI is occasionally of additional benefit in imaging the mediastinum, mainly for vascular lesions or for tumors associated with the heart or great vessels. Remarkable definition of vascular structures is obtainable in several views, without the need for contrast material injection (Fig. 16-105). The powerful magnetic field employed contraindicates the use of MRI in patients with pacemakers or cerebrovascular metal clips and complicates examination of critically ill patients on monitors and elaborate life support systems. Fortunately, most metallic hardware likely to occur in the mediastinum (prosthetic valves, vascular clips, sternal wires) does pose a major hazard.

Plane tomography has been replaced by CT and MRI. A barium swallow can demonstrate invasion, compression, or displacement of the esophagus, resulting from intrinsic or extrinsic lesions. Arteriography rarely is necessary in addition to CT and MRI, but contrast material injection of the aorta or pulmonary artery provides information regarding blood supply and anatomic relationship to critical vascular structures that is sometimes not obtainable by any other method. For preoperative evaluation of major vascular disorders (aneurysms), angiography is the diagnostic standard. Venous angiography can provide specific information about the extent of involvement and nature of collateral channels in superior vena caval obstruction but is difficult to justify unless operation and reconstruction are anticipated. Myelography, previously considered an essential part of the evaluation of posterior mediastinal tumors lying very close to the vertebral foramina, has been replaced by CT and MRI of the spine.

Radioisotope scanning can provide very specific information when an anterior mediastinal mass is suspected of being a substernal goiter. Endoscopy of the esophagus or tracheobronchial tree can add observations on displacement, compression, or erosion by adjacent mass lesions and occasionally provides biopsy material. Percutaneous transbronchial or transesophageal needle biopsy sometimes can be used to obtain a diagnosis. Mediastinoscopy and parasternal mediastinotomy are frequently also used for diagnosis. These are the preferred approaches to anterior mediastinal masses suspected to be lymphomas or thymomas because fine-needle aspiration usually does not yield a large enough specimen for definitive tissue diagnosis.

Small, anatomically discreet and encapsulated mediastinal masses are best managed by definitive resection without preoperative biopsy. The operative mortality for resection of such mediastinal lesions is very low. In a large series reported by Davis and associates from Duke, mortality has been 0.8 percent since 1930, without a single death in 236 patients over the most recent period of 26 years. For small mediastinal masses, operation provides definitive diagnosis and simultaneous definitive treatment. Large mediastinal masses, especially those which appear on CT to involve adjacent mediastinal structures, should be biopsied first because most of these will be treated non-operatively (e.g., lymphomas) or with a combined modality approach using preoperative chemotherapy (e.g., germ cell tumors or thymomas).

Neurogenic Tumors

Neurogenic tumors arise from sympathetic ganglia or intercostal nerves and are almost always found in the posterior mediastinum lying in the paravertebral gutter. The peak incidence is in adulthood. Because only 10 to 20 percent of adult neurogenic tumors are malignant, presentation as an incidental finding in an asymptomatic young adult is common. A higher proportion (20 to 40 percent) of childhood tumors are malignant. Chest wall pain caused by nerve compression or bony erosion is the most common symptom. Hemiparesthesia, hemiparesis, and other signs of spinal cord compression can be seen in tumors with “dumbbell” extension through the intervertebral foramina. Hormonally active tumors are most often childhood malignancies, which can produce hypertension, flushing, diarrhea, diaphoresis, anorexia, and fever.

Neurilemoma

Neurilemmomas (schwannomas) account for 40 to 60 percent of all neurogenic tumors. They arise from mature Schwann cells in intercostal nerves and have a hard, yellowish, well-encapsulated gross appearance consistent with the fact that most are benign. Some form “dumbbell” extensions through the intervertebral foramina (Fig. 16-106).

Neurofibroma

Neurofibromas contain elements of both nerve sheath and nerve cells and account for about 10 percent of all neurogenic tumors. They are poorly encapsulated but resemble neurilemmomas radiographically. Mediastinal neurofibromas can be one feature of generalized neurofibromatosis (vonRecklinghausen's disease), in which case the risk of malignant degeneration to neurosarcoma is increased. Advanced age also increases the risk of malignancy. Malignancy is present in 25 to 30 percent of tumors of this type and carries a poor prognosis because of rapid growth and aggressive local invasion.

Neuroblastoma

Neuroblastomas are the most poorly differentiated tumors arising from the sympathetic nervous system. Only about 10 percent occur as a primary lesion in the mediastinum. More than 75 percent occur in children under 4 years of age, and many are hormonally active, producing vanillylmandelic acid in sufficient quantity to present with a systemic symptom complex often consisting of hypertension, fever, vomiting, and diarrhea. Bone, liver, and lymph node metastases, as well as direct spinal cord invasion with neurologic deficits, are not infrequent at the time of diagnosis. Tumors presenting in such advanced stages usually are unresectable, but the tumors are generally radiosensitive, and debulking followed by radiation therapy can produce long-term survival. Tumors presenting in the mediastinum and those presenting in the first year of life have a more favorable prognosis.

Ganglioneuroma, Ganglioneuroblastoma

Ganglioneuromas arise from mature nerve cells in sympathetic ganglia and are benign tumors that usually present in a younger age group than tumors of neural sheath origin. Radiographically, ganglioneuromas have a triangular configuration, with the base toward the mediastinum, and may be completely obscured by the vertebrae in the lateral projection. They are poorly encapsulated and can be difficult to resect because of adherence to adjacent structures. Ganglioneuroblastomas consist of a mixture of mature and immature cells and are rare tumors that share features of neuroblastoma. These usually are seen in patients who are under 3 years of age and are rare in adults.

Paraganglionic Tumors

Pheochromocytomas are chromaffin paraganglionic tumors that characteristically secrete catecholamines. Intrathoracic primaries are unusual, occurring in about 1 percent of all pheochromocytomas. As with all extraadrenal locations, intrathoracic tumors are more frequently “silent” (nonsecreting) than their adrenal counterparts but are also more often malignant—about 30 percent of extraadrenal pheochromocytomas are malignant. Chemodectomas are nonchromaffin paraganglionic tumors that rarely secrete catecholamines and arise from chemoreceptor tissue around the aortic arch, vagus, and aorticosympathetics. They are quite rare, and 15 to 30 percent are malignant.

Treatment

Operation is indicated in most posterior mediastinal neurogenic tumors. The region is best approached through a standard posterolateral thoracotomy. Benign tumors should be excised completely. Preoperative evaluation of all posterior mediastinal tumors includes careful evaluation of the intervertebral foramina and vertebral bodies, which is most easily done initially with a CT scan. MRI is required to confirm intraspinal extension (see Fig. 16-106). When intraforaminal extension exists, resection is performed as a combined thoracic surgical and neurosurgical procedure with spine resection as necessary to provide exposure to the origin of the tumor at the level of the cord.

Malignant tumors are excised if possible. Radical operations for neuroblastoma are approached selectively, keeping clearly in mind the age of the patient, the radiosensitivity of the tumor, and the possibility of spontaneous maturation. Resection of an active (secretory) pheochromocytoma requires attention to the perioperative medical management of paroxysmal hypertension.

Thymoma

In adults, thymoma is the most common anterior mediastinal mass and ranks second in frequency among tumors and cysts of the mediastinum. Thymoma is rare in children and has equal sex distribution, with a peak age incidence between 40 and 60 years. About one-third of patients are asymptomatic at the time of diagnosis. Symptomatic patients present with mass effects on adjacent organs or with systemic effects referable to one of the paraneoplastic syndromes associated with thymoma. Of the former, common examples include cough, chest pain, dyspnea, and superior vena caval obstruction. Of the latter, myasthenia gravis is the most common, although hypogammaglobulinemia and red cell aplasia have

been described. It is most often stated that the incidence of myasthenia gravis is 10 to 50 percent in patients with thymoma. Conversely, thymoma is seen in only 8 to 15 percent of patients with myasthenia gravis. Myasthenic patients with thymoma have a poorer prognosis than patients without thymoma and are less likely to benefit from thymectomy.

Thymoma does not have a characteristic radiographic appearance, and diagnosis usually is made when the mass is excised (Fig. 16-107). The most prevalent histologic classification is based on the relative proportions of lymphocytic and epithelial elements, so the tumor is described as lymphocytic, epithelial, or mixed. Histology does not contribute to the distinction between benign and malignant, which is based on invasive gross characteristics. Distant metastases occur but are uncommon. Local tumor growth with involvement of surrounding mediastinal structures usually is followed by the development of pleural metastases. The pattern of disease progression is reflected in the staging system used for thymoma (Table 16-19).

CT scanning provides important assessment of the extent of the primary tumor. Biopsy should be performed when the tumor is large and is not encapsulated so that preoperative chemotherapy can be given. Smaller tumors that can be removed completely (stage I or II) should be managed by surgical resection alone.

This usually is accomplished via a median sternotomy. Fifty to sixty-five percent of thymomas are benign and subject to curative resection, which should encompass the entire thymus and all adjacent mediastinal adipose tissue. Postoperative irradiation is of unknown benefit and does not compensate or substitute for a complete surgical resection (Fig. 16-108).

Lymphoma

Mediastinal involvement is present in about 50 percent of patients with Hodgkin's and non-Hodgkin's lymphoma, and lymphoma is the most common mediastinal malignancy. Lymphoma is located most frequently in the anterior mediastinum (Fig. 16-109). Hilar nodes in the middle mediastinum are involved less commonly, and posterior mediastinal location is rare. Chemotherapy and radiation are the standard treatment for lymphomas, and resection is almost never indicated. Surgery is performed primarily as a diagnostic procedure, either as an initial biopsy before treatment or to determine if residual active tumor remains after chemotherapy and radiation.

Teratodermoid Tumors

Teratomas account for less than 10 percent of all mediastinal tumors, with almost all found in the anterior mediastinum. By definition, teratomas consist of multiple tissue types not normally found at the site of the tumor. They often are partially cystic and consist primarily of ectodermal elements that can include hair, teeth, and sebaceous glands. Teratomas are thought to arise from branchial cleft and pouch cells associated with the thymus. The mediastinum is second to the gonads as the most frequent location of teratomas in adults. The sex ratio is roughly equal, and age distribution peaks in early adulthood.

In modern series, about two-thirds of patients are asymptomatic at presentation, and the majority of symptoms are nonspecific mass effects such as chest pain, cough, and dyspnea. The classic pathognomonic presentation with cough productive of hair and sebum has become a rarity, since most tumors are detected before eroding into the tracheobronchial tree. As with other neoplasms of the region, malignant teratocarcinomas are more likely to present with symptoms related to aggressive invasion of adjacent vital structures.

Typical radiographic appearance is that of a large, well-circumscribed anterior mediastinal mass. Twenty to forty percent of teratomas are calcified, most often appearing as a nonspecific opacity in the cyst wall, although occasionally due to the presence of teeth or bone. CT scanning is very helpful in delineating involvement of adjacent structures and in confirming fat density in the center of the cystic mass (Fig. 16-110). Elevated serum levels of alpha-fetoprotein (AFP) and human chorionic gonadotropin (hCG) suggest malignancy.

When the mass is small and discreet and likely to be a benign teratoma, surgical excision through a median sternotomy is the best method of diagnosis and treatment. Eighty percent are benign, and resection is curative. Even with benign forms, resection can be difficult because of the tendency for the tumors to be adherent to surrounding structures, most commonly pericardium, lung, great vessels, and thymus, and incomplete resection is occasionally necessary. For benign tumors, recurrence is rare even after partial excision.

Germ-Cell Tumors

Primary extragonadal germ cell tumors are rare. Although they can be seen in the pineal, sacrococcygeal, and paraaortic regions, they are found most often in the anterior mediastinum, where they comprise less than 1 percent of all mediastinal tumors. The histogenesis of germ cell tumors outside the gonads is poorly understood, but a theory of origin from pluripotential primordial germ cells in the mediastinum is favored. Mediastinal teratoma probably should be viewed as the end point of benign differentiation in this germ cell line but usually is considered separately because clinical behavior is different.

Five distinct cell types are recognized. Seminoma and embryonal cell carcinoma are most common, followed by choriocarcinoma, malignant teratoma, and endodermal sinus (yolk sac) carcinoma. These tumors usually are seen in young adults, with a male-to-female ratio of at least 4:1. Because the tumors are highly malignant, it is not surprising that 80 to 90 percent of patients are symptomatic when the diagnosis is made. The most frequent symptoms are nonspecific and result from tumor expansion encroaching on adjacent structures to produce cough, dyspnea, chest pain, or superior vena caval syndrome.

Standard posteroanterior and lateral chest x-rays detect over 90 percent of such tumors. A CT scan provides important information about tumor extent and invasion of adjacent structures. Serum tumor markers suggest the diagnosis and are important to obtain before treatment as a basis for monitoring relapse and response to treatment. All patients with choriocarcinoma have elevated serum human chorionic gonadotropin levels, as will some patients with seminoma and embryonal cell carcinoma. Alpha-fetoprotein levels can be elevated, most commonly in embryonal cell and yolk sac tumors.

The possibility of metastasis from a gonadal tumor must be excluded before a mediastinal germ cell tumor is declared primary. Primary gonadal tumors rarely metastasize only to the mediastinum and most often spread through retroperitoneal lymphatics. A gonadal primary can be excluded with reasonable accuracy if there is no evidence of retroperitoneal involvement by CT scan or lymphangiography and if a gonadal mass is not detectable by palpation and ultrasound examination.

Previously, most patients with mediastinal germ cell tumors underwent exploration through a median sternotomy and an attempt at complete resection. In a large series from the Mayo Clinic, complete resection was achieved in 44 percent of 56 patients. Because of improved chemotherapy during the past decade, most patients are now managed with chemotherapy initially. Surgical resection is used to excise the residual mediastinal mass after chemotherapy and to make a determination about the need for additional treatment postoperatively. Long-term survival is linked primarily to the tumor response to the initial chemotherapy.

Mesenchymal Tumors

Tumors of mesenchymal origin constitute about 7 percent of all mediastinal tumors and cysts, with most occurring in the anterior mediastinum. Lipomas are most common and are characteristically soft masses without fixation to surrounding structures that can reach enormous size without producing symptoms. Fibromas are more dense and less common but have similar clinical behavior. The malignant forms (liposarcoma and fibrosarcoma) are seen rarely.

Tumors of lymph-vascular and blood-vascular origin also are classified as mesenchymal neoplasms. Tumors of blood-vascular origin consist of hemangiomas (capillary, cavernous, and venous) and rare malignant hemangiopericytomas. The most common lymph-vascular tumor is a lymphangioma (cystic hygroma). Most vascular tumors present as smooth, often lobulated masses of uniform density on chest x-ray and appear as cystic masses on CT scan.

The complete list of mesenchymal mediastinal tumors also includes mesothelioma, hamartoma, myxoma, mesenchymoma, leiomyoma, and leiomyosarcoma, xanthogranuloma, and rhabdomyosarcoma.

Endocrine Tumors

Thyroid and parathyroid tumors appearing in the mediastinum are most properly considered within the context of their usual cervical manifestations. Less than 10 percent of parathyroid adenomas are located in the mediastinum, and most are approachable through a cervical incision. Because of their embryologic origin from the third branchial cleft, they usually are in close association with the upper pole of the thymus gland. Parathyroid tumors rarely present as a mediastinal mass.

Mediastinal thyroid tissue usually is a direct substernal extension of the cervical gland. Aberrant mediastinal thyroid tissue with agenesis of the cervical gland is exceedingly rare but does provide the rationale for obtaining a radionuclide thyroid scan in any patient with an undiagnosed mass high in the anterior mediastinum.

Mediastinal Cysts

Congenital cysts constitute approximately 20 percent of all primary mediastinal mass lesions and account for the vast majority of middle mediastinal primary lesions. On chest x-ray they appear as opaque densities that may be indistinguishable from neoplasms except on the basis of typical location. On CT scan a mass with near water density occurring in a characteristic location is virtually diagnostic and provides a strong rationale for routine use of CT scanning in mediastinal lesions. In a recent review of experience with mediastinal cysts in 34 children, Snyder and associates found that the accuracy of their preoperative diagnosis increased from 50 percent before the use of CT scanning to 100 percent thereafter.

Pericardial Cysts

These cysts are the most common type occurring in the mediastinum. They usually are detected as an incidental finding in an asymptomatic patient and very frequently appear at the right costophrenic angle as a smooth-walled cystic mass 3 to 6 cm in diameter. They contain a clear fluid and occasionally communicate with the pericardium. Histologically, they are lined with a single layer of mesothelial cells. The location and appearance of pericardial cysts are so characteristic, especially on CT scan, that close observation is becoming a defensible option, although most are still resected for diagnosis.

Bronchogenic Cysts

Bronchogenic cysts are most frequently located just posterior to the carina or main stem bronchi, although they can be found elsewhere in the mediastinum or more peripherally in the lung (Fig. 16-111). Communication with the tracheobronchial tree can occur to produce an air-fluid level, serving to distinguish them completely from pericardial cysts but allowing for confusion with lung or mediastinal abscess in certain cases. Chest x-ray and CT scan usually demonstrates a cystic mass in the characteristic location, although bronchogenic cysts can contain a viscid fluid difficult to distinguish from a solid mass by CT scan alone. A contrast esophagram may show compression of the esophagus by an anterior mass. Histologically, they are lined with ciliated respiratory epithelium and contain varying amounts of cartilage, smooth muscle, and mucous glands. They are most frequently symptomatic in children, producing cough, dyspnea, and stridor in more than half. All bronchogenic cysts should be resected and usually are approached through a posterolateral thoracotomy. When they have formed a communication with the tracheobronchial tree, the chronic infection that frequently results can make resection through dense

inflammatory adhesions very difficult.

Enteric Cysts

Enteric cysts are located in the posterior mediastinum adjacent to the esophagus. They are occasionally embedded in the muscularis of the esophagus but rarely communicate with the esophageal lumen. The cysts have a smooth wall with a muscular coat and a lining recognizable as intestinal mucosa, although it may be ciliated, and they contain a clear, colorless mucoid fluid. When lined with an aberrant gastric mucosa, peptic ulceration can lead to perforation of adjacent bronchus or esophagus, producing hemoptysis or hematemesis, and erosion into adjacent lung can produce a lung abscess. A rare association with vertebral anomalies has been described in which the enteric cyst is attached to the spinal cord of meninges, and a patent tract may exist that can be demonstrated by myelography.

Approximately 60 percent of enteric cysts are recognized in patients under 1 year of age when symptoms of tracheal and esophageal compression are prominent. Less than one-third of children with enteric cysts are asymptomatic. Complete evaluation of children with a suggestive presentation includes chest x-ray and esophagram followed by a CT scan with contrast material in the esophagus. Resection is always indicated. The lesions are approached through a posterolateral thoracotomy, with the choice of side determined by the level of involvement and the appearance of projection into either hemithorax.

Mediastinitis

Acute Mediastinitis

Acute mediastinitis is a fulminant infectious process with high morbidity and mortality characterized by rapid spread through the areolar planes of the mediastinum. The mediastinal pleura confines the process to the mediastinum only temporarily, with a breach occurring into one or both pleural cavities early in the course of the infection in most cases, after which the negative pressure of the pleural space helps to rapidly spread the infection throughout. The rapid spread of infection is promoted by several factors. One is the separation of tissue planes produced by air forced into soft tissues adjacent to a perforated hollow viscus, most often the esophagus, further promoted by the digestive action of salivary and gastric enzymes. Another is the pressure gradient established from the atmosphere to the negative pressure of the pleural space once the pleura is penetrated, which pulls the infection through the mediastinum from its source and into the pleural space. A third factor is the presence of naturally continuous fascial planes connecting the deep cervical compartments with the mediastinum, along which oropharyngeal infection can spread.

The infection is initiated most frequently by esophageal perforation, resulting from instrumentation, trauma, foreign body, suture line leak, or spontaneous postemetic rupture (Boerhaave's syndrome). Tracheal rupture or perforation is a less common cause in which dissemination of air through the soft tissues is massive, and infection is likely to be a secondary development. Direct necrotizing spread of infection without violation of an intrathoracic viscus is seen most commonly with aggressive oropharyngeal infections involving the deep cervical space but also has been described in association with infections of ribs, sternum, and vertebrae.

Chest pain, dysphagia, respiratory distress, and cervical upper thoracic subcutaneous crepitus are the chief hallmarks of the process during the earliest stages of infection when it is most important to diagnose the problem and begin treatment. Evidence of fulminant systemic infection is certain to appear within 24 h, and florid sepsis with hemodynamic instability supervenes rapidly in untreated patients.

The chest x-ray may be normal very early in the process, although mediastinal and subcutaneous air becomes apparent in most patients. The mediastinal contour is usually wide, and pleural effusion with or without pneumothorax appears very frequently. A contrast esophagram, for which water-soluble contrast material usually is recommended, is essential when esophageal perforation is known or suspected. Esophagoscopy rarely is indicated in acute perforation. Infections resulting from esophageal perforation and those descending from a perioral source usually are caused by a mixture of gram-positive and gram-negative aerobic and anaerobic organisms representing the spectrum of oral flora. Initial antibiotic coverage should be broad enough to cover all possibilities until cultures are available.

Treatment must be early and aggressive because mortality is directly linked to length of time between injury and esophageal repair or diversion. Antibiotics and fluid resuscitation are begun immediately. Chest tubes are placed for pneumothorax or effusion. The primary problem, such as esophageal perforation, is treated according to accepted principles, either separately or in combination with drainage procedures. For preservation of the cervical esophagus, direct drainage of the neck is required, entering the deep cervical space through an incision parallel to the sternocleidomastoid muscle and retracting the muscle laterally to expose the carotid sheath and pretracheal and retrovisceral spaces. Perforations of the thoracic esophagus are approached via a thoracotomy, which allows primary repair of the site of perforation and buttressing of this area with adjacent well-vascularized tissue (e.g., intercostal muscle or pericardial fat pad). The thoracotomy should be performed via the pleural space into which the esophageal perforation has drained, usually the right side for perforations in the middle or upper thoracic esophagus and the left side for perforations of the distal esophagus. Thorough debridement and drainage of the pleural space and mediastinum in conjunction with closure of the perforation are essential. Esophageal diversion by cervical esophagostomy, gastrostomy, and ligation or division of the gastroesophageal junction is required only when primary closure has failed or cannot be performed because of severe mediastinal inflammation related to a delay in diagnosis (usually 48 h or more).

Mediastinitis is seen in 1 to 4 percent of patients after open heart surgery and has accounted for an increasingly large proportion of all cases of mediastinitis as open heart procedures have increased in frequency. It follows a more indolent course than the entities discussed above, is rarely associated with crepitus and mediastinal air on x-ray, and has the bacteriologic spectrum of other wound infections, with *S. aureus* and *Streptococcus epidermidis* predominating. Several techniques have improved the treatment of this complication, which include sternal debridement and reclosure with mediastinal irrigation for infections diagnosed early and the rotation of bilateral pectoralis major muscle flaps into the sternal defect for more severe infections.

Chronic Mediastinitis

Chronic inflammation and fibrosis in the mediastinum (sclerosing mediastinitis, fibrosing mediastinitis) are thought to result most often from granulomatous infection such as tuberculosis or histoplasmosis, although identification of an organism in individual patients is rare. It has been postulated that the process begins as an inflammatory reaction in the tissues surrounding involved lymph nodes. The process is likely to remain clinically silent unless it progresses to produce obstruction of the esophagus, airways, superior vena cava, or other mediastinal vascular structures. The chest x-ray may show mediastinal widening but often is normal. CT scanning combined with angiography may be necessary to define the process. Operative exploration frequently is required to establish a diagnosis and also can be undertaken to relieve obstruction. In a series of 22 patients, medical treatment with the antifungal agent ketoconazole was surprisingly effective at controlling progression of disease when combined with operation.

Superior Vena Caval Obstruction

Superior vena caval obstruction is caused by bronchogenic carcinoma in 85 percent of cases. In the remainder, the cause is another mediastinal tumor, fibrosing mediastinitis, thoracic aortic aneurysm, or caval thrombosis secondary to chronic indwelling catheters or instrumentation. At least 40 percent of bronchogenic carcinomas producing superior vena caval obstruction are small cell tumors. Obstruction can be caused by compression or direct invasion. The clinical syndrome produced is easily recognizable, consisting of venous distention, facial edema, and plethora, often accompanied by headache and respiratory symptoms. In rare patients, associated airway compression or laryngeal edema can be life-threatening, but there is otherwise little evidence to support the commonly held notion that superior vena caval obstruction is inherently dangerous. Seizures, intracranial venous thrombosis, and other nonspecific cerebral consequences are unusual and highly associated with the presence of brain metastases. Survival in patients with obstruction due to carcinoma usually is measured in weeks to months, and it can be difficult to separate the dismal prognosis and aggressive behavior of the primary disease from the effects of superior vena caval obstruction alone. As with venous obstruction elsewhere in the body, compensatory venous collaterals develop promptly and largely ameliorate the condition, a fact that also complicates objective assessment of treatment modalities.

The vascular diagnosis can be confirmed by venography, but CT scanning with venous contrast material is equally effective and provides additional information regarding surrounding structures that can be diagnostically valuable. Invasive diagnostic procedures, such as mediastinoscopy, bronchoscopy, and lymph node biopsy, may be more difficult because of elevated venous pressure. Respiratory complications related to venous engorgement and edema of the tracheobronchial mucosa can occur but are almost always manageable in the hands of a careful anesthesiologist. The clinical tradition favoring emergency radiation therapy for the clinical syndrome before performing a biopsy is no longer accepted. A tissue diagnosis should be obtained before starting treatment.

Because the majority of cases are caused by an incurable neoplasm, palliative radiation with or without combination chemotherapy is by far the most common treatment modality. Rare cases of benign etiology are treated occasionally with venous bypass, but without large numbers of reportable patent conduits. Bypasses from the jugular vein to the atrium or distal superior vena cava have been accomplished with femoral vein and with a spiral graft constructed from excised saphenous vein (Fig. 16-112). Saphenojugular bypass also has been described, in which the saphenous vein is routed to the neck through a subcutaneous tunnel and left attached at the saphenous bulb for outflow. All invasive treatments have in common the difficulty of predicting which patients will be unable to establish sufficient venous collaterals over time without operation.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 17 - Congenital Heart Disease

Aubrey C. Galloway
Michael Artman
Stephen B. Colvin

GENERAL CONSIDERATIONS

Introduction

Congenital structural malformations of the heart or great vessels constitute the most common form of heart disease in children. The incidence of congenital heart disease is nearly 1 percent of live births (approximately 8/1000). Despite major advances in diagnosis, medical therapies, and operative treatment during the past decade, congenital heart disease is a leading cause of morbidity and mortality in infants and children. The highest proportion of deaths from birth defects in the first year of life is attributable to congenital heart defects. Although this chapter focuses on congenital malformations, acquired diseases also can occur in children, including conditions such as rheumatic fever (mitral and aortic valve regurgitation), Kawasaki's disease (coronary artery aneurysms) and others that may require operative intervention.

Given the intricacy of the embryonic development of the heart and cardiovascular system, it is not surprising that structural malformations occur frequently. A functioning cardiovascular system is essential to the survival of the fetus, and consequently the heart and vasculature develop early in gestation. Most fetal heart structures are formed between the third and the eighth week of pregnancy. The primitive fused heart tube undergoes looping, the first break in symmetry of the developing embryo, and septation in a highly ordered and tightly controlled fashion to form a normal four-chambered functioning heart during this 4- to 5-week period. Cellular commitment, differentiation, and migration are genetically controlled processes that may be affected by environmental or epigenetic influences. The developing heart receives important contributions from neural crest derivatives, and signals from the extracellular matrix are probably involved in directing cell migration. The developing branchial arch system contributes to the development of the aorta and ductus arteriosus. Disturbances in any of these processes may have profound effects on the final structure of the heart and great vessels after development is complete. The wide spectrum of congenital heart defects underscores the complexity of these normal cardiovascular developmental processes.

It is increasingly apparent that disordered developmental mechanisms are likely to be responsible for a high percentage of congenital cardiac defects. Basic developmental errors in the genes regulating cardiac development result in primary malformations that may be lethal to the embryo. Defects may arise from environmental influences during intrauterine development such as teratogens (e.g., retinoic acid), toxic agents, infections, and alterations in blood flow. Embryos with the most severe malformations are not viable, and only a fraction of all embryos with cardiovascular malformations survive to birth. Single-gene mutations may account for a large proportion of congenital heart defects. The role of chromosomal abnormalities such as trisomy 21, Down syndrome with frequent atrioventricular septal defects, has been known for some time, but the specific gene(s) involved has yet to be determined. Examples of recognizable single-gene defects include Marfan syndrome, supravalvular aortic stenosis, long QT syndrome, familial hypertrophic cardiomyopathy, Williams syndrome, Holt-Oram syndrome, and Noonan's syndrome. Considerable interest has been focused on the finding of a strong relationship between monosomy of a locus on chromosome 22 and conotruncal malformations. Deletions in the 22q11 region have been associated with DiGeorge syndrome, velocardiofacial syndrome (Shprintzen's syndrome), and conotruncal anomaly face syndrome. These syndromes share common features, and the phenotypic abnormalities represent a spectrum of disorders attributable to a related developmental field defect. Patients with microdeletions of chromosome 22q11 exhibit a variety of conotruncal and aortic arch anomalies, including truncus arteriosus, interrupted aortic arch, and tetralogy of Fallot. Although the specific genes involved remain to be identified, studies of the molecular genetics of congenital heart defects hold great promise for improved understanding, diagnosis, and treatment of affected infants.

The fetal circulation has several distinctive features that may allow the fetus to grow and develop normally during intrauterine life with a congenital heart defect that is lethal if untreated after birth. During fetal development, the lungs are collapsed and have a high vascular resistance, and pulmonary blood flow is low. Much of the oxygenated blood returning from the placenta through the inferior vena cava to the right atrium flows through the foramen ovale into the left atrium and the left ventricle. Most of the blood expelled from the right ventricle into the pulmonary artery is shunted through the ductus arteriosus into the descending thoracic aorta. Fetuses with severe right ventricular or left ventricular outflow tract obstruction, ventricular septal defects, or aortic arch anomalies thrive in utero. Shortly after birth the circulatory patterns change as a consequence of the normal transition to extrauterine life. It is during this transition time that many structural congenital abnormalities become apparent. At birth the lungs expand to assume the role of oxygenation. There is normally a rapid fall in pulmonary vascular resistance, and the ductus arteriosus closes in the first few days after birth. The ductus remains patent in only a small proportion of individuals but is one of the most common forms of congenital heart disease. The foramen ovale becomes functionally closed within a few days or weeks, although it may remain open anatomically for several months. The foramen ovale is a slitlike channel that seals when left atrial pressure becomes higher than right atrial pressure. Persistent patency of the foramen ovale, usually an innocuous defect, occurs in perhaps 10 to 20 percent of adults and has been implicated occasionally in thromboembolic stroke. With elevation of right atrial pressure above left atrial pressure from any cause, the foramen ovale may be stretched open and create a right-to-left shunt from the right atrium to the left atrium. Right-to-left shunting can be of sufficient magnitude to produce cyanosis from shunting of unoxygenated blood into the systemic circulation in infants with critical pulmonary valve stenosis.

Although a broad spectrum of congenital heart defects has been recognized and characterized, in large pediatric cardiac clinics several of the most common malformations comprise the majority of the abnormalities seen. Isolated ventricular septal defect is the most common anomaly, representing 20 percent or more of all patients. Other relatively common malformations (each occurring in 10 to 15 percent of patients) are patent ductus arteriosus, atrial septal defect, pulmonic stenosis, aortic stenosis, coarctation of the aorta, tetralogy of Fallot, and transposition of the great arteries.

Classification

Congenital heart disease may be classified by the type of anatomic abnormality present, which in turn produces a distinct physiologic disturbance. Five major groups exist: (1) *obstructive left-sided lesions* that restrict systemic blood flow; (2) lesions producing *increased pulmonary blood flow* by left-to-right shunting of blood across a cardiovascular defect; (3) lesions producing *cyanosis* by shunting right-to-left away from the lungs into the systemic circulation; (4) *complex malformations with mixed physiology*; and (5) *anomalous origin of vessels and vascular rings*.

Pathophysiology

The physiologic consequences of congenital heart disease vary from mild to severe, depending on the anatomy. The mildest forms consist of abnormal physical findings with minimal derangements of physiology. In some instances, such as with mild pulmonic stenosis or a small ventricular septal defect, treatment may be unnecessary or may be delayed until later in life. With more severe defects, symptoms may be more consequential. Physiologic abnormalities, such as pressure gradients across stenotic valves, shunts through septal defects, and elevated pulmonary artery pressure, can be detected by echocardiography or measured by cardiac catheterization. If untreated, patients may develop congestive heart failure or severe hypoxemia. Eventually physiologic abnormalities worsen and corresponding anatomic changes occur, such as cardiac hypertrophy, irreversible ventricular dysfunction, or pulmonary vascular disease.

Obstructive Left-Sided Lesions

The most common disorders are aortic valvular stenosis and coarctation of the aorta. These lesions impede emptying of the left ventricular chamber, resulting in systolic pressure overloading and corresponding hypertrophy of the ventricle. As the ventricular response is predominantly concentric hypertrophy, cardiac enlargement may be difficult to detect by clinical examination, and the chest radiograph may be normal or only slightly abnormal. The electrocardiogram (ECG) is helpful in identifying the presence of left ventricular hypertrophy. The echocardiogram not only can assess wall thickness and myocardial mass but also can provide a noninvasive estimate of the pressure gradient across the stenotic aortic valve. Similar information can be obtained with magnetic resonance imaging (MRI) techniques. MRI is helpful in defining the aortic anatomy in cases of coarctation of the aorta. MRI often provides superior images of the aorta, especially in older children and adolescents in whom echocardiographic imaging of the aortic arch and descending aorta is limited when using conventional transthoracic approaches.

With progressive left ventricular hypertrophy, susceptibility to ventricular arrhythmias develops, and sudden death may occur in patients with aortic stenosis, often during exercise, when myocardial oxygen demands increase disproportionately to oxygen delivery. Newborn infants with critical aortic stenosis or coarctation of the aorta may exhibit severe, life-threatening heart failure. In these cases, emergency intervention is indicated. Otherwise, cardiac failure is uncommon in older children and, if present, usually is a late and preterminal manifestation. Surgical intervention should be timed to prevent the development of severe hypertrophy and ventricular dysfunction. Left-sided obstructive lesions are managed in close cooperation with the interventional cardiologist in order to provide optimal surgical and catheterization-based treatments.

Left-to-Right Shunts

The direction and magnitude of flow through an unrestrictive ventricular septal defect or patent ductus arteriosus depends on the relative vascular resistances in the pulmonary and systemic circulations. Because systemic vascular resistance is normally greater than pulmonary vascular resistance, a defect in the ventricular septum or a connection between the aorta and the pulmonary artery (e.g., patent ductus arteriosus) results in a shunt of oxygenated blood from the left-sided circulation to the right side (hence the term left-to-right shunt). Shunting through an uncomplicated atrial septal defect is also left-to-right because of differences in ventricular compliance that favor flow into the right atrium and pulmonary circulation. A left-to-right shunt produces an increase in pulmonary blood flow. Systemic blood flow often is preserved unless there is marked pulmonary blood flow with significant signs and symptoms of congestive heart failure. Cyanosis is not a feature of isolated left-to-right shunts. The most common defects producing left-to-right shunts are ventricular septal defect, atrial septal defect, patent ductus arteriosus, and atrioventricular septal defects.

Pulmonary Congestion

A left-to-right shunt becomes physiologically significant when the pulmonary blood flow is 50 percent to two times greater than the systemic flow (see Cardiac Catheterization section for calculation of Q_p/Q_s). Large shunts may produce a pulmonary blood flow three to four times greater (or more) than systemic blood flow, with a pulmonary blood flow exceeding 10 to 15 L/min/m² of body surface area. During infancy and early childhood, the resulting pulmonary congestion increases the work of breathing and basal energy expenditure. Infants with large left-to-right shunts may tire during feeding and be unable to achieve sufficient caloric intake for normal growth. The predominant presentation in infancy is failure to thrive with signs of increased respiratory effort. Pulmonary congestion is associated with increased susceptibility to lower respiratory tract infections, and recurrent bouts of pneumonia may occur in the first few years of life. Beyond early childhood, however, high pulmonary blood flows often produce surprisingly little disability for a period of time. With careful questioning of the family, it often becomes apparent that the child has diminished exercise tolerance with easy fatigability relative to his or her playmates. Sometimes the degree of impairment is subtle, and it is only in retrospect after surgical correction that the family recognizes a change in the child's exercise capacity.

With the increase in pulmonary blood flow there is a corresponding enlargement of the left ventricle because of increased pulmonary venous return to the left atrium, and the right ventricle becomes volume overloaded in the presence of a large atrial septal defect. Volume overloading of the ventricle generally produces chamber dilation, rather than hypertrophy, which can be recognized on clinical examination and on the chest radiograph. Changes in the electrocardiogram are often less prominent than those seen with concentric hypertrophy. Echocardiography is useful for evaluating chamber sizes more precisely.

Medical management of large left-to-right shunts with pulmonary congestion includes vasodilator and diuretic therapy, often with digoxin therapy in addition, although the value of digoxin is uncertain unless left ventricular failure is present. Left ventricular afterload reduction with angiotensin-converting-enzyme inhibitors decreases the amount of left-to-right shunting, while the diuretics decrease fluid retention and limit pulmonary congestion by diminishing *volume* overload. In many cases, shunting through a left-to-right shunt diminishes with time as the child grows and the defect becomes smaller. Medical therapy may be used to temporize, and not all children with a left-to-right shunt need surgical repair. However, if an infant fails to grow despite medical therapy or if there are complications because of a large left-to-right shunt, then operative intervention is indicated.

Increased Pulmonary Vascular Resistance

It is important to distinguish between pulmonary hypertension and pulmonary vascular disease. Infants and children with a large unrestrictive ventricular septal defect have essentially equal pressures in the left and right ventricles, because the ventricles are functionally connected by the septal defect. Consequently, the pressures in the right ventricle and pulmonary arteries are elevated to systemic levels, a condition defined in its strictest sense as pulmonary hypertension. The pulmonary vascular resistance normally is low, and there is markedly increased pulmonary blood flow even though the patient has pulmonary hypertension. It is more meaningful from a physiologic and prognostic perspective to characterize each patient with respect to their pulmonary vascular resistance rather than the pulmonary artery pressure.

Elevated pulmonary blood flow and pressure ultimately produce changes in the pulmonary vasculature, resulting in a progressive increase in pulmonary vascular resistance. Pulmonary vascular resistance is calculated with the following formula (in which Q_p is measured in L/min/m² body surface area):

$$PVR = \frac{\text{Mean pulmonary arterial pressure} - \text{Left atrial pressure}}{Q_p}$$
 (Pulmonary blood flow)

- Left atrial pressure / Q_p (Pulmonary blood flow)

Normal pulmonary vascular resistance is less than 2.5 Wood units by this formula. Pulmonary hypertension resulting from an increase in pulmonary blood flow subsides as soon as the cardiac defect producing the increase in blood flow is corrected. Elevated pulmonary vascular resistance may take longer to resolve because it is a result of abnormal reactivity of the pulmonary vascular bed with thickening of the media and intima. If corrected early, pulmonary vascular disease progression is halted and the vascular changes may regress. If left untreated, children and adolescents with large left-to-right shunts can develop irreversible pulmonary vascular disease. When pulmonary vascular resistance is fixed above 10 Wood units and the shunt is balanced or changes to predominantly right-to-left, a condition known as Eisenmenger's syndrome, the operative risk becomes exceedingly high. In evaluating pulmonary hypertension, the significant physiologic measurement is the degree of change in the pulmonary vascular resistance, as calculated from the relation between flow and pressure, and not the absolute level of the pulmonary artery pressure. If the pulmonary resistance is elevated, it is helpful to determine whether or not the vessels are reactive to oxygen, nitric oxide, or other pulmonary vasodilators. Even if the baseline resistance is high, a child may be considered a good candidate for surgery if the pulmonary vascular bed remains reactive and is capable of dilating in response to appropriate stimuli.

It is rare for permanent pulmonary vascular changes to occur before 1 to 2 years of age in uncomplicated ventricular septal defect or patent ductus arteriosus, but there is significant individual genetic variation in susceptibility to the development of these changes. Some children with a large ventricular septal defect and a large increase in pulmonary blood flow do not develop any increase in pulmonary vascular resistance, while others with smaller defects develop significant pulmonary vascular changes at an early age. Infants with trisomy 21 and ventricular septal defect or atrioventricular septal defect are at increased risk for early development of fixed pulmonary vascular disease, so repair at approximately 2 to 3 months of age is advised. Similarly, defects such as truncus arteriosus or transposition of the great arteries with ventricular septal defect may produce permanent pulmonary vascular disease in some infants before 6 months of age. Most lesions associated with an increase in pulmonary vascular resistance should be surgically corrected in the first 3 to 12 months of life. More complex defects, such as transposition or truncus arteriosus, generally require operation in the first few weeks of life. With a simple atrial septal defect, the early onset of pulmonary vascular disease almost never occurs, and repair in infancy generally is not necessary.

Cyanotic Lesions

Right-to-left shunting of systemic venous blood directly back into the systemic circulation results in arterial hypoxemia and cyanosis. Cyanosis usually occurs because of the combination of an anatomic obstruction that results in *decreased pulmonary blood flow* and an intracardiac defect that allows right-to-left shunting of unoxygenated blood. The classic example is the tetralogy of Fallot, a combination of ventricular septal defect and right ventricular outflow tract obstruction. Other cyanotic disorders with decreased pulmonary blood flow include pulmonary atresia with intact ventricular septum, tricuspid atresia, and, in many cases, Ebstein's anomaly. Some more complex malformations, such as transposition of the great vessels, double-outlet right ventricle, truncus arteriosus, and single ventricle, produce mixing of unoxygenated and oxygenated blood, with bidirectional shunting and cyanosis, despite normal or increased pulmonary blood flow.

Cyanosis resulting from significant right-to-left shunting produces a large number of physiologic disturbances because of the tissue anoxia resulting from chronic hypoxemia. Most disturbances result from deficient oxygen transport to tissues of the body. Cardiac failure is rare in an uncomplicated cyanotic patient, in contrast to its more frequent occurrence in patients with large left-to-right shunts. The combination of cardiac failure and cyanosis can occur with complex lesions that produce mixing of blood with bidirectional shunting, yet have nonrestricted pulmonary blood flow. These lesions include transposition of the great vessels, double-outlet right ventricle, truncus arteriosus, and single ventricle.

The degree of cyanosis depends on the degree of anoxia and the blood hemoglobin concentration, since the visible intensity of cyanosis is determined by the amount of reduced hemoglobin in the circulation. It has been estimated that about 5 g of reduced hemoglobin is required to produce visible cyanosis. Normally in the capillaries about 2.25 g of reduced hemoglobin is present. With an average hemoglobin concentration of 15 g/dL of blood, a decrease in arterial oxygen from the normal range of nearly 95 to 75 percent is needed to produce visible cyanosis. In the presence of anemia, a more severe degree of anoxia is required to produce visible cyanosis, while with polycythemia and hemoglobin concentrations of 20 g/dL or more, severe cyanosis occurs with lesser degrees of anoxia. Cyanosis may be difficult to detect in newborn infants because of physiologically high levels of fetal hemoglobin.

Central cyanosis results from a defect in oxygenation of blood in the lungs or from an intracardiac right-to-left shunt. Cyanosis resulting from a pulmonary abnormality usually can be suspected because of the presence of signs and symptoms of respiratory distress. Cyanosis due to lung disease generally improves when the patient breathes 100% oxygen. This is a quick and helpful test to perform when confronted with a newborn baby with cyanosis. A rise in the arterial partial pressure of oxygen to above 150 mmHg makes the diagnosis of cyanotic congenital heart disease much less likely, though it does not rule it out. In the catheterization laboratory, pulmonary disease is suspected if the oxygen saturation of blood in the pulmonary veins is less than 95 percent. Pulmonary insufficiency from cardiac disease occurs only with severe pulmonary congestion from cardiac failure or far-advanced pulmonary vascular disease.

In contrast to cyanosis from pulmonary causes, cyanosis from an intracardiac shunt permits direct entry of venous blood into the systemic circulation, causing cyanosis that is unresponsive (or only minimally responsive) to an increase in the fraction of inspired oxygen. The intensity of the cyanosis is related to the volume of pulmonary blood flow, for ultimately cyanosis depends on the relative proportions of unoxygenated and oxygenated blood in the arterial circulation. Even though a large intracardiac shunt is present, an increase in pulmonary blood flow to produce a larger amount of oxygenated blood can substantially reduce cyanosis and improve oxygen transport. This was the rationale for palliative systemic-to-pulmonary artery shunts to increase pulmonary blood flow.

Clinical features of chronic cyanosis are seen much less commonly now that early diagnosis and effective surgical treatment are available for most forms of cyanotic congenital heart disease. There are still children who escape early detection and present with many of the classic features of chronic cyanosis. Two distinctive changes that appear with chronic cyanosis are clubbing of the digits and polycythemia, and the triad of cyanosis, clubbing, and polycythemia is a familiar one in children with congenital heart disease. Clubbing of the digits, or hypertrophic osteoarthropathy, is an unusual change in the appearance and structure of the digits, consisting of a rounding of the tips of the fingers and toes and a thickening of the ends, associated with deposition of fibrous tissue. In addition, there may be a pronounced convexity of the fingernails. Histologically, the fingers have increased numbers of capillaries, with a large number of tiny arteriovenous aneurysms. Clubbing usually is not prominent until a cyanotic child is 1 to 2 years of age, but in some instances of severe anoxia it may evolve within several weeks. Clubbing gradually subsides after correction of the intracardiac defect.

Polycythemia is a normal physiologic response of the bone marrow to chronic anoxia. Up to a point this response is beneficial; an increase in red cell and hemoglobin concentration increases the ability of the blood to transport oxygen. Hematocrit concentrations of 60 to 70 percent are not uncommon with chronic cyanosis; values exceeding 80 percent are noted in extreme cases. There is a parallel rise in viscosity of the blood, with restriction to the flow of blood as the hematocrit level rises. Once the hematocrit level exceeds 75 to 80 percent, the blood's increased viscosity may constitute a significant hazard for the development of cerebral venous thrombosis and neurologic injury. In chronic polycythemia defects in blood coagulation may develop, with abnormalities in several components of the blood-clotting mechanism. These patients may have increased susceptibility to hemorrhage after surgical procedures.

A decrease in exercise tolerance, with dyspnea on exertion, is a characteristic feature of cyanotic heart disease, because the circulation is unable to increase oxygen transport during exercise. The severity of the disability or its progression can be measured with formal exercise testing. Cyanotic children quickly learn that dyspnea on walking can be lessened by assuming a squatting position. Physiologic studies indicate that squatting produces an increase in peripheral vascular resistance with a corresponding increase in pulmonary blood flow by diminishing the degree of right-to-left shunt. Squatting is most commonly seen in older children with unrepaired tetralogy of Fallot, but it may occur in other cyanotic conditions.

Periodic episodes of unconsciousness, termed *hypercyanotic spells*, are a sign of cerebral anoxia. They often appear in the third or fourth month of life in severely cyanotic children. These episodes characteristically occur at different times, not always associated with exertion, and evolve as episodes of crying, deepening cyanosis, and loss of consciousness lasting a few minutes to a few hours. Emergency treatment to improve the oxygen content of the arterial blood is indicated if hypercyanotic spells occur. Medical treatment includes placing the infant in the knee-chest position, administration of oxygen, morphine for sedation, and phenylephrine to increase systemic resistance and reduce the magnitude of the right-to-left shunt. Urgent surgical placement of a systemic-to-pulmonary shunt or early total correction of the defect is indicated once hypercyanotic spells develop. In selected cases, pulmonary blood flow can be increased in the catheterization laboratory by dilating a stenotic pulmonary valve.

Another cause of neurologic injury in cyanotic children is brain abscess. The increased susceptibility may be partly related to direct access of bacteria in the venous circulation to the arterial circulation through the right-to-left shunt in the setting of polycythemia and sluggish capillary flow. A localized infarct with subsequent bacterial infection may explain the evolution in some patients. Cerebral injury may occur because of paradoxical thromboembolism through an intracardiac defect. In this instance, a thrombus migrating through the venous circulation, which would normally lodge in the pulmonary bed, traverses an intracardiac defect and reaches the cerebral circulation.

In older children with severe cyanosis there is a striking increase in the bronchial circulation through the development of aortopulmonary collateral vessels as a compensatory response to the chronic decrease in pulmonary blood flow. The myriad of collateral vessels, often constituting a mass of varicosities in the mediastinum, are of surgical significance because of the risk of bleeding during operation, and because they produce overperfusion of the lungs if left in place once a surgical shunt is created. It is helpful in this setting to

coordinate the operative intervention with the occlusion of aortopulmonary collateral vessels in the catheterization laboratory. The use of specially designed coils delivered through a catheter to occlude the major collateral vessels can reduce the time in the operating room and minimize flooding of the lungs once a controlled surgical shunt is created. In patients undergoing total definitive operative repair, preoperative occlusion of major aortopulmonary collateral vessels simplifies the operative procedure and minimizes flooding of the field during cardiopulmonary bypass.

Examination

History

A thorough history and physical examination are important components of the evaluation of an infant, child, or adolescent with known or suspected heart disease. The parents or caretakers are the source of most of the essential information related to the child's health. The family often is very concerned because they lack information or they harbor inaccurate perceptions regarding the prognosis of heart disease in a child. The thought of heart surgery being performed on their child provokes extreme anxiety. It is essential to establish good rapport and close communication with the patient and the parents in order to minimize these anxieties and ensure proper treatment and compliance. Most parents want to be well informed in order to participate in decisions that may dramatically affect their child.

Initial historical information should be collected regarding pertinent maternal and familial factors. The general state of the mother's health during pregnancy should be ascertained along with questioning about medications, infections, and lifestyle. A careful family history is essential to provide clues as to possible hereditary disorders. The perinatal history should include questions about birth weight, illnesses at birth, length of hospital stay after birth, and whether or not a murmur or cyanosis was detected in the newborn nursery.

Significant congenital heart disease often presents in infancy, so a careful analysis of the first few weeks and months of life is essential. Often the presence of heart disease in infants may go unrecognized because the signs of heart disease lack specificity. For example, feeding difficulties may result from a variety of conditions other than congestive heart failure. Significant arterial desaturation can be present without obvious cyanosis. If the infant is the first child, new parents may not recognize abnormal infant behavior since they have no previous experience with a normal baby. Specific questions about feeding include the type of feeding (formula versus breast milk), frequency of feeds, amount consumed with each feeding (or time spent at each breast), and duration of each feeding should be recorded. It is important to ask if the infant seems to breathe harder or perspire during feeding. The pattern of growth (height, weight, and head circumference) should be ascertained. It is useful to plot the child's growth parameters on an appropriate growth chart in order to determine the pattern of growth. The age at which a cardiac murmur was detected for the first time should be noted. The time of appearance of cyanosis is significant as well as an assessment of whether or not the degree of cyanosis is progressing.

A decrease in exercise tolerance, manifested by dyspnea on exertion, is a common symptom and an indication of the severity of the disorder. This may be subtle, so it is important to ask the parents to compare the child's activity level with that of his or her playmates. Sometimes the school can provide useful information. Often the parents indicate that the child does not exhibit exercise intolerance, but on careful questioning it becomes apparent that the child's activity is self-limited and the child fails to engage in strenuous forms of play activities. Frequent squatting while playing can be readily identified by the parents, although they may think of this as normal for their child and usually do not volunteer this information. Older children and adolescents should be allowed to participate in the history gathering and answer questions related to activity and exercise tolerance. Older children should be asked about the presence of symptoms such as shortness of breath, chest pain, palpitations, pre-syncope, and syncope. Previous signs or symptoms of neurologic episodes such as cyanotic spells, cerebral embolism, brain abscess, or other signs of cerebral injury should be noted. An inquiry should be made about infections such as pneumonia, bacterial endocarditis, or rheumatic fever.

Physical Examination

General inspection should be performed to assess the overall appearance of the infant or child looking for phenotypic features characteristic of common syndromes associated with congenital heart defects (e.g., trisomy 21, Turner's, Noonan's, velocardiofacial, or Williams syndromes). The examiner should be alert to evidence of cyanosis, respiratory difficulties (respiratory rate, effort, and use of accessory muscles), diaphoresis, or whether the infant is resting comfortably. The approach to the physical examination depends on the age of the patient. It is often helpful to examine the infant while the mother or father holds the baby to provide a familiar, warm, and comforting environment. Every effort should be made not to disturb the infant because crying complicates auscultation of the heart.

The vital signs, including blood pressure, should be recorded. Height, weight, and head circumference should be carefully measured. On examination of the chest, any deformity of the left hemithorax, indicating long-standing cardiac enlargement, should be noted. The chest is palpated for precordial activity and to determine whether a thrill is present. A palpable thrill is particularly important, as it indicates significant underlying cardiac disease. A finger placed in the suprasternal notch is used to feel for the thrill associated with left ventricular outflow tract obstruction. It is imperative to carefully feel the pulses in all four extremities and to simultaneously palpate the upper and lower extremity pulses. It is sometimes easier to feel the pedal pulses than the femoral pulses in infants. Diminished intensity of the lower extremity pulses provides evidence of coarctation of the aorta. Bounding pulses suggest patent ductus arteriosus or aortic regurgitation. Cyanosis may be obvious or may require close scrutiny for detection. Inspection of the nail beds is more reliable when performed using natural light. The fingers should be carefully inspected from above and from the side for evidence of clubbing.

Auscultation should be performed in a consistent, systematic fashion. Patience is required with infants and toddlers. The entire precordium should be evaluated, as well as both sides of the back and the midaxillary areas bilaterally. The first and second heart sounds should be characterized and the presence or absence of additional sounds should be noted. A single second sound often is found in pulmonary valve atresia and tetralogy of Fallot. An excessively loud second heart sound signifies pulmonary hypertension or transposition of the great arteries. When a loud murmur is present, it should be characterized with regard to the type of murmur, intensity, location, and transmission. Many of the common defects produce characteristic murmurs. For example, a harsh holosystolic murmur loudest at the lower left sternal border is typical for a ventricular septal defect. An ejection systolic murmur at the upper left sternal border with radiation into the lungs and a preceding ejection click is diagnostic of pulmonary valve stenosis. A pulmonary ejection murmur with a fixed widely split second heart sound is characteristic of an atrial septal defect. Diastolic murmurs are less frequent in infants but when present are especially significant. A diastolic inflow rumble may result from increased pulmonary blood flow from a large ventricular septal defect. The murmur of a patent ductus arteriosus typically has a diastolic component. Less commonly, a diastolic murmur may be due to aortic or pulmonic regurgitation.

Auscultation of the lungs usually is normal in infants and children in the absence of pulmonary infection, even in cases of congestive heart failure due to a large left-to-right shunt. The presence of rales from cardiac failure in children suggests impaired left ventricular function with elevated pulmonary venous and capillary pressures. Characteristically, no abnormalities are found in the lungs with right-to-left shunts producing cyanosis. The hallmark of congestive failure in children is hepatic enlargement, developing and regressing rapidly as the degree of circulatory failure changes in response to therapy. Estimation of the presence and extent of hepatic enlargement is of particular importance. In contrast to adult forms of cardiac disease, hepatic enlargement occurs more commonly than the rales or peripheral edema found in adults with heart disease. If an infant or child presents with peripheral edema, it is more likely to be a result of renal disease (e.g., nephrotic syndrome) than to a structural cardiac defect.

Diagnostic Tests

The use of diagnostic tests should be goal-directed, based on the history and physical examination. The chest radiograph, electrocardiogram, and echocardiogram are not used as screening tests because of the lack of sensitivity and specificity (chest radiograph and electrocardiogram) or the expense (echocardiogram). Echocardiography has assumed a predominant role in the diagnosis and management of heart disease in infants and children, but the chest radiograph and electrocardiogram continue to be useful.

Chest Radiograph

On the chest radiograph, cardiac position, size, and contour, aortic arch position, and vascularity of the lung fields should be noted. The lungs and pleural spaces should be examined. Bony abnormalities may occur in certain syndromes or may be secondary to previous thoracic operative procedures. In older children with coarctation of the aorta, rib notching caused by collateral flow may be apparent. Cardiac size is best expressed as the cardiothoracic ratio, with a ratio greater than 0.5 (0.6 in infants) suggesting cardiac enlargement. Enlargement of specific cardiac chambers can be estimated, but this is more precisely done with echocardiography. Enlargement of the left atrium occurs with mitral insufficiency, ventricular septal defect, patent ductus arteriosus, or any form of left ventricular failure. Left ventricular enlargement is characteristic of aortic disease, mitral insufficiency, coarctation of the aorta, patent ductus arteriosus, and ventricular septal defect. Right atrial enlargement is especially prominent in Ebstein's malformation and also occurs in tricuspid atresia. Selective enlargement of the right ventricle is seen frequently with atrial septal defect.

Characteristic changes in cardiac contour may be seen in certain malformations. The boot-shaped heart of tetralogy of Fallot results from hypertrophy of the right ventricle in association with a right ventricular outflow tract. The egg-shaped heart of transposition of the great vessels is caused by enlargement of the right ventricle and right atrium, with a narrow mediastinal shadow due to the anterior-posterior relation of the aorta and pulmonary artery. With total anomalous connection of the pulmonary veins to the superior vena cava, a characteristic figure-of-eight or snowman appearance has been described. The size of the pulmonary vessels and the pulmonary vascularity also are important. With left-to-right shunts producing a significant increase in pulmonary flow, the vessels are enlarged with engorgement of the lung fields. The appearance may be strikingly different from conditions with a normal or decreased pulmonary blood flow, such as in tetralogy of Fallot.

Electrocardiogram

The electrocardiogram is the best tool for determining the cardiac rhythm. The normal ECG changes significantly during the first few days and weeks after birth, with continued changes progressing throughout adolescence. It is important to interpret the ECG in an age-appropriate fashion, since what appears to be markedly abnormal for an adult might be completely normal for an infant or child. A considerable amount of information can be gained on the presence and degree of hypertrophy, although the ECG in infants and children is not highly sensitive. Selective hypertrophy of the left ventricle, as in aortic valvular stenosis, or selective hypertrophy of the right ventricle, as in pulmonic valvular stenosis, can be identified and roughly correlated with the degree of stenosis. A large ventricular septal defect generally produces electrocardiographic features of biventricular hypertrophy. An atrial septal defect usually produces evidence of right ventricular enlargement. Evidence of ischemic changes or infarction can be provided by the ECG, especially if serial tracings are available for comparison.

Echocardiogram

Echocardiography is the cornerstone of noninvasive assessment of infants and children with congenital heart disease. Two-dimensional imaging and a thorough Doppler evaluation provide information on the structure and function of the heart. In contrast to adults, infants and children generally have excellent acoustic windows, so that conventional transthoracic imaging provides high-quality diagnostic information. Often views from the subcostal approach are extremely helpful, which usually is not the case in adults. The patient must be still during the examination; sedation is sometimes necessary in infants and young children in order to obtain diagnostic images. Because transthoracic and subcostal imaging generally provides all of the necessary information, transesophageal echocardiography is not required in children nearly as frequently as it is in adults.

Transthoracic echocardiography has emerged as an important tool in the operating room during pediatric cardiac surgery to confirm preoperative diagnoses, monitor ventricular function, and assess the adequacy of the repair. With the newer, smaller probes, transesophageal echocardiography can readily be applied to infants weighing as little as 3 kg. Intraoperative transesophageal echocardiography has become the accepted practice in most large centers providing cardiac surgery for infants and children. Fetal echocardiography is useful for detecting significant cardiac structural defects before birth. This has proved to be helpful in preparing for the immediate management of newborns

with congenital heart disease. Newer techniques of three- and four-dimensional echocardiography are used in selected pediatric patients but have yet to gain widespread application.

An echocardiogram and Doppler study generally can provide information regarding cardiac and great vessel anatomy, including cardiac situs, atrioventricular concordance, ventriculoarterial concordance, ventricular and septal wall thickness, chamber size and configuration, and aortic or pulmonary artery size. Doppler evaluations are helpful in detecting and quantitating intracardiac or great vessel shunts and the presence and severity of atrioventricular and semilunar valve regurgitation or stenosis. The pressure gradient across a stenotic valve or coarctation correlates well with pressure measurement obtained at catheterization. The magnitude of a left-to-right shunt can be estimated by a variety of techniques. In the presence of a ventricular septal defect or patent ductus arteriosus, estimates of systolic right ventricular and pulmonary arterial pressure are relatively straightforward. Many children have a small amount of tricuspid valve regurgitation, often inaudible but detectable by Doppler, which allows convenient estimation of right ventricular pressure.

Because of the reliability, accuracy, and ease of modern echocardiography, many infants and children are treated medically or surgically without the need for invasive studies (e.g., cardiac catheterization). Patients with straightforward obstructive lesions, atrial septal defect, ventricular defect, or patent ductus arteriosus are often in this category. Even infants with complex defects such as transposition of the great arteries, atrioventricular septal defect, hypoplastic right heart syndrome, truncus arteriosus, and interrupted aortic arch usually are taken to the operating room without undergoing cardiac catheterization. MRI may be helpful in defining the anatomy, especially extracardiac features (aortic arch, coarctation, pulmonary arteries, systemic and pulmonary venous connections).

Cardiac Catheterization

Pediatric cardiac catheterization is becoming more of an interventional procedure than a traditional diagnostic procedure. In most centers, the proportion of diagnostic catheterizations is declining as the number of interventional procedures increases. Cardiac catheterization is increasingly used in children for the diagnosis and treatment (e.g., radiofrequency catheter ablation) of simple and complex arrhythmias. Many common malformations can be treated safely and effectively in the catheterization laboratory. Balloon dilation generally is the preferred treatment for uncomplicated isolated valvular pulmonic stenosis. Patent ductus arteriosus usually is treated with coil occlusion of the ductus in the catheterization laboratory. Although not yet approved for general use, catheterization techniques likely will be used in the near future in closing most secundum atrial septal defects. Balloon dilation of stenotic pulmonary arteries (including placement of stents in the pulmonary arteries) plays an important role in the comprehensive management of patients with complex lesions. The role of interventional techniques continues to emerge for more complicated cases such as aortic stenosis, mitral stenosis, and coarctation of the aorta. Coordination and planning of complementary catheterization- based and surgical approaches is necessary to provide optimal care for infants and children with heart disease.

Diagnostic cardiac catheterization may be helpful when, for example, precise measurement of pulmonary vascular resistance is necessary. During cardiac catheterization, intracardiac pressures and pressure gradients from obstructive lesions can be determined, shunt magnitudes can be calculated, ventricular or vascular morphology can be visualized, and vascular resistances can be calculated. Diagnostic catheterization should be considered in complicated postoperative patients, such as infants being prepared for stage 2 and 3 reconstruction for hypoplastic left heart syndrome, in whom uncertainty exists regarding the cardiac status.

In the normal heart, depending on the age of the child, the right atrial systolic pressure is less than 5 mmHg, the right ventricular systolic pressure ranges from 15 to 30 mmHg, the left atrial pressure and the left ventricular diastolic pressure each range from 5 to 10 mmHg, and the left ventricular systolic pressure ranges from 80 to 120 mmHg. Accurate quantification of intracardiac shunts is based on the Fick principle. Normally pulmonary blood flow (Q_p) equals systemic blood flow (Q_s) and the Q_p:Q_s ratio is 1. In patients with left-to-right shunts the Q_p:Q_s ratio is increased.

Q_p/Q_s =Aortic oxygen sat. - Mixed venous oxygen sat. / Pulmonary venous oxygen sat. - Pulmonary arterial oxygen sat.

A Q_p:Q_s ratio of 1.5 or less is associated with mild physiologic disturbances and may not require operative intervention. A Q_p:Q_s ratio of 1.5 to 1.8 is on the borderline, and usually a Q_p:Q_s ratio greater than 1.8 to 2.0 is an indication for operative correction. It is important to consider all information, especially the pulmonary vascular resistance, and not just the pulmonary artery pressure or Q_p:Q_s ratio alone, when considering a patient for operative intervention.

The physiologic evaluation of the degree of valvular stenosis is obtained by calculating the functional cross-sectional area of the stenotic valve orifice. This is done by measuring the flow rate and pressure differential across the involved orifice. A valvular diameter more than 50 to 75 percent below the predicted normal for the child's body surface area is markedly abnormal. For example, a normal cross-sectional aortic valve area is 2 cm²/m² body surface area, while an area less than 0.5 cm²/m² represents severe aortic stenosis. Similar calculations are possible for the mitral valve, which has a normal area of 2.5 to 3.0 cm²/m². Significant symptoms are noted with a mitral valve diastolic gradient greater than 10 mmHg and with a mitral valve area of less than 0.8 cm²/m².

PRINCIPLES OF PREOPERATIVE, INTRAOPERATIVE, AND POSTOPERATIVE CARE

Preoperative Management

In cases of newborn infants with life-threatening congenital heart disease the initial diagnosis usually is made in the referring hospital, usually by echocardiography. Immediate initiation of proper preoperative therapy in the referring institution and careful transport are essential for optimal results. Close attention should be given to oxygenation, acid-base status, fluids, body temperature and respiratory mechanics before and during transport. Metabolic acidosis should be corrected immediately in patients with severely restricted systemic or pulmonary blood flow. Patients who are dependent on a patent ductus arteriosus should be started on an intravenous infusion of prostaglandin E₁ (0.05mg/kg/min). Treatment with digitalis, diuretics, or inotropic agents also is frequently necessary. If the infant is hypoventilating or severely tachypneic with respiratory distress, mechanical ventilation should be initiated. Close attention should also be paid to body temperature, oxygen saturation, and respiratory mechanics during transport.

Once the seriously ill infant arrives in the neonatal or pediatric cardiac intensive care unit, vigorous critical care management is initiated. It is no longer advisable to rush the patient to the operating room early in unstable condition. Instead, correction of the underlying heart failure and metabolic defects is carried out over 2 to 3 days, improving perfusion, correcting acidosis, and stabilizing hemodynamics before surgery. This approach allows the child to arrive in the operating room in stable metabolic and hemodynamic condition, which significantly lowers the operative risk.

Intraoperative Management

On arrival in the operating room general anesthesia is induced and central venous access and arterial pressure monitoring are secured. A percutaneous, double-lumen internal jugular central venous catheter usually is inserted. Secondary sites for central venous access are the subclavian or femoral veins. Similarly, a percutaneous radial artery line is inserted for arterial monitoring if feasible. Otherwise, a radial, brachial, or femoral artery cutdown is performed.

The patient is placed in the supine position with a roll or sandbag between the shoulders. A small Foley catheter is placed into the bladder and rectal and tympanic membrane temperature probes are placed. Close attention must be given to the patient's temperature, fluid status, and oxygenation. Intraoperative transesophageal echocardiography usually is performed to confirm the diagnosis and to rule out any unsuspected lesions.

After the preparation and sterile draping of the patient, a sternotomy incision is made. Special pediatric cannulas are available for arterial and separate vena caval cannulation. Cardiopulmonary bypass is established with moderate (22 to 25°C) or profound (15 to 17°C) hypothermia. The heart usually is arrested with cardioplegia for myocardial protection, and the intracardiac repair is performed.

A major advance responsible for the excellent results obtainable with total correction of congenital heart problems is the frequent use of profound (deep) systemic hypothermia during cardiopulmonary bypass. Lower temperatures give a measure of end organ protection and allow perfusion with lower flow rates. In patients with complicated anatomy, deep hypothermia and total circulatory arrest may be used, allowing the surgeon excellent visualization to perform a precise, complex intracardiac repair. The techniques used are based on the principle that the energy requirements and oxygen consumption of the body are reduced relative to decreases in body temperature. For example, at normothermia, systemic flow rates during cardiopulmonary bypass must approach the normal cardiac index of 2.5 to 3.5 L/min/m² to supply the oxygen demands of the body. At 25°C total-body oxygen consumption drops by more than 50 percent and pump flow rates of 1.5 to 1.7 L/min/m² are adequate for tissue and organ perfusion. More strikingly, at 15°C oxygen consumption is less than 25 percent of normal. At this temperature low pump flow rates of 1.0 to 1.5 L/min/m² are adequate, and total cessation of flow (circulatory arrest) is safe for 45 to 60 minutes. The use of low systemic flow rates or total circulatory arrest has allowed safer, more precise intracardiac correction of complex abnormalities in small infants or neonates with a still, dry, bloodless field.

After the cardiac defect is repaired, the child is rewarmed. Before discontinuing bypass, additional central venous, left atrial, or pulmonary artery pressure monitoring catheters are placed as necessary. Temporary pacemaker wires are routinely placed on the right atrium and right ventricle, often placing two atrial and two ventricular pacemaker wires if atrioventricular pacing is required. Inotropic agents are used as necessary for impaired cardiac function. Once all monitoring is in place and the child is warm and ventilated, bypass is discontinued.

The hemodilution and systemic inflammatory effects of cardiopulmonary bypass have been found to result in significant postoperative total-body volume overload because of increased capillary leakage. This may lead to transient dysfunction of the heart, lungs, and kidneys. Several centers have advocated the use of ultrafiltration to prevent massive postoperative volume overload, minimize edema, and hasten myocardial functional recovery. Ultrafiltration during or immediately after cardiopulmonary bypass removes excess water, resulting in a higher postoperative hematocrit level, decreased bleeding, fewer transfusion requirements, and improved pulmonary function. Modified ultrafiltration removes excess fluid and activates complement and cytokines, resulting in diminished neutrophil activation with less inflammatory mediated endothelial cell injury or capillary damage. Aggressive early postoperative ultrafiltration improves cardiac performance, improves pulmonary compliance and alveolar oxygen transport, and diminishes the risk of multiorgan injury.

Postoperative Care

Important principles in postoperative pediatric cardiac intensive care are constant observation, monitoring of the electrocardiogram, continuous measurement of central venous, left atrial, and arterial pressure, continuous monitoring of systemic oxygen saturation with a pulse oximeter, repeated measurement of arterial and mixed venous blood gases, monitoring of fluids, electrolytes, and urine output, and management of mechanical ventilation and respiratory therapy (Fig. 17-1).

The child's body temperature should be maintained and a nasogastric tube inserted to prevent gastric distention. Clinical changes occur very rapidly in small children and neonates; reaction times must be rapid and the treatment precise if optimal results are to be achieved. Postoperative care requires close communication between the surgeon, the pediatric cardiologist, the intensivists, and the critical care nurses.

Four key areas of postoperative management are: (1) hemodynamics, (2) fluids, electrolytes, and renal function, (3) arrhythmias, and (4) mechanical ventilation—respiratory problems.

Hemodynamics

Hemodynamic monitoring of the postoperative patient requires continuous measurement of preload (central venous and left atrial pressures), systemic arterial pressure, systemic oxygen saturation, mixed venous oxygen saturation, and urine output. Physical examination also is helpful in determining the adequacy of perfusion, which is reflected in the temperature of the extremities. Cool extremities reflect a modest reduction in cardiac output, whereas cold feet and legs often indicate a severe depression in cardiac output. The mixed venous oxygen saturation reflects the level of peripheral oxygen extraction, indirectly reflecting the cardiac output. A mixed venous oxygen saturation of less than 40 percent and partial pressure of oxygen less than 25 mmHg suggests a significantly depressed cardiac output. The cardiac output can be measured directly in selected patients using a thermistor and thermodilution techniques.

Cardiac preload and intravascular volume status should be assessed by monitoring of central venous pressure or left atrial pressure as indicated. These measurements reflect the loading conditions of the ventricles. Separate monitoring of various chambers often is necessary because the left-sided cardiac function and the right-sided cardiac function may respond differently after repair of complicated congenital lesions. Attempts should be made to optimize the cardiac filling pressures. If the preload is decreased and intracardiac pressures are low, blood transfusion is recommended if the patient's hematocrit level is less than 40 percent and fresh frozen plasma if the hematocrit level is greater than 40 percent, continuing transfusion until the filling pressures are optimized. If the intracardiac pressures are elevated and the cardiac output is depressed, this usually reflects a problem with cardiac contractility or cardiac tamponade.

A common cause of decreased preload and hemodynamic instability in the early postoperative period is bleeding. Often a coagulopathy exists after cardiac surgery in infants, frequently because of the combination of hypothermia, hemodilution, and platelet dysfunction. It is not uncommon for the prothrombin time to be elevated to 15 to 20 s immediately postoperatively, with abnormal clot formation. This demands immediate correction with fresh frozen plasma and platelets. If bleeding persists after clotting studies normalize, a surgical source must be suspected and consideration must be given to surgical reexploration.

Cardiac tamponade must be suspected whenever hemodynamics deteriorate in a postoperative patient. The classic signs of tamponade are a low systemic blood pressure, a pulsus paradoxus, a high central venous pressure, and a widened mediastinal shadow on the chest radiograph. In many patients the classic signs are not present, and tamponade should be suspected and ruled out whenever signs of a low cardiac output are present.

If decreased preload and tamponade are not present, low cardiac output is most likely a result of depressed myocardial contractility, which may occur in any patient after intracardiac repair. Common causes of depressed contractility are myocardial edema secondary to cardiopulmonary bypass and ischemia-reperfusion, poor myocardial protection, hyperkalemia, hypoxia, drug toxicity, and intrinsic cardiac disease. Mechanical intracardiac defects also must be ruled out. Once obvious correctable causes have been excluded, the patient with poor cardiac contractility should be started on inotropic support. The most commonly used inotropic agents are dopamine (5 to 20 mg/kg/min), dobutamine (5 to 20 mg/kg/min), epinephrine (0.05 to 0.5 mg/kg/min), and norepinephrine (0.05 to 0.5 mg/kg/min). Phosphodiesterase inhibitors such as amrinone and milrinone have also been used, but these agents probably work mostly through vasodilation and a reduction in afterload because the myocardial phosphodiesterase—cyclic AMP system may be poorly developed in neonates. Inotropic agents should always be given through a central venous line.

Fluids, Electrolytes, and Renal Function

Careful monitoring of fluid intake and output is essential in critically ill postoperative patients. Fluid and blood administration must be balanced with urine output, blood losses, and insensible losses. The intraoperative procedure often is the initial cause of volume overload, secondary to hemodilution during cardiopulmonary bypass. The combination of cardiopulmonary bypass and hypothermia also often induces some degree of renal dysfunction, which limits the patient's ability to clear fluids postoperatively. Most neonates or infants undergoing complex intracardiac repair or hypothermic circulatory arrest are started on low-dose dopamine (3 to 5 mg/kg/min) immediately postoperatively to enhance renal blood flow.

The maintenance fluids are usually given with D5/NS, restricting sodium and volume intake. In small neonates, D10/NS is frequently used to maintain adequate serum glucose. Total fluids are restricted over the first 48 h. Normal maintenance fluid requirements are calculated as 100 mL/kg/24 h for the first 10 kg, plus 50 mL/kg/24 h for the second 10 kg, plus 20 mL/kg/24 h for weight above 20 kg. After operation, children are given 50 to 75 percent of maintenance calculations. Potassium is seldom replaced in the early postoperative phase unless the serum potassium level is less than 3.0 mEq/L in the presence of good cardiac output and urine output. When necessary, potassium chloride is given in one to two doses of 0.1 mEq/kg over 30 min each, after which the serum potassium level is remeasured.

Urine output is a very good indicator of organ perfusion, as low cardiac output often results in an immediate drop in urine output in infants. Some degree of renal dysfunction often occurs secondary to cardiopulmonary bypass. Most patients are started on renal doses of dopamine immediately after operation. Patients with poorly perfused, cool extremities, a low mixed venous oxygen saturation, and a dropping urine output should be started on inotropic support. Diuretics (furosemide 0.2 to 2 mg/kg) also are given to increase the urine output. An hourly urine output of less than 0.5 mL/kg should be of great concern, because this often reflects low cardiac output syndrome, impending renal dysfunction, or both. Because the problems associated with volume overload, myocardial edema, and hyperkalemia are often fatal, aggressive therapy with inotropic agents and diuretics is critical. If the urine output remains below 0.5 mL/kg/h and the patient does not respond within 1 to 3 hours, early peritoneal dialysis is recommended.

Arrhythmias

Cardiac arrhythmias after congenital cardiac surgery are not uncommon. The diagnosis is made by the standard 12-lead electrocardiogram, or by an atrial electrocardiogram where the atrial pacing wire is connected to the chest lead. The ECG is assessed for heart rate, rhythm, the presence of P waves, atrioventricular (A-V) conduction (or dissociation), and QRS morphology. In some patients, a transient bolus of adenosine aids in treatment and diagnosis. All serious arrhythmias should be treated promptly in consultation with the pediatric cardiologist and intensivist, because most postoperative arrhythmias are associated with a drop in cardiac output with progressive hemodynamic compromise.

Postoperative arrhythmias may be classified as bradyarrhythmias—either sinus node dysfunction or third degree atrioventricular block; and as tachyarrhythmias—either primary atrial tachycardia, atrioventricular reentry tachycardia, His bundle or junctional ectopic tachycardia, ventricular tachycardia, or ventricular fibrillation.

The postoperative *bradycardias* have two primary causes. The first, *sinus node dysfunction*, often results from injury to the sinoatrial (S-A) node. Patients with S-A node dysfunction usually develop ectopic or absent atrial activity with junctional or ventricular escape rhythms. This rhythm often can be overridden and treated by atrial pacing. Postoperative bradycardia also may occur secondary to *third-degree heart block*, which may develop secondary to surgical trauma or intrinsic disease. Patients with third-degree heart block and A-V dissociation require A-V sequential pacing. These patients should not be discharged from the hospital until the heart block resolves or a permanent pacemaker is placed.

Atrial tachycardia may arise from an *ectopic atrial focus* (with an abnormal P wave) or from *atrial flutter (atrial reentry tachycardia)*. Intravenous adenosine (100 to 300 mg/kg) may help establish the diagnosis and provide initial treatment. These arrhythmias usually are treated with drugs to slow A-V conduction, such as digoxin or verapamil, or by Class I or Class II antiarrhythmic agents.

Atrioventricular reentry tachycardia usually involves accessory pathways as seen in Wolf-Parkinson-White syndrome. A-V reentry tachycardia may require cardioversion and should be treated with Class I or Class II antiarrhythmic agents.

An increasingly recognized tachyarrhythmia after congenital heart surgery is *His bundle automatic tachycardia* or *junctional ectopic tachycardia (JET)*. This arrhythmia is characterized by a rapid heart rate (180 to 220) and A-V dissociation, with a QRS morphology that is similar to that noted during sinus rhythm. While JET frequently lasts for only 72 to 96 h, it is unfortunately associated with a high mortality rate that is a consequence of progressive cardiac dysfunction unless it is promptly treated and converted. Treatment involves sedation to decrease endogenous catecholamines, minimizing the use of exogenous catecholamines, neuromuscular paralysis, systemic cooling to 32 to 35°C and intravenous antiarrhythmia therapy with amiodarone. Amiodarone is a Class III agent that slows depolarization and repolarization. The loading dose is 5 mg/kg over 1 h, with a maintenance dose of 5 to 7 mg/kg/h.

Ventricular tachycardia or *ventricular fibrillation* may appear without warning but can be corrected if therapy can be started within 1 to 3 min. Immediate cardiopulmonary resuscitation (CPR) with closed-chest massage should be initiated. These patients should be treated with electric cardioversion (2 watt-sec/kg) and with intravenous Class I (lidocaine, procainamide) or Class III (bretylum, amiodarone) agents. Electrolyte abnormalities should always be ruled out by checking for potassium, magnesium, phosphorous, and other abnormalities.

Mechanical Ventilation—Respiratory Problems

Excellent respiratory and ventilator management is essential in the postoperative care of patients after congenital heart surgery. Many of these patients present with increased pulmonary blood flow preoperatively, congestive heart failure, and pulmonary edema. Respiratory management may be further complicated by the reactive pulmonary vasculature and reactive small airways found in neonates and infants. The secondary effects of cardiopulmonary bypass result in pulmonary capillary leakage with decreased lung compliance. While early extubation is recommended frequently in patients at low risk, most neonates and infants with complex lesions are ventilated for at least 24 h postoperatively.

The child usually is intubated with a nasotracheal tube, which should be properly secured. Because the tracheobronchial passages and the endotracheal tube are small, secretions are difficult to remove and dislodgment can occur easily. Meticulous suctioning is required. One of the most common causes of cardiac arrest or bradycardia postoperatively is acute hypoxia and respiratory acidosis from a plugged or dislodged endotracheal tube. The critical care team should be acutely aware of this possibility, and if the ventilated child develops acute hypoxia or bradycardia the patient should be disconnected from the ventilator, attached to a breathing bag with 100% oxygen, and suctioned. If breath sounds are not subsequently heard, strong consideration should be given to changing the endotracheal tube. Other mechanical considerations such as atelectasis or tension pneumothorax must be ruled out.

Mechanical ventilation is performed by delivering a tidal volume of approximately 10 mL/kg. Barotrauma is of great concern in children with immature or noncompliant lungs, and care is taken not to exceed a peak inspiratory pressure of 30 to 35 mmHg, because this can result in overdistention of the alveoli and cause pulmonary injury. Conversely, children in the first year of life are dependent on the distending ventilation pressure to keep the alveoli open. Lung compliance is poor postoperatively, and there is a strong tendency toward atelectasis. This may be prevented by positive end-expiratory pressure (PEEP), usually at 3 to 6 mmHg. Other than tidal volume and inspiratory pressure, the respiratory rate and inspired oxygen concentration are variables that can be manipulated.

Serial measurement of arterial blood-gas tensions is essential to assess oxygenation and adequacy of ventilation. Continuous monitoring of oxygen saturation may also be used,

along with monitoring of end tidal carbon dioxide tension. The desired arterial oxygen tension varies with the cardiac physiology and the degree of residual shunting. The appropriate level for a particular patient should be determined, and prolonged high levels of inspired oxygen should be avoided, because this results in oxygen toxicity with damage to the type II pneumocytes. Elevated arterial P_{CO_2} exceeding 40 to 45 mmHg occur from hypoventilation, resulting in acidosis and pulmonary vasoconstriction. In some patients this is desirable, but for most patients the tidal volume and respiratory rate should be adjusted to obtain a normal Pa_{CO_2} .

In patients with incompletely corrected physiology and continued bidirectional shunting, pulmonary and systemic blood flows are determined by the relative resistances of the pulmonary and systemic vascular beds. The extreme example of this is in the management of the patient with hypoplastic left heart syndrome who has undergone a stage 1 Norwood operation (outflow correction by conversion to single-ventricle physiology, with a central-systemic-to-pulmonary shunt for pulmonary blood flow—see section below). If the systemic oxygen saturation is too high and the Pa_{CO_2} is low, pulmonary blood flow is increased and the lungs become flooded at the expense of a decreased systemic cardiac output. Ideally, these patients should have balanced pulmonary and systemic flows, which is best achieved with a systemic oxygen saturation of 75 to 85 percent ($Pa_{O_2} = 40$ mmHg) and a Pa_{CO_2} of approximately 40 to 45 mmHg. Extra in-line carbon dioxide may be added to the ventilator circuit to achieve this goal without hypoventilating the patient. The systemic and mixed venous oxygen saturations are sequentially monitored to evaluate the pulmonary (Q_p) versus the systemic (Q_s) blood flows.

Pulmonary hypertensive crisis is a severe complication that is sometimes seen postoperatively in patients with reactive pulmonary vascular disease. This typically results in a suprasystemic pulmonary artery pressure, with a low systemic arterial blood pressure. While the central venous pressure and pulmonary arterial pressure are elevated, the left atrial pressure remains low because blood flow through the lungs is diminished. Patients at risk for pulmonary hypertensive crisis are those with large left-to-right shunts, such as patients with atrioventricular septal defect, transposition of the great vessels with ventricular septal defect, truncus arteriosus, total anomalous pulmonary venous return, or a large ventricular septal defect. These patients usually have preoperative pulmonary hypertension and an increased pulmonary vascular resistance secondary to medial hypertrophy in the pulmonary arterioles. The best treatment is prevention. Care should be taken to avoid hypoventilation or acidosis, maintaining a Pa_{O_2} above 85 to 90 mmHg and a Pa_{CO_2} of 34 to 38 mmHg, with a slight respiratory alkalosis. Hypoxia and acidosis from any cause are potent pulmonary vasoconstrictors. Patients with pulmonary hypertension and a high pulmonary vascular resistance usually are paralyzed and sedated with fentanyl for 24 to 48 h to minimize endogenous catecholamine production. Careful suctioning is used, because stimulation of the airway or hypoxia during suctioning can initiate a crisis. If possible, exogenous catecholamines and alpha-adrenergic agonists are avoided. Medications used to prevent or treat severe pulmonary hypertension are isoproterenol, nitroglycerine, nitroprusside, tolazoline, prostacyclin, amrinone or milrinone, and inhaled nitric oxide. Inhaled nitric oxide is a potent pulmonary vasodilatory that has been found to be highly efficacious in many patients.

OBSTRUCTIVE LEFT-SIDED LESIONS

Coarctation of the Aorta

Coarctation is a common congenital malformation, occurring in 10 to 15 percent of patients with congenital heart disease. It is more common in males (3:1 ratio). Coarctation of the aorta often represents a spectrum of disease, from isolated obstruction adjacent to the ligamentum arteriosus to diffuse hypoplasia extending proximally into the aortic arch. Classically, preductal or “infantile type” coarctation represents a diffuse narrowing of the aorta between the subclavian artery and the ductus arteriosus, usually with ductus-dependent blood flow distally. These children present with acute heart failure as newborns. Postductal or “adult type” coarctation involves a more localized narrowing at the site of insertion of the ligamentum arteriosus. Patients with localized coarctation are often asymptomatic, presenting with hypertension later in childhood. The extent and severity of coarctation varies widely, from diffuse aortic arch hypoplasia to a mild isolated aortic stenosis.

The ligamentum arteriosus is attached to the medial surface of the aorta near the site of coarctation in older patients. The stenotic area may have two or three component parts. The most frequent is a localized shelf, consisting of an infolding of the aortic media into the lumen. This is most visible on the aortic wall opposite the ligamentum arteriosum. In the lumen, a thickened ridge of intima may be present and may increase the severity of the stenosis.

Distal to the coarctation, the aorta usually is dilated. In adults, a true aneurysm forms in a small proportion of patients. Large, dilated intercostal arteries entering the distal aorta, providing collateral circulation around the site of obstruction, are a striking feature. In older patients, these large arteries produce “notching” of the ribs that is visible on the chest radiograph. Rarely, they may become aneurysmal and rupture.

Patients with infantile coarctation present early in life. In these patients a varying degree of “tubular hypoplasia” is common, consisting of a narrowing of the aorta between the coarctation and the left subclavian artery, often extending proximally to involve the distal aortic arch. The condition usually is fatal unless treated and is associated with other cardiac defects, such as ventricular septal defect, bicuspid aortic valve, and mitral valve anomalies in 75 percent of the cases. In this condition, a large patent ductus perfuses the distal aorta with blood from the pulmonary artery. The coarctation is located proximal to this. The association of coarctation of the aorta, subaortic stenosis, parachute mitral valve, and supraaortic left atrial ring is known as Shone's complex.

Pathophysiology and Clinical Manifestations

Neonates usually present with severe congestive heart failure, which is exacerbated by the presence of a ventricular septal defect or by an additional valvular lesion. If the coarctation is severe, closure of the ductus arteriosus precipitates extreme acidosis and renal shutdown as well as pulmonary congestion. These patients must be vigorously resuscitated, often requiring intubation, prostaglandin infusion, intravenous dopamine, intravenous furosemide, and correction of acidosis before urgent operation.

After the first year of life, congestive heart failure rarely occurs before the age of twenty. Hypertension, however, becomes a significant concern. Children in their early teens often have significant hypertension with changes in the aortic wall, including aneurysm formation. Without treatment the average life expectancy is only 30 to 40 years. The four most common causes of death in patients who do not undergo surgical treatment are rupture of the aorta, cardiac failure, rupture of intracranial aneurysms, and bacterial endocarditis.

The diagnosis often is made by a routine school physical examination uncovering hypertension. Headache, epistaxis, and leg fatigue are the most frequent symptoms. Claudication in the lower extremities is uncommon. Before surgery these patients should be treated with beta-adrenergic blockers to lessen ventricular strain, lower the hypertension, and minimize the risk of postoperative rebound hypertension.

The classic combination of hypertension in the upper extremities with absent or decreased pulses in the lower extremities of a child immediately suggests coarctation. If weak femoral pulsations are present, direct measurement of the blood pressure in the upper and lower extremities may be necessary to confirm the diagnosis. Prominent pulsations from collateral circulation may be visible in the neck and over the muscles of the shoulder girdle. A systolic murmur usually is audible over the left hemithorax.

In neonatal coarctation the diagnosis is accurately established by echocardiography, and emergency repair usually is done on the basis of this test alone. In older patients the chest radiograph may help to establish the diagnosis by demonstrating bilateral notching of the ribs posteriorly. Notching is unusual in patients under 6 years of age but is almost always present by age fourteen. The electrocardiogram characteristically shows signs of left ventricular hypertrophy and often left ventricular strain. In most patients the diagnosis can be made accurately from the clinical findings in combination with the echocardiogram. MRI or enhanced magnetic resonance angiography can establish the diagnosis with certainty and outline the anatomy preoperatively.

Operative Indications

Neonates and infants are treated urgently, at the time of presentation. Usually they are stabilized for 24 to 48 h in the intensive care unit to minimize congestive heart failure, correct acidosis, and improve renal perfusion. Concomitant defects may be treated at the time of coarctation repair or staged for treatment at a later time, depending on the severity. In older children the ideal age for operation is between 3 and 4 years. Children with mild to moderate coarctation are treated whenever proximal hypertension or significant left ventricular hypertrophy begins to develop. Although balloon dilation has been used successfully to treat recurrent coarctation, this technique is less frequently used for primary coarctation repair, and operative correction usually is the treatment of choice.

Operative Technique

A left posterolateral thoracotomy in the fourth intercostal space is used. The coarctation usually is readily seen, with the typical medial indentation at the site of insertion of the ligamentum arteriosum and large, tortuous intercostal arteries entering the distal aorta. The mediastinal pleura is incised, after which the vagus nerve is retracted medially, noting the course of the recurrent laryngeal nerve encircling the ligamentum arteriosus. The aortic arch proximal to the left subclavian artery, the left subclavian artery, the ligamentum arteriosum, and the distal aorta are serially mobilized. The intercostal arteries should be isolated and preserved, but they may be divided if necessary.

Once the vessels have been adequately mobilized, the proximal aorta, the left subclavian artery, and the distal aorta are occluded with vascular clamps, after which the coarctation is repaired. Repair is performed with any one of three techniques: (1) subclavian flap arterioplasty (Fig. 17-2); (2) resection and end-to-end anastomosis using absorbable vascular sutures to allow for growth (Fig. 17-3); or (3) wide resection with beveled hemiarch anastomosis (Fig. 17-4). The technique used depends on the particular anatomy and the length of coarctation; the beveled hemiarch technique is used for patients with long tubular coarctation and diffuse distal arch hypoplasia.

Once the aorta is clamped, the distal aortic pressure may be measured to determine the adequacy of flow through collateral channels. This is less important in neonates but is critical in older children with moderate coarctation. A distal aortic pressure of more than 50 to 55 mmHg suggests adequate collateral flow. In this situation, the length of the cross-clamp time becomes less critical because the collaterals provide adequate perfusion of the distal spinal cord. If the distal aortic pressure is less than 45 to 50 mmHg, the child is at increased risk for spinal cord ischemia if the cross-clamp time exceeds 20 min. In these patients the operation must be planned and executed in an expedient manner or some form of distal shunt must be used.

The anastomosis usually is done with continuous absorbable monofilament suture. Some surgeons have advocated placing interrupted sutures in the anterior wall of the anastomosis to avoid a “purse-string” effect and to allow for further growth, but we have found this unnecessary in most cases. After completion of the anastomosis and removal of vascular clamps, the blood pressure should be measured proximal and distal to the anastomosis to confirm that no significant gradient remains. If repair is done for neonatal coarctation in association with ventricular septal defect, pulmonary artery pressure is measured after repair of the coarctation. If the pulmonary artery pressure remains elevated, immediate pulmonary artery banding is done to restrict pulmonary blood flow or, alternatively, formal repair of the ventricular septal defect is performed in 7 to 10 days.

In adults, extensive degenerative changes in the aorta, calcification, and fibrosis make insertion of a prosthetic graft necessary in most patients, because direct anastomosis cannot be done. In such instances the coarctation may be excised and an end-to-end interposition graft or an end-to-side graft from the distal arch and subclavian artery to the aorta beyond the site of coarctation, bypassing the coarctation site, may be used.

Results

Repair of isolated coarctation has an operative mortality of 1 percent or less, approaching zero in older children. Repair of neonatal coarctation in children with associated cardiac

lesions is more difficult, with a mortality of 10 to 15 percent, depending on the severity of the associated conditions. Complications include paraplegia (0 to 0.5 percent), renal failure, bleeding, paradoxical postoperative hypertension with intestinal ischemia, chylothorax, and late reoperation for recurrent coarctation.

Paradoxical hypertension typically occurs in the first 48 to 72 h after operation, usually in older patients with severe, long-standing coarctation. These patients almost always have had severe hypertension preoperatively. The syndrome is related to an increase in arterial pressure in visceral arteries, previously functioning with a lower mean pressure. These patients may develop severe abdominal pain and persistent, severe hypertension postoperatively. In extreme cases intestinal necrosis can occur. This risk of this complication is lessened by starting beta-adrenergic blockers preoperatively and continuing beta blockers and sodium nitroprusside postoperatively to tightly control hypertension, which usually eliminates this complication.

Some residual, persistent hypertension is common in patients operated on after 5 years of age and increases with age. This is related to an up-regulation of the renin-angiotensin system. Residual hypertension occurs much less frequently in patients operated on before 3 to 4 years of age.

Regardless of the technique used, recurrent coarctation occurs in a small number of patients within 10 to 15 years. The incidence ranges from 2 to 15 percent, depending on the patient's age and the severity of the coarctation at the time of initial repair. Balloon dilation of the recurring coarctation is the initial treatment of choice for most patients, although reoperation is feasible.

Interrupted Aortic Arch

Interrupted aortic arch (IAA) is a rare defect. Neonates with this condition present with severe congestive heart failure and acidosis, with hypoperfusion of the lower half of the body. These symptoms begin 2 to 3 days after birth, when the patent ductus arteriosus begins to close. Without surgery death usually occurs within 7 to 10 days. Approximately 30 percent of patients with Type B IAA have DiGeorge's syndrome, manifested by an absence of thymic tissue, hypocalcemia, and immunologic abnormalities.

Pathologic Anatomy and Physiology

IAA is classified by the anatomic site of aortic interruption. In *Type A IAA*, which occurs in 40 percent of the cases, the interruption is distal to the left subclavian artery, similar in location to that of patients with infantile type coarctation. It is sometimes difficult to differentiate between severe infantile coarctation and Type A IAA, because they are physiologically similar. *Type B IAA*, the most common form (55 percent), results in total interruption of the arch between the left carotid and left subclavian arteries. *Type C IAA* accounts for only 5 percent of the cases, with the interruption occurring proximally between the innominate artery and the left carotid artery.

The majority of patients with IAA have a large *ventricular septal defect (VSD)*. There also is an association between IAA and other left-sided obstructive lesions. These patients often exhibit varying degrees of hypoplasia of the *left ventricular outflow tract (LVOT)* and the ascending aorta. The aortic valve is bicuspid in 40 to 50 percent of patients, the aortic annulus may be moderately to severely hypoplastic, and a significant number of patients with Type B IAA have subaortic obstruction.

Children with IAA are totally dependent on ductal flow for perfusion of the lower half of the body. As long as the ductus arteriosus is open, the condition frequently is not apparent immediately after birth. As the patent ductus begins to close, however, the child develops poor perfusion of the lower body with rapidly progressive metabolic acidosis and renal insufficiency. Because of the impedance to forward blood flow produced by the interrupted aorta, severe pulmonary congestion and heart failure also rapidly develop. The pulmonary congestion is worsened by left-to-right shunting across the ventricular septal defect.

Diagnosis

Prompt diagnosis and treatment are essential; these children rapidly become acidotic and die. It is estimated that approximately 10 percent of the children with IAA die before operation. The diagnosis is almost always made with echocardiography. MRI scanning should be performed to determine the site of arch interruption and to assess the remainder of the ascending aorta and the LVOT. Cardiac catheterization rarely is necessary.

Treatment

Once the diagnosis of IAA is made an immediate infusion of prostaglandin E₁ is begun to maintain ductal patency. Acidosis should be corrected and inotropic agents are often begun to lessen the degree of heart failure and enhance renal perfusion. Intubation and mechanical ventilation frequently are required as well. Once the acidosis is corrected and the urine output improves, urgent operation is performed.

Type A lesions are corrected through the left side of the chest with an incision in the fourth intercostal space, similar to the repair for aortic coarctation. The proximal aortic arch, left subclavian artery, and distal aorta are mobilized widely. Resection and beveled hemiarch repair are performed as for neonatal coarctation with tubular arch hypoplasia. If a ventricular septal defect is present, placement of a pulmonary artery band may be considered, or the child is allowed to recover and an early complete repair is performed. A complete repair through a midline approach may be considered in some patients if the VSD is large.

Most children with IAA Types B and C should undergo early complete repair through a midline sternotomy with deep hypothermia and circulatory arrest. The aorta is mobilized, the coarcted segment is resected, and primary end-to-end anastomosis usually can be performed. The associated ventricular septal defect is repaired through a right atriotomy incision, and other cardiac defects are corrected at this time. When severe LVOT obstruction is present in association with IAA, the subvalvular obstruction may be corrected with septal myotomy and myectomy. Augmentation of the ascending aorta and arch also may be considered in patients with multiple levels of obstruction.

A small proportion of patients have such severe associated LVOT and subvalvular obstruction that they are considered to be within the spectrum of hypoplastic left heart syndrome. This group of patients generally has an aortic annulus less than 4 mm in diameter with severe subvalvular obstruction. In these patients consideration should be given to performing a conversion to a single-ventricle type of physiology and performing a stage 1 Norwood procedure.

Results

The operative mortality for Type A IAA is 5 to 10 percent, depending on the severity of the associated cardiac pathology. Repair of Types B and C IAA with VSD has an operative mortality of 10 to 25 percent. Late survival rate is approximately 60 to 80 percent at 5 years. Reoperation for recurrent obstruction at the repair site may be necessary, and patients with subaortic obstruction may develop progressive LVOT obstruction. Recurrent stenosis at the repair site may be treated with balloon dilation, whereas progressive LVOT obstruction requires reoperation.

Jonas and associates reported the results of a multi-institutional study of the repair of 183 patients with IAA and VSD. Survival was 73 percent at 1 month and 63 percent at 4 years. The risk of death was increased by low birth weight, younger age, Type B anatomy, smaller VSD, and subaortic obstruction. Some reintervention was required for a coexisting obstructive lesion in 23 percent of the patients within 3 years. In patients with multiple levels of left ventricular outflow obstruction, augmentation of the ascending aorta and arch during initial surgery had a positive effect on survival.

Aortic Stenosis (Valvular/Subvalvular/Supravalvular)

Categories

Congenital Aortic Stenosis

Congenital aortic stenosis is a relatively common abnormality, representing 8 to 10 percent of all patients with congenital heart disease. The stenosis may be discrete or it may encompass diffuse parts of the left ventricular outflow tract. Patients with critical neonatal aortic stenosis may have diffuse endocardial fibroelastosis within the left ventricle, which may become true hypoplastic left heart syndrome if severe. Aortic stenosis frequently is associated with coarctation of the aorta, pulmonary stenosis, mitral valve abnormalities, patent ductus arteriosus, and ventricular septal defect.

Aortic stenosis is three to four times more frequent in males than in females. No causative factors are known for valvular stenosis. The variant of supravalvular stenosis appears caused by a genetic linkage, resulting in abnormal elastic tissue. Supravalvular stenosis is known to be associated with peripheral pulmonary stenosis and Williams syndrome, which is characterized by abnormal facies, mental retardation, and hypercalcemia. Subaortic obstruction may be a result of genetically linked hypertrophic cardiomyopathy.

Congenital aortic insufficiency is rare. When found, it is usually secondary to other abnormalities, such as ventricular septal defect, Marfan syndrome or another connective tissue disease, congenital stenosis with secondary insufficiency, or rheumatic disease.

Valvular Stenosis

Approximately 75 percent of patients with congenital valvular stenosis have a bicuspid aortic valve with varying degrees of annular hypoplasia. A bicuspid valve is a common anomaly occurring in nearly 2 percent of the population, usually without significant stenosis early in life. A common anatomic finding is fusion of the right and left cusps, with the undeveloped commissure represented by a median raphe that may or may not extend to the ventricular wall. Thickening of the valve cusp from abnormal tissue is common and may contribute to the stenosis. Mild poststenotic dilatation of the ascending aorta is common.

In infants with severe valvular stenosis, more severe deformities are common, such as a single-cusped valve, a small annulus with annular hypoplasia, and diffuse ventricular fibroelastosis. At the severe extreme, the left ventricle may be so underdeveloped as to become nonsalvageable, and the patient is considered to have hypoplastic left heart syndrome.

Subvalvular stenosis is rare. The pathology ranges from a discrete fibrous ring with localized obstruction to diffuse fibromuscular obstruction. Gradations exist between these two extremes.

Discrete ringlike stenosis is present in about one-half of patients with subvalvular obstruction. In these patients the proximal muscular outflow tract may be secondarily narrowed from muscular hypertrophy. The stenotic ring usually is connected to the base of the aortic cusp with a small raphe of fibrous tissue. Two other anatomic relationships also are of importance. Beneath the noncoronary cusp the stenotic ring is attached to the ventricular septum, where the conduction bundle may be injured if appropriate landmarks are not observed. Beneath the left coronary cusp, the ring is often attached to the base of the anterior leaflet of the mitral valve, which must be protected during excision of the ring.

Diffuse fibromuscular subvalvular stenosis or *tunnel-like subaortic obstruction* results from diffuse fibromuscular narrowing of the subvalvular left ventricular outflow tract. The fibromuscular obstruction usually is concentric and severe, encompassing a long "tunnel" of the outflow tract. Symptoms usually develop early in life and are severe. The obstructive

gradient is fixed and not dynamic, which helps in differentiating it from idiopathic hypertrophic subaortic stenosis.

Idiopathic Hypertrophic Subaortic Stenosis (IHSS) is an inherited hypertrophic cardiomyopathy that results in asymmetric septal hypertrophy, systolic anterior motion (SAM) of the anterior leaflet of the mitral valve, and dynamic left ventricular outflow tract obstruction. In patients with IHSS, symptoms gradually increase with age as the septal hypertrophy increases. The symptoms are angina, dyspnea, and syncope. A systolic murmur of medium intensity near the apex, but not prominent at the base of the heart, may be the first sign of disease. With progressive disease, atrial fibrillation, systemic emboli, and sudden death are the most significant events. Sudden death is distressingly common, presumably from an arrhythmia.

Supravalvular Stenosis

There is considerable variation in the anatomic type of supravalvular obstruction. Peterson and colleagues, reviewing 68 cases, found three types: *hourglass* (45 cases), *diffuse hypoplasia* (14 cases), and *membranous* (9 cases). Associated abnormalities are frequent, including a close association with Williams syndrome. Peripheral pulmonary stenosis should always be ruled out in patients with supravalvular aortic stenosis. In addition to peripheral pulmonary stenosis, focal stenosis of the aortic arch also may be present. Approximately 30 percent of the patients have associated involvement of the aortic valve cusps, usually because of distortion causing aortic regurgitation. Coronary abnormalities are found in a significant number of the patients.

Pathophysiology

The physiologic abnormality is directly related to the severity of the obstruction. At one extreme is the neonate with critical aortic stenosis. These babies present with severe heart failure and metabolic acidosis, often requiring intubation, inotropic support, and prostaglandin infusion to maintain ductal patency. Intensive therapy must be initiated before surgery to “unload” the heart, correct acidosis, and improve visceral organ perfusion. After stabilizing the patient, urgent intervention is indicated with balloon angioplasty or operative correction.

Milder forms of congenital aortic stenosis often have little immediate physiologic significance. These patients may remain asymptomatic for many years, slowly developing the classic findings of left ventricular hypertrophy and cardiomegaly. Progressive stenosis can lead ultimately to the development of significant concentric ventricular hypertrophy, with decreased exercise capacity, chest pain, arrhythmias, or congestive heart failure. A mean gradient of less than 50 mmHg usually does not produce enough disability to require operation, but correction should be considered in asymptomatic patients if the gradient is more than 50 to 60 mmHg and the cross-sectional area is less than 0.5 cm^2 . Another indication for operation is deteriorating left ventricular systolic function according to echocardiographic or radionuclide studies. Moderate levels of stenosis of uncertain significance can be further assessed by exercise testing. A finding of decreased exercise capacity, an abnormal drop in the ejection fraction in response to exercise, arrhythmias, or pulmonary congestion on exercise testing usually is an indication for operation. The goal is to initiate intervention on the basis of deteriorating but reversible changes in physiology—before the onset of irreversible cardiac damage. Operation is indicated in almost every patient once symptoms develop. In patients with subvalvular stenosis, operation is indicated once the gradient exceeds 40 to 50 mmHg or with the first sign of aortic valve involvement, before the development of significant aortic insufficiency, in an attempt to preserve the aortic valve.

Clinical Manifestations

Neonates with severe aortic stenosis present with congestive heart failure, acidosis, and low-output syndrome in the first days of life. Many older children with significant stenosis are asymptomatic, emphasizing the importance of echocardiography or catheterization to measure the severity of the abnormality. The most common symptoms are fatigue, dyspnea, angina, arrhythmias, and syncope, found in 30 to 50 percent of patients. Usually these are found in patients with a gradient of more than 50 to 60 mmHg and significant left ventricular hypertrophy.

The three most frequent physical findings are a systolic ejection murmur, a forceful left ventricular impulse, and a narrow pulse pressure. The systolic murmur is a harsh, ejection-type murmur, heard best in the second right interspace, widely transmitted to the neck and arms. A palpable thrill may be present. With long-standing disease and left ventricular hypertrophy the ventricular impulse is forceful and heaving. The pulse pressure often is decreased when stenosis is severe, although it may be normal. A diastolic murmur may indicate concomitant aortic insufficiency.

The electrocardiogram, indicating the degree of left ventricular hypertrophy, provides a moderately sensitive noninvasive guide to the severity of aortic stenosis, but it has distinct limitations. The usual abnormalities are signs of left ventricular hypertrophy and subsequent depression of the ST segment and inversion of T waves. With a gradient of more than 50 mmHg, a left ventricular strain pattern may be seen. Some patients may have severe obstruction with few ECG abnormalities. The chest radiograph may show signs of pulmonary congestion and cardiomegaly, or it may be normal in patients with more moderate stenosis.

The echocardiogram is helpful in establishing the diagnosis and should be performed immediately once the diagnosis is suspected. It is accurate in determining the transvalvular gradient and the annular size and in assessing the subvalvular left ventricular outflow tract. In patients with supravalvular stenosis, an MRI scan should be done to further delineate the anatomy and to assess the aortic arch and peripheral pulmonary arteries. If the diagnosis is suspected or the gradient is of uncertain significance, exercise testing and cardiac catheterization should be considered.

Treatment

Valvular Aortic Stenosis

Neonatal aortic stenosis requires urgent intervention. Balloon valvuloplasty and surgical valvotomy using cardiopulmonary bypass have been successful. Initial therapy depends on several factors, including the degree of valvular degeneration, the size of the aortic annulus, and the presence of associated pathology. Initial balloon valvuloplasty has been most successful in children with well-formed bicuspid valves and an aortic annulus of adequate size. When feasible this approach obviates taking the critically ill neonate to the operating room. In neonates with more primitive valvular anatomy, such as those with an extremely thickened bicuspid valve or a single-cusped valve, and in patients with a small, nearly hypoplastic annulus, surgical therapy yields the best initial results. In either case, initial valvuloplasty should be considered palliative; further surgical correction is almost always required later in life.

Surgical valvotomy is performed with the patient on cardiopulmonary bypass and cardioplegic arrest. The ascending aorta is incised with a curved incision extending down into the noncoronary sinus. Calibrated Hegar dilators are useful for measuring the diameter of the stenotic orifice before and after commissurotomy. Nomograms based on the child's body surface area are used to estimate the appropriate annular size. Cautious valvotomy is then performed, cutting the fused stenotic area up to the commissure (Fig. 17-5). Overcorrection can result in tearing the valve and producing aortic insufficiency. The annulus usually grows as increased antegrade flow is established.

With valvular stenosis, the fused commissures are carefully incised with a small knife blade, carefully dividing the fused commissures exactly along the center of the fibrous raphe in order to leave a thick margin on each of the two cusps that are separated. The incisions may be carried to the aortic wall or terminated 1 to 2 mm from the wall, depending on the anatomy. Commissural incisions should be limited to where the commissures are well formed. With the classic bicuspid valve, the commissure between the right and left cusps is underdeveloped. Usually no incision at all is made in this area, leaving the valve as a bicuspid valve. Primitive, thickened, hypoplastic leaflet tissue is partially excised to decrease the thickness of the leaflets and improve leaflet mobility. In older children with isolated stenosis from a bicuspid valve, balloon valvuloplasty often is the treatment of choice.

Valve replacement is reserved for children with significant annular hypoplasia or those with recurrent stenosis. When significant annular hypoplasia is present, a more extensive annular enlarging—valve replacement procedure often is necessary. Also, progressive aortic insufficiency or mixed recurrent stenosis with progressive insufficiency may develop in children who have previously undergone correction of valvular or subvalvular stenosis, necessitating reoperation for valve replacement.

Most children who require valve replacement are now treated with the Ross procedure (pulmonary autotransplant). When significant annular hypoplasia is present, extended aortic root replacement using a cryopreserved human homograft or a combination of Ross procedure and root enlargement is used. The use of prosthetic valve replacement is avoided whenever possible in children, since xenografts (bioprostheses) have a high early failure rate from progressive calcification, and mechanical prostheses necessitate long-term anticoagulation therapy.

The operative risk for repair of critical aortic stenosis in neonates is significant, with mortality ranging from 10 to 30 percent, depending on the size of the aortic annulus and the degree of left ventricular hypoplasia. For older children, there is little operative risk in repair of valvular aortic stenosis, with mortality less than 1 percent. Late reoperation often is required for patients undergoing correction early in life, but it may be postponed for many years if the initial procedure is successful. Valve replacement often is required as the secondary procedure in 10 to 15 years. The operative mortality is approximately 1 percent for the Ross procedure and other valve replacement operations not requiring extensive annular enlargement.

Hsieh and associates reported long-term results in 59 patients with a mean follow-up of 18 years. Of the 13 patients who died, sudden death occurred in seven, at least four of whom were known to have significant residual disease. Actuarial analysis indicated that the probability of reoperation increased from 2 percent at 5 years to 44 percent at 22 years. Dobell et al. reported more discouraging results; one-third of the group required a repeat operation within 10 years.

Elkins and associates reviewed 301 patients treated with balloon or surgical valvuloplasty. A subsequent valve replacement was required in 160 patients, whose progress was then followed to evaluate the efficacy of valve replacement. Survival was significantly better after the Ross operation than after other forms of valve replacement. Similarly, the risk of reoperation was lower after the Ross operation (87 percent freedom from reoperation at 9 years) than after other forms of valve replacement (55 percent freedom from reoperation at 9 years). The Ross operation offers significant advantages over other forms of valve replacement in children, with better rates of survival and freedom from reoperation and a better quality of life.

All patients undergoing treatment for congenital aortic stenosis should be seen at yearly intervals indefinitely, with echocardiograms performed to assess valve function and to detect any growth of the aortic annulus. Exercise testing may be indicated before clearing patients for vigorous physical exercise. The long-term prognosis varies, with many patients going over 20 years without subsequent problems. Most reports indicate that at least 30 to 50 percent of the patients require a subsequent operative procedure within 10 to 20 years. Once valve replacement is required, the late results after the Ross procedure are exceedingly good, although these patients must also be observed in long-term follow-up, as late operation may subsequently be necessary.

Subvalvular Stenosis

For treatment of patients with a *discrete subvalvular ring*, the valve cusps can be carefully retracted and the fibrotic ring excised (Fig. 17-6). Excellent visualization is required to prevent injury to the base of the aortic valve, the mitral valve, or the conduction bundle in the ventricular septum. The ring may consist of thin, fibrous tissue, which is easily removed, or it may be a thick, fibrotic structure. The fibrous ring often involves the base of the left coronary cusp of the aortic valve and the anterior leaflet of the mitral valve and

must be removed carefully from these areas. Usually the fibrous ring involves approximately 270 degrees of the circumference of the left ventricular outflow tract. Complete excision is required.

In many patients with a known fibrous ring, more diffuse subvalvular obstruction also is present, and failure to address this finding results in an increased incidence of late recurrence and restenosis. The surgeon should carefully look for fibromuscular obstruction during surgery and, if any is found, a block of muscle from the septum should be resected, similar to the operation for *IHSS*. Subvalvular obstruction can recur after initial operative correction, particularly in patients who have diffuse fibromuscular obstruction. The incidence of late operation is lessened by the more aggressive use of septal myotomy and myectomy during the initial operative correction of subvalvular obstruction.

Tunnel-like or *diffuse subvalvular obstruction* requires extensive resection of fibromuscular tissue from the left ventricular outflow tract. The operation is similar to the myotomy and myectomy performed for *IHSS*. This is adequate in most cases, although patients with severe *LVOT* obstruction may have a combination of annular hypoplasia and subvalvular obstruction and require aortoventriculoplasty. Konno described the initial aortoventriculoplasty procedure, using Dacron patch enlargement of the LVOT and aortic annulus and aortic valve replacement with a mechanical prosthesis. Patients who require aortoventriculoplasty are more often treated with extended root replacement using a homograft or with the combination of the Ross procedure and patch enlargement of the outflow (Fig. [17-7](#)). These procedures obviate the need for anticoagulation therapy, and the Ross procedure also has the advantage of growth potential.

IHSS

Most patients with IHSS are treated medically initially, with operation reserved for those with symptomatic, severe obstruction that does not improve with medical therapy. Beta-adrenergic blockers or calcium channel blockers usually are used. Surgical myotomy and myectomy is indicated in symptomatic patients with an outflow gradient of 50 mmHg or greater despite medications.

Using a transaortic approach, a rectangular block of ventricular muscle is excised from the septum, extending from just below the aortic valve, down the septum several centimeters, stopping at the level of the bottom of the anterior leaflet of the mitral valve. The myectomy may be 0.5 to 1.5 cm wide and should stay to the left of the right coronary ostium to avoid penetrating the conduction bundle. Upon completing the procedure, relief of the gradient should be confirmed by measuring left ventricular and aortic pressures.

Supravalvular Stenosis

With the hourglass type of stenosis, widening the stenotic area by adding a patch of Dacron or pericardium produces excellent correction. Before the patch is placed, the fibrous ridge above the sinotubular junction should be excised as completely as possible. Usually a bilobed patch is placed with extension into the right and the noncoronary sinuses of Valsalva (Fig. [17-8](#)). This opens the sinotubular junction and provides effective relief of obstruction. An alternative method is to excise totally the supravalvular ridge of stenosis and perform an end-to-end anastomosis between the distal aorta and the aortic root, advancing small "tongues" of the distal aorta into the noncoronary and right coronary sinuses of Valsalva for enlargement. The incidence of reoperation is low after correction of supravalvular stenosis.

Congenital Mitral Valve Disease

Congenital abnormalities of the mitral valve are rare, accounting for less than 1 percent of all cases of congenital heart disease. Rheumatic heart disease, endocarditis, and cardiomyopathy can produce mitral valve pathology during childhood. Severe mitral insufficiency of ischemic origin usually is seen in children only in association with an aberrant origin of the left coronary artery.

Pathology

Four general types of congenital pathologic mitral valve abnormalities were described by Ruckman and Van Praagh: (1) Typical congenital mitral stenosis, with varying degrees of obliteration and fusion of the chordae tendineae and subvalvular apparatus and mild-to-moderate deficiency of leaflet tissue, is the most common form (50 percent) of congenital mitral valve disease. Mitral insufficiency may be present concomitantly because of annular dilation, leaflet prolapse, abnormal clefts, leaflet tissue deficiency, and subvalvular restriction. (2) True "parachute mitral valve" is relatively uncommon. The term "parachute mitral valve" refers to the insertion of all chordae tendineae into a single, shortened papillary muscle. This results in restricted leaflet mobility with valvular obstruction. More commonly the pathology is somewhere along a continuum between typical congenital mitral stenosis, which results from fusion or obliteration of "fanlike" chordae, and "parachute mitral valve" as a result of true single papillary muscle. (3) A more severe form of congenital mitral stenosis develops when the mitral annulus is significantly hypoplastic. Mitral stenosis with annular hypoplasia may be part of diffuse left ventricular outflow tract obstruction. Hypoplastic left heart syndrome represents the most extreme form of this condition. Some degree of annular hypoplasia is present in approximately 40 percent of patients with congenital mitral stenosis. (4) Supraannular mitral stenosis is a relatively rare condition resulting from a supraannular ring of connective tissue in the left atrium, producing physiologic stenosis.

Associated cardiac malformations occur in 75 percent of patients with congenital mitral obstruction. These include ventricular septal defect (30 percent), valvular aortic stenosis (29 percent), aortic atresia and hypoplastic left heart syndrome (29 percent), less severe forms of subvalvular *LVOT* obstruction (30 to 60 percent), and coarctation of the aorta (27 percent). Abnormal left ventricular muscle with endocardial fibroelastosis occurs in most patients with mitral stenosis and diffuse LVOT obstruction and in nearly 50 percent of all patients with congenital mitral stenosis. The combination of supraannular mitral ring, parachute mitral valve, left ventricular outflow obstruction, and aortic coarctation is called Shone's complex.

Clinical Manifestations

Symptoms of pulmonary venous congestion often appear in infancy and include dyspnea, orthopnea, and pulmonary edema. If the condition is untreated approximately one-half of the patients die within 6 months after the appearance of symptoms. Those with less severe obstruction, better left ventricular development, and fewer associated lesions may not develop symptoms until 2 to 4 years of age and rarely remain asymptomatic until 10 to 12 years of age.

The chest radiograph and electrocardiogram often demonstrate an enlarged left atrium, pulmonary congestion, and P mitrale of stenotic lesions. With mitral insufficiency, cardiomegaly and pulmonary congestion are often present. The echocardiogram is the primary diagnostic tool, with transesophageal echocardiography used as necessary. With transesophageal echocardiography lesions in the left atrium, the valve, and the subvalvular apparatus can be readily identified and the degree of stenosis and insufficiency accurately assessed. Cardiac catheterization may be recommended before operative repair in some patients.

Treatment

Operation should be timed to avoid significant deterioration in left ventricular function. Because of the possibility that repair may not be feasible and valve replacement may be required, the anatomy should be carefully analyzed echocardiographically before surgery. If valve repair is not feasible, the patient may be managed medically until later in the clinical course. If valve repair is feasible, operation should be recommended at the first sign of deteriorating ventricular function. While operation is clearly indicated, once congestive heart failure or significant symptoms develop, a more optimal time for intervention would be before the development of symptomatic left ventricular dysfunction, such as the first echocardiographic sign of diminishing systolic function. If valve repair is accomplished at this time, the long-term functional results are excellent.

Results

Coles reported operative experiences with reparative procedures in 48 patients, with an operative mortality of only 2.9 percent. This study suggests improved late results when valve repair was possible compared with a similar group of patients undergoing mitral valve replacement. Kodoba reported a 5-year survival rate of only 43 percent after mitral valve replacement in the first year of life, and a 3-year freedom-from-reoperation rate of only 45 percent. The most important factors affecting survival are the severity of associated defects and the adequacy of left ventricular function. After operation late trends suggest improved results when mitral valve repair is feasible, and this is preferred over valve replacement whenever possible. Our own experience with more than 70 cases of congenital mitral valve procedures over the past 15 years suggests that valve repair should be feasible in more than 90 percent of patients, with an operative mortality of approximately 1 percent. The long-term results after valve repair are excellent.

Cor Triatriatum

Pathology

This abnormality may be viewed as a variant of total anomalous pulmonary venous drainage, except that the unreabsorbed common pulmonary venous sinus empties into the left atrium through a restricted aperture rather than through a vertical vein to the right side of the heart. The common venous chamber is superior and posterior to the left atrium, with a diaphragm separating this chamber from the true left atrium, where the left atrial appendage enters. A small opening in the thick muscular diaphragm is the only communication between the upper pulmonary venous chamber and the lower true atrial chamber. This severely obstructs pulmonary venous return and produces supraannular mitral obstruction, with left-to-right shunting across any defect in the atrial septum. An atrial septal defect is present in about 70 percent of cases, generally communicating with the common upper pulmonary venous chamber, resulting in a left-to-right shunt.

Clinical Manifestations

Cor triatriatum produces severe pulmonary congestion with pulmonary artery hypertension, identical to the pathophysiology of mitral stenosis. Congestive heart failure is usually quite severe, and the patients' symptoms are pronounced. Gradients as high as 20 mmHg have been recorded between the venous chamber and the lower left atrial chamber. The mortality is high when pulmonary venous obstruction is present, and if the anomaly is not corrected operatively 70 to 75 percent of these infants die in the first year of life.

The chest radiograph shows pulmonary congestion with right ventricular enlargement. Right ventricular hypertrophy is evident on the electrocardiogram, varying with the degree of pulmonary hypertension. Standard or transesophageal echocardiography usually is diagnostic, outlining the abnormal chambers. MRI is helpful in establishing the diagnosis in some cases. Cardiac catheterization is often unnecessary, although it permits measurement of pulmonary artery pressure and precise delineation of other, associated anomalies; catheterization may be used if the diagnosis is uncertain. The differential diagnosis includes total anomalous pulmonary venous return, congenital pulmonary vein stenosis, and mitral stenosis. Occasionally cor triatriatum is encountered unexpectedly in a patient with an atrial septal defect and an inordinate degree of left-to-right shunting.

Treatment

The need for early operation varies, depending on the severity of pulmonary venous obstruction. Operation should be performed promptly in patients with severe obstruction because of the rapid development of pulmonary hypertension. Less severely obstructed patients with significant left-to-right shunting may be operated on later in childhood, as is the case for correction of isolated atrial septal defect.

During infancy hypothermic circulatory arrest may be used, although standard perfusion techniques are used in older patients. The defect is approached through the right atrium, enlarging the atrial septal defect and excising the left atrial membrane to create a common, unobstructed left atrial chamber. This eliminates all obstruction to pulmonary venous

drainage and allows unrestricted flow through the normal mitral valve. The atrial septal defect is closed with a patch of pericardium. The operative risk is low, with a mortality of less than 1 percent, and late results are excellent, with surviving patients exhibiting normal physiology.

INCREASED PULMONARY BLOOD FLOW (LEFT-TO- RIGHT SHUNTS)

Patent Ductus Arteriosus

Patent ductus arteriosus is one of the most common forms of congenital heart disease, constituting about 10 percent of all cases of congenital heart disease. The increased incidence of patent ductus in premature infants is widely recognized, varying inversely with the birth weight and gestational age, ranging in frequency from of 25 to 100 percent. The presence of a patent ductus arteriosus represents a normal physiologic state in severely premature infants. In older, full-term infants, the incidence of patent ductus is much less common, representing a true pathologic event. Patent ductus arteriosus is two to three times more common in females than in males.

The ductus arteriosus, which develops as an embryologic remnant of the sixth left aortic arch, is an important normal fetal pathway connecting the pulmonary artery at its bifurcation to the aorta just beyond the origin of the left subclavian artery. Through this channel in embryonic life, blood bypasses the collapsed lungs, flowing directly from the pulmonary artery into the aorta. With the expansion of the lungs at birth, the ductus normally closes within a few days, becoming the fibrotic ligamentum arteriosum.

The physiologic stimuli responsible for closure of the ductus have been studied in detail. Changes in the oxygen tension of the arterial blood exert a profound stimulus on the closure. An important mechanism of closure lies the distinctive histologic structure of the wall of the ductus, which is different from that of the pulmonary artery or the aorta. As the ductus closes, the wall of the ductus contracts, the internal elastic membrane fragments, and smooth muscle projects into the lumen as progressive fibrosis obliterates the patent channel. Naturally occurring prostaglandins oppose closure of the ductus. In full-term infants, the ductal tissue is more responsive to oxygen, leading to ductal closure, whereas in premature infants the prostaglandin effect is more prominent, resulting in increased ductal patency.

Closure of a patent ductus can be significantly hastened by the administration of indomethacin, which blocks the synthesis of the prostaglandins that normally oppose contraction of ductal smooth muscle. In contrast, ductal patency can be maintained when necessary by an intravenous infusion of prostaglandin E₁.

Pathology and Pathophysiology

The diameter of a ductus typically is 5 to 7 mm, though it ranges from as little as 2 to 3 mm to more than 10 mm. The length of the ductus ranges from 5 to 10 mm. Associated anomalies occur in approximately 15 percent of cases; the most common are ventricular septal defect and coarctation of the aorta, although a patent ductus may occur in association with many other congenital cardiac defects.

Depending on the diameter of the ductus, a varying amount of blood is shunted from the aorta to the pulmonary artery, constituting a left-to- right shunt. In a large ductus, the Q_p:Q_sratio may be 3:1 or greater. The severity of symptoms is directly proportional to the size of the shunt.

In infancy the high pulmonary vascular resistance of fetal life subsides gradually in the first month of life. In some infants with patent ductus, the increased pulmonary blood flow causes the pulmonary vascular resistance to remain elevated, resulting in pulmonary hypertension. The pulmonary resistance usually decreases immediately to normal after closure of the patent ductus, but only a partial regression may occur if the pulmonary vascular resistance has been elevated for a long time. In some patients the pulmonary vascular resistance may become fixed by as early as 5 to 6 years of age, resulting in right-to-left shunting and cyanosis. Patients with patent ductus arteriosus have an increased susceptibility to bacterial endocarditis.

Clinical Manifestations

In premature infants a large patent ductus may cause serious heart failure, whereas most full-term infants and older children are asymptomatic. When symptoms are present, the most common are palpitations, dyspnea, and decreased exercise tolerance. Symptoms of congestive heart failure may develop in adults. Approximately 40 percent of patients with a large patent ductus die before the age of 45 years if untreated.

The hallmark of a patent ductus arteriosus is the continuous “machinery” murmur, one of the most distinctive signs in clinical medicine. A wide pulse pressure usually is found with a large ductus, produced by a decrease in diastolic pressure. In the extremely large ductus, the diastolic pressure may approach very low levels and be associated with peripheral vascular findings similar to those of severe aortic insufficiency. Signs of pulmonary congestion may be present.

The diagnosis may be made in premature infants from the widened pulse pressure detected through an umbilical arterial catheter and confirmed by echocardiography. The chest radiograph usually shows increased pulmonary blood flow. In older children the diagnosis is made from the characteristic murmur and by echocardiography. Catheterization usually is used only in the treatment of children with patent ductus arteriosus and is not necessary for diagnostic purposes. In adult patients diagnostic catheterization is performed to measure the pulmonary vascular resistance and to assess ductal length and any calcification, which may make repair difficult.

Treatment

Ductal closure can be achieved in most premature infants with indomethacin therapy, producing ductal closure in approximately 90 percent of patients with two or three courses of therapy. Operation is indicated in premature infants with severe respiratory insufficiency who do not respond to indomethacin therapy. In full-term infants who do not have congestive heart failure, ductal closure should be performed at some time between 6 months and 2 years of age. The intraluminal insertion of a specially designed coil in the cardiac catheterization laboratory results in successful occlusion of the patent ductus the in vast majority of older children. Surgical division is recommended in children who have a ductus of large diameter or of extremely short length, making coil occlusion more risky or unfeasible. The mortality of surgical ductal closure approaches zero.

Mikhail and colleagues reported on experiences with operative closure through a small posterolateral thoracotomy in 306 infants with an average age of 11 days, without any deaths. Some centers recently have used thoroscopic approaches for ductal closure with excellent results, although additional clinical benefit over standard techniques is hard to document. Coil occlusion in older children also carries a low operative risk, with a mortality close to zero, and achieves excellent intermediate-term results.

Operative Technique

For operative closure, the patent ductus is exposed through a short, posterolateral incision in the fourth intercostal space. In premature infants the patent ductus is double ligated with a heavy silk or doubly occluded with metallic clips; operating time usually is less than 30 min and the procedure is extremely safe. In older children the preferred technique is complete surgical division. A partial-occlusion clamp is placed on the aorta adjacent to the ductus and an angled vascular clamp is placed on the pulmonary-artery side of the ductus. The ductus is divided sharply and each side is oversewn with a continuous double layer of monofilament suture. Care is taken to ensure that the suture line is flush with the aorta to avoid late aneurysm formation because of an unobliterated ductus diverticulum.

Aortopulmonary Window

Aortopulmonary window is a rare abnormality. At the Toronto Children's Hospital, only 23 of 15,000 patients with congenital heart disease who were seen over a period of 20 years had an aortopulmonary window.

Pathology and Pathophysiology

Embryologically, the defect results from incomplete development of the spiral septum dividing the primitive truncus arteriosus into the aorta and pulmonary artery. Persistent truncus arteriosus is a more severe malformation of similar cause. The opening, or “window,” between the aorta and the pulmonary artery may vary in diameter from 5 to 30 mm. Usually it is located proximally, near the ostium of the coronary arteries. At least 30 percent of patients have a severe additional cardiac malformation.

The large left-to-right shunt is similar to that of a large patent ductus arteriosus or ventricular septal defect. The clinical course is rapidly progressive because of the tremendous amount of shunting, with early development of severe congestive heart failure. Increased pulmonary vascular resistance occurs quickly, by 6 months of age.

Clinical Manifestations

The clinical findings may be identical to those of a large patent ductus arteriosus, with a continuous murmur and wide pulse pressure. Often, however, only a systolic murmur is present because of the severe pulmonary hypertension. The differential diagnosis includes large patent ductus, ventricular septal defect, and truncus arteriosus.

The diagnosis is readily made by echocardiography and can be confirmed by MRI if necessary. Cardiac catheterization usually is unnecessary unless the child is older and a fixed pulmonary vascular resistance is suspected.

Treatment

Operation should be performed as soon as the diagnosis has been established because of the rapidity of development of irreversible pulmonary vascular disease. A transaortic approach usually is used, and large defects are closed with a prosthetic patch. The pulmonary artery is clamped separately. Care is taken to avoid injury to the coronary arteries and the pulmonary valve. In patients operated on in infancy or before the development of severe pulmonary vascular disease, the results have been excellent. The operative risk is proportional to the increase in pulmonary vascular resistance, and the overall operative mortality is less than 5 percent.

Ventricular Septal Defects

Ventricular septal defect (VSD) is a common form of congenital heart disease, constituting 20 to 30 percent of congenital defects. There are no known etiologic factors. Associated anomalies are common. These include patent ductus arteriosus, coarctation of the aorta, atrial septal defect, right ventricular outflow tract obstruction, tetralogy of Fallot, double-outlet right ventricle, transposition of the great arteries, truncus arteriosus, and aortic insufficiency from prolapse of an aortic valve cusp into the ventricular septal defect.

Pathology and Pathophysiology

Ventricular septal defects have been classified according to the position in the ventricular septum: (1) perimembranous, (2) posterior inlet or atrioventricular canal type, (3) outlet or supracristal, and (4)muscular (Fig. 17-9).

Perimembranous VSDs (Fig. 17-10) are the most common in patients requiring surgery, accounting for 80 percent. These defects involve the membranous septum and include the malalignment type of defects seen in tetralogy of Fallot. The bundle of His is located along the posterior, rightward rim of the septum, where it bifurcates into left and right conduction bundles. The defect may extend superiorly into the outlet septum next to the aortic valve annulus, which often is visible through the upper part of the defect. Operative

closure with a Dacron patch frequently is possible by working via the right atrium through the tricuspid valve (Fig. [17-11](#)). In some cases in which the defect is high, a transverse right ventriculotomy incision may be used for exposure. The base of the tricuspid valve leaflet may be temporarily detached for better exposure of the upper part of the defect. Posteriorly, the sutures often are woven into the base of the tricuspid valve leaflet to avoid the conduction tissue along the crest of the septum.

A rare perimembranous defect produces shunting from the left ventricle to the right atrium, either directly or through defective tricuspid tissue along the base of the annulus. The resulting shunt usually is large because of the great difference in pressure between the left ventricle and the right atrium.

Inlet or atrioventricular canal type defects are perimembranous defects that extend posteriorly, beneath the conal papillary muscle and the tricuspid valve, involving the inlet septum. The conduction tissue runs adjacent to the rim of the septum, similar to perimembranous defects, and care must be taken to avoid heart block during repair.

Outlet or supracristal defects are in the infundibular septum, adjacent to the pulmonary and aortic valves. Consequently, aortic insufficiency is a common feature because of deficient support in the aortic annulus combined with a Venturi effect from the left-to-right shunt. These defects are safely away from the conduction bundle, so heart block is not a risk, but concomitant repair of the aortic valve frequently is necessary.

Muscular *VSDs* are the most common, although many close spontaneously and do not require surgery. Closure usually occurs before 2 to 3 years of age. Muscular defects are located inferiorly, anywhere in the muscular septum, and are often multiple. They are safely away from conduction tissue. The rare “Swiss cheese” type of muscular VSD consists of multiple communications into the right ventricle, making visualization and closure difficult.

VSDs vary from as small as 3 to 4 mm to larger than 3 cm in size. The defect may be classified as restrictive or nonrestrictive, depending on whether the right ventricular pressure is elevated to systemic levels or not. In general the nonrestrictive defect is equal in diameter to the aortic annulus. In patients with a large, nonrestrictive VSD the right ventricular pressure is equal to systemic pressure and the left-to-right shunt ($Q_p:Q_s$) may be 4:1 or greater, producing pulmonary congestion and heart failure. In large defects the degree of shunting is dictated by the relative levels of pulmonary and systemic vascular resistance. With a long-standing defect, the pulmonary resistance may increase significantly, leading to irreversible pulmonary vascular changes. When an irreversible increase in pulmonary vascular resistance develops, a right-to-left shunt may develop, leading to cyanosis. This pathologic condition is referred to as Eisenmenger’s syndrome.

Small restrictive septal defects ($Q_p:Q_s < 1.5:1$), which produce little or no increase in the right ventricular pressure, have few immediate physiologic consequences. The long-term risk of bacterial endocarditis is increased, apparently because of endocardial damage from the jet of blood through the defect. Some cardiologists recommend closure of small defects to lower the lifelong risk of endocarditis, but this is controversial and not routinely performed.

Clinical Manifestations

Patients with small defects usually are asymptomatic even though a loud murmur and thrill may be present. With large defects, severe heart failure, dyspnea, and pulmonary congestion are common. These patients may have frequent respiratory symptoms and pneumonia or may show a lag in growth and development. Severe cardiac failure usually occurs within the first few months of life or much later in adulthood. Patients with increased pulmonary vascular resistance may be deceptively asymptomatic for several years, until cyanosis and hemoptysis develop.

The chest radiograph may be normal or may show cardiomegaly with pulmonary congestion. The electrocardiogram usually shows signs of left ventricular or biventricular hypertrophy. The echocardiogram is almost always diagnostic, demonstrating the location and size of the defect and any associated cardiac anomalies. The echocardiogram can be used to estimate the magnitude of left-to-right shunting and the pulmonary artery pressure. More precise information can be obtained by cardiac catheterization, but this usually is not necessary unless the pulmonary pressure is near the systemic pressure, requiring an accurate assessment of pulmonary vascular resistance.

Treatment

Small defects should simply be observed, since 60 to 70 percent will close in early life. More than one-half of small muscular defects close before 3 years of age and about 90 percent by 8 years.

The treatment of large defects depends on the presence of cardiac failure or increasing pulmonary vascular resistance. Severe cardiac failure in infancy may be fatal, and prompt operation is performed in the first few weeks or months of life. Even without severe heart failure, children with large left-to-right shunts are operated on electively between 3 and 6 months of age, since irreversible changes in the pulmonary vascular bed may develop early, even during the first year of life. When pulmonary artery pressure is already at systemic levels and the pulmonary vascular resistance is increased, the operative risk is higher, but long-term benefit can be obtained if the patient still has a reactive pulmonary vascular bed with left-to-right shunting. The criteria for inoperability vary significantly but generally include a nonreactive, fixed pulmonary vascular resistance of greater than 10 Woods units. When operation is performed at less than 6 to 9 months of age, reversibility usually is noted.

Operative Technique

Operation is performed through a median sternotomy with extracorporeal circulation. If a patent ductus arteriosus is present, it must be closed at the beginning of operation. The *VSD* usually is closed through a right atrial approach and occasionally through a short transverse ventriculotomy. The transatrial approach is preferable because it minimizes the risk of ventricular dysfunction and arrhythmias. A prosthetic patch is routinely used (see Fig. [17-11](#)). The critical part of the operation is to avoid heart block by identifying key anatomic guides at the posterosuperior margin of the defect. With a still, dry field, the fibrous trigone located at the bottom of the noncoronary sinus can be identified on inspection through the ventricular septal defect. The conduction bundle passes through this trigone and then along the area where the membranous septum joins the muscular septum posteriorly (see Fig. [17-10](#)). Placing sutures to the right of an imaginary line projected between the fibrous trigone and the papillary muscle of the conus usually avoids injury to the conduction bundle.

With today’s techniques, the risk of heart block is very small. Outlet (infundibular) defects and muscular defects can be readily repaired with a prosthetic patch because there is no danger of injury to the conduction bundle. After cardiopulmonary bypass has been established, residual shunts may be detected by measuring the oxygen content of blood samples drawn simultaneously from the right atrium and the pulmonary artery. Intraoperative transesophageal echocardiography also may be used to confirm complete closure of the defect.

A small proportion of patients with *VSDs*, usually those with outlet defects, develop severe aortic insufficiency from prolapse of an aortic valve cusp into the underlying defect. The insufficiency often is progressive as the aortic cusp herniates to a greater degree, virtually tamponading the underlying septal defect. Surgical correction often can be done by aortic valvuloplasty.

Results

The operative mortality in *VSD* repair is less than 2 to 3 percent, depending on the size and the age of the patient and the pulmonary vascular resistance. The risk of perioperative heart block is less than 2 percent. Residual defects may develop around the repair in a small percentage of patients but are usually of no hemodynamic consequence. Late results are excellent; most patients have normal ventricular function and exercise capacity after repair.

Atrial Septal Defects

A variety of malformations involve the atrial septum or the pulmonary veins and result in a left-to-right shunt of blood from the systemic to the pulmonary circulation. These include ostium secundum defects, sinus venosus defects with partial anomalous pulmonary venous return, and ostium primum defects (incomplete atrioventricular septal defects). The physiologic abnormality is identical with secundum defects and with sinus venosus defects, consisting simply of a left-to-right shunt. Atrial septal defects are among the most common cardiac malformations, representing 10 to 15 percent of all cases of congenital heart disease. They are more than twice as frequent in females than in males. Embryologically, the secundum defects result from failure of the septum secundum to develop completely.

Pathology and Pathophysiology

Atrial septal defects (ASDs) vary widely in size and location. The majority of *ASDs* are ostium secundum defects (Fig. [17-12](#)), located in the middle part of the atrial septum in the area of the ostium secundum. *ASDs* may range from openings as small as 1 cm in diameter to the virtual absence of the atrial septum with a common atrium. Most are 2 to 3 cm in diameter. A patent foramen ovale should not be considered an atrial septal defect, for it is a normal opening in 15 to 25 percent of adult hearts. Because of its slit-like construction, a normal foramen ovale allows shunting of blood only from right to left. A “high” subcaval defect near the orifice of the superior vena cava is commonly referred to as a sinus venosus defect (Fig. [17-13](#)), and it is usually associated with anomalous entry of one or more right pulmonary veins into the superior vena cava. This most common form involves drainage of the upper and middle lobe veins, which enter the superior vena cava below the entry site of the azygos vein. More unusual defects include common atrium, different forms of “unroofing” of the coronary sinus, and low ostium secundum defects that extend toward the inferior vena cava. Right pulmonary veins, which enter the right atrium directly, may be found with a common atrium or with a large posterior ostium secundum defect. Part of the septum may be fenestrated, but true multiple defects are rare.

An unusual variant of partial anomalous pulmonary veins entering the inferior vena cava has been described as a “scimitar” syndrome, emphasizing a characteristic radiologic appearance resulting from the shadow of the anomalous vein, which is parallel to the right border of the heart. The malformation is associated with hypoplasia of the right lung and anomalous origin of the pulmonary arteries from the aorta. A left-to-right shunt is present.

A rare variant of atrial septal defect is an ostium secundum defect combined with mitral stenosis, known as Lutembacher’s syndrome. The mitral stenosis retards the flow of blood from the left atrium to the left ventricle and produces an enormous left-to-right shunt through the septal defect, with massive dilatation of the pulmonary arteries. Some mitral valve prolapse occurs in 10 to 20 percent of patients with atrial septal defects, because of underfilling of the left ventricle. This usually resolves simply by repairing the atrial septal defect, with less than 5 percent having significant mitral insufficiency that must be corrected at operation.

An atrial septal defect results in a left-to-right shunting because of compliance and resistance characteristics between the left and right sides of the cardiac circulation. The thick-walled left ventricle is less compliant than the right ventricle, and the systemic vascular resistance is higher than the pulmonary resistance. The shunt is left-to-right despite the absence of a pressure difference across the defect. The size of the septal defect exerts influence only if it is small and restrictive. During infancy the left-to-right shunt remains small until the pulmonary vascular resistance drops at 1 to 3 months of age and the right ventricle becomes more compliant, increasing the shunt. Most older children with defects of significant size have shunts of 3:1 or greater.

Pulmonary vascular changes do not develop early in life, in contrast to patients with ventricular septal defects, and pulmonary hypertension infrequently develops, occurring in less than 5 percent of children with atrial septal defects. Increased pulmonary vascular resistance may occur during the fourth or fifth decade, however, if the defect is left untreated.

Cardiac failure occurs very late, usually in the sixth decade, after the onset of pulmonary hypertension and atrial fibrillation. Without treatment, life expectancy is decreased by 15 to 20 years.

Clinical Manifestations

Symptoms are uncommon in the first few years of life because the shunt is small until the right ventricular hypertrophy of infancy has subsided. Frequently children with large shunts are physically active and asymptomatic. The most frequent symptoms are fatigue, palpitations, and exertional dyspnea. In adults, signs of congestive heart failure or arrhythmias gradually appear, often with the first pregnancy or later in life.

A soft systolic murmur usually is audible in the second or third left intercostal space. In the first few years of life this murmur may be faint or considered a functional murmur, but it increases after 2 years of age. The murmur arises from increased flow through the right ventricular outflow tract. The second pulmonic sound is characteristically widely split and “fixed.”

The chest radiograph may show mild to moderate cardiac enlargement, principally the result of a large right ventricle. The right atrium and pulmonary artery are also prominent, with increased vascularity in the lung fields. The electrocardiogram usually is characteristic with a right-axis deviation. The echocardiogram usually is diagnostic, and cardiac catheterization is seldom necessary.

Treatment

Because most children with atrial septal defects are asymptomatic, operation is frequently recommended on the basis of a large echocardiographic defect with significant left-to-right shunting. Most defects 1 to 2 cm in size or larger should be closed at 2 to 4 years of age. The only contraindication to operation is increased pulmonary vascular resistance and Eisenmenger’s syndrome, which occurs much later in life.

Operative Technique

The operation is performed with the patient on extracorporeal perfusion (cardiopulmonary bypass) through a standard sternotomy incision or through a less invasive right minithoracotomy incision. Complete repair with excellent long-term results is achieved with both approaches, and the surgical team should use the approach that they are most familiar with; the overall operative mortality should approach zero in patients undergoing atrial septal defect repair.

Once operative exposure is obtained and the patient is placed on cardiopulmonary bypass, the heart is electrically fibrillated or arrested with cardioplegia, so that the defect can be repaired precisely. The right atrium is opened (Fig. 17-14), and the defect is closed with continuous suture or with a small pericardial patch. Most ostium secundum defects in children can be closed primarily. Large ostium secundum defects with a near common atrial chamber may require patch closure with pericardium (Fig. 17-15). Large defects with partial anomalous venous drainage into the right atrium are corrected by placement of a pericardial patch in such a way as to recreate the atrial septum, redirecting the anomalous veins into the newly created left atrial chamber.

Sinus venosus defects also are closed by rerouting the anomalous pulmonary venous blood from the right upper and middle lobe veins across the high sinus venosus defect, using an intracardiac pericardial baffle. This simultaneously corrects the partial anomalous venous drainage and closes the septal defect. Care must be taken to avoid obstructing drainage of the anomalous veins or the superior vena cava. Native atrial tissue often is used to enlarge the caval-atrial junction and prevent obstruction of the superior vena cava after sinus venosus repair, using an inverted Y-to-V advancement flap of right atrial appendage onto superior vena cava.

Several centers have performed closure of secundum defects in the catheterization laboratory, using either “clamshell” or “angel-wing” devices. Early reports have been encouraging in properly selected patients, although some problems with device fracture or migration have occurred in a small number of patients. The long-term risks of clot formation and endocarditis also remain uncertain. Device closure of atrial septal defects is an exciting new technical development that may be feasible in a significant number of patients with ostium secundum–type defects.

Results

Operative mortality for surgical treatment of uncomplicated atrial septal defect approaches zero. In our experience at the New York University Medical Center, no deaths have occurred in children undergoing repair of secundum or sinus venosus defects over the past 15 years, which supports the policy of routine closure of virtually all ASDs. Long-term survival is primarily related to age and pulmonary vascular resistance. Murphy and associates found late survival rates to be equal to those of the normal population in patients operated on early in life before the development of atrial fibrillation, heart failure, and pulmonary hypertension. The 30-year survival rate was 97 percent for patients whose defect was repaired early in life but only 40 percent for those whose repair was made after 40 years of age.

Outcomes data support a policy of routine closure of ASDs, even in older adult patients, as long as the patient does not have a fixed increased pulmonary vascular resistance. The results reported by Fiore and associates for surgical repair in adults described survival rates that were significantly higher than in nonsurgically treated patients. If the pulmonary vascular resistance is significantly elevated, the operative risk is increased and long-term benefit is less compared with early repair. Early operation during childhood is indicated whenever feasible.

Incomplete Atrioventricular Septal Defects

The terms *incomplete atrioventricular septal defect*, *incomplete atrioventricular canal defect*, and *ostium primum atrial septal defect* are used interchangeably. These defects are relatively uncommon, accounting for 4 to 5 percent ASDs. Associated abnormalities include unroofed coronary sinus, patent ductus arteriosus, persistent left superior vena cava, coarctation of the aorta, and left ventricular outflow tract obstruction. Approximately 20 percent of patients with incomplete atrioventricular septal defect have Down syndrome. Except for this unusual association, no etiologic factors are known.

Pathology and Pathophysiology

The two significant anatomic defects are a cleft in the anterior leaflet of the mitral valve and a low, crescent-shaped defect in the lower atrial septum (Fig. 17-16). The cleft in the anterior leaflet of the mitral valve may be partial, extending for a short distance from the ventricular septum, or complete, separating the entire anterior leaflet into halves. Chordae tendineae usually are attached to the margins of the cleft, constituting a “trileaflet valve,” with little or no mitral insufficiency. In other patients significant mitral insufficiency occurs. As described by Perloff, failure of the endocardial cushions to fuse produces a configuration in which the atrioventricular valves are in an abnormally low position and the aortic valve is in an abnormally high anterior position, resulting in the characteristic “gooseneck” deformity seen on the left ventricular angiogram. In extreme cases this may lead to left ventricular outflow tract obstruction. The endocardial cushion defect in the septum results in inferior displacement of the conduction system, which produces a characteristic left-axis deviation on the electrocardiogram.

The physiologic abnormalities are a left-to-right shunt combined with mitral insufficiency. When mitral insufficiency is minimal, the physiologic abnormality is identical to that of a large atrial septal defect of the ostium secundum type. When mitral insufficiency is moderate or severe, left ventricular failure and pulmonary hypertension appear early in life and produce a much more severe impairment of cardiac function than is seen in ostium secundum–type septal defects. An increase in pulmonary vascular resistance with pulmonary hypertension may develop but is more frequent with complete atrioventricular septal defects.

Clinical Manifestations

When mitral insufficiency is minimal, the clinical picture is similar to that of ostium secundum atrial septal defect. With significant mitral insufficiency, cardiac failure with pulmonary congestion and dyspnea may be fatal in the first year of life unless surgically corrected.

The chest radiograph usually shows a moderate cardiac enlargement, involving the right and left ventricles. Increased pulmonary vascularity is common. The electrocardiogram is useful in typically showing a left-axis deviation with counterclockwise rotation, compared with right-axis deviation in patients with ostium secundum defects. The echocardiogram shows classic abnormalities and demonstrates any valvular insufficiency. The echocardiogram is sufficient to confirm the diagnosis, and cardiac catheterization is performed only if significant pulmonary hypertension is present and the degree of reversibility is in question.

Treatment

In most patients, operative correction of the defect should be performed between the ages of 1 and 4 years. When pulmonary hypertension and cardiac failure or significant valvular insufficiency are present, operation may be necessary in infancy. Rarely, an adult is seen in the fourth or fifth decade without symptoms of heart failure and an unsuspected incomplete atrioventricular septal defect. These patients are those in whom mitral insufficiency has been minimal, with a clinical course similar to an ostium secundum–type of ASD. In most patients the combination of a left-to-right shunt and mitral insufficiency results in progressive cardiac enlargement and pulmonary congestion during childhood.

Operative Technique

Operation is performed through a median sternotomy or a small right anterior thoracotomy incision, with the patient on extracorporeal circulation. The operative objectives are correction of the mitral insufficiency and closure of the atrial septal defect, while taking care to avoid heart block from injury to the conduction bundle along the posterior margin of the defect.

The right atrium is widely opened, and the crescent-shaped septal defect is examined. The cleft in the anterior leaflet of the mitral valve is inspected carefully, as is the underlying ventricular septum, ruling out an unsuspected complete atrioventricular defect.

The authors prefer to close the cleft in the anterior mitral valve leaflet with three or four interrupted figure-of-eight sutures (Fig. 17-17). Although some surgeons do not close the cleft, routine closure lowers the risk of late postoperative valvular insufficiency. After repair of the cleft in the mitral leaflet, the atrial septal defect is repaired, usually with a patch of pericardium. Along the posterior rim of the defect, near the apex of the triangle of Koch, the penetrating conduction bundle is very superficial. To avoid heart block, the sutures are inserted superficially and leftward along the annulus of the mitral valve, or well rightward of the coronary sinus, redirecting coronary sinus flow into the left atrium. Both patch techniques are successful in minimizing the risk of heart block. Repair of the tricuspid valve usually is not necessary.

Results

The operative mortality is 1 to 2 percent for the group of patients without additional risk factors. Factors that increase the operative risk are hypertension, mitral insufficiency, and left ventricular outflow tract obstruction. If repair with adequate correction of valvular insufficiency is accomplished before the development of severe pulmonary hypertension, the recovery usually is uneventful, similar to that for ostium secundum ASD. The risk of heart block is approximately 1 to 2 percent. Late complications include recurrent mitral

insufficiency, which occurs in less than 10 percent of patients, with less than 5 percent of patients requiring subsequent reoperation.

The functional results and late survival rates are excellent in the majority of patients. King and associates reported long-term results in 199 patients, the majority of whom had maintained an excellent result after operation. Late survival was 96 percent at 20 years, which matched the general population.

Complete Atrioventricular Septal Defects

Pathology and Pathophysiology

Complete atrioventricular septal defect also is referred to as *complete atrioventricular canal defect* or *complete endocardial cushion defect*. The malformation results from failure of fusion of the endocardial cushions in the central portion of the heart, causing a large defect involving the atrial and ventricular septum (Fig. [17-18](#)). The central portion of the annulus between the mitral and tricuspid valves also fails to form, creating a single, six-leaflet atrioventricular valve. Down syndrome is present in approximately 80 percent of patients with complete atrioventricular septal defect.

The variations in valvular deformity in patients with complete atrioventricular septal defect was described by Rastelli in 1966. The Rastelli classification divides the deformity into three groups, types A, B, and C, on the basis of the presence of clefts or septal attachments in the bridging superior common leaflet. Since the posterior leaflets also can have varying pathology, it is more practical to visualize the common atrioventricular valve as a six-leaflet structure, overlying a large septal defect that involves the ventricular and the atrial septum. The bridging and chordal attachments of the various leaflets are described separately. When all components are present but the VSD is restrictive, the term *intermediate or transitional atrioventricular septal defect* is used. When the single atrioventricular valve overrides one ventricle by more than 50 percent, with underdevelopment of the other ventricle, the lesion is termed *unbalanced*. When the defect is severely unbalanced, a biventricular repair may not be feasible, and the patient is treated as having single-ventricle physiology. The physiologic defect is a large left-to-right shunt at the atrial and ventricular levels, resulting in pulmonary hypertension and cardiac failure in infancy. The severity of the symptoms depends on the extent of insufficiency of the atrioventricular valve and the size of the VSD component. When the VSD is large or when atrioventricular valvular insufficiency is severe, the clinical course is a malignant one, with heart failure, pulmonary hypertension, and death within 6 months without operative correction. More than one-half of the children with complete atrioventricular septal defect die by 1 to 2 years of age without treatment.

The diagnosis is established principally by echocardiography. Cardiac catheterization should also be done for evaluation of patients with severe pulmonary hypertension and elevated pulmonary vascular resistance, particularly if the child is more than 6 months of age or if other associated lesions are suspected. The catheterization shows a typical “gooseneck” deformity in the left ventricular outflow tract.

Treatment

The use of pulmonary artery banding to diminish pulmonary blood flow in patients with complete atrioventricular septal defect, previously a common practice, now is uncommonly used except in children with extremely unbalanced defects in whom a biventricular repair is not possible. Early complete repair is recommended for most patients, usually before 6 months of age. If possible, repair is done electively before the development of severely increased pulmonary vascular resistance or significant atrioventricular valvular insufficiency. If the patient develops significant congestive heart failure early in life, repair should be performed earlier, at 2 to 3 months of age.

Operative Technique

The operative procedure consists of placement of a prosthetic patch to correct the underlying ventricular septal defect (Fig. [17-19 A](#)), reattachment of the atrioventricular valve leaflets in such a way as to create separate left-sided and right-sided valves (Fig. [17-19 B](#)), and closure of the ostium primum atrial septal defect component (Fig. [17-19 C](#)). Care is taken to avoid injury to the conduction bundle, as in repair of incomplete atrioventricular septal defect. The repair may be performed with a single-patch technique, using a single piece of Dacron for the VSD and ASD components, or with a two-patch technique, using separate patches for the VSD and ASD components. The authors prefer the two-patch technique because it allows more precise correction of any atrioventricular valvular insufficiency and may be associated with less late leaflet dehiscence, since the leaflet tissue is not divided. Excellent results also are obtained with the single-patch method.

Results

The operative mortality in repair of complete atrioventricular septal defect is 3 to 5 percent. Factors that increase the risk are the presence of significant preoperative atrioventricular valvular insufficiency, elevated pulmonary vascular resistance, unbalanced anatomy, and associated lesions such as left ventricular outflow tract obstruction or tetralogy of Fallot.

The authors' experience at the New York University Medical Center includes more than 70 patients undergoing complete repair over 15 years, with an operative mortality of less than 5 percent. Many of these patients had increased pulmonary vascular resistance and moderate-to-severe valvular insufficiency preoperatively. At late follow-up the New York Heart Association (NYHA) functional status of 75 percent was class I or II, and 95 percent were free of late valvular insufficiency.

Bando and associates reported on their experience from 1974 to 1995 using the two-patch technique with 203 patients. The operative mortality decreased substantially over the 20-year period, averaging 3 percent for the post-1990 period; the 10-year survival rate was 91.3 percent. After repair, the left-sided valve remained competent in 94 percent of the patients on late echocardiographic follow-up. Similar results were reported by Hanley and associates in 301 patients operated on at Boston Children's Hospital using the single-patch technique, with an operative mortality of 3 percent after 1987 and a late risk of recurrent valvular insufficiency of 7 percent.

CYANOTIC LESIONS

Any intracardiac defect producing right-to-left shunting of unoxygenated venous blood results in systemic oxygen desaturation and cyanosis. Typical cyanotic conditions, such as tetralogy of Fallot, pulmonary atresia, and tricuspid atresia, have significantly decreased pulmonary blood flow. These patients may require a palliative systemic-to-pulmonary shunt to improve pulmonary blood flow if early total repair is not feasible. Other conditions, such as transposition of the great vessels, double-outlet right ventricle, and truncus arteriosus, have the unusual combination of increased pulmonary blood flow and moderate cyanosis.

Palliative Shunts

The concept of improving pulmonary blood flow surgically for patients with cyanotic heart disease was introduced in 1944. Blalock and Taussig demonstrated that dramatic improvement in oxygenation could be obtained in patients with tetralogy of Fallot by anastomosis of the subclavian artery to the pulmonary artery to create an artificial arterial systemic-to-pulmonary shunt. The operation was developed after Taussig had observed that symptoms worsened in tetralogy of Fallot patients whenever a patent ductus arteriosus spontaneously closed. The procedure, which was termed the Blalock-Taussig operation, represents one of the milestones in cardiac surgery. In 1946 Potts described another systemic-to-pulmonary shunt, a direct anastomosis between the descending aorta and the left pulmonary artery. In 1954 Glenn described an anastomosis of the superior vena cava to the right pulmonary artery, and in 1962 Waterston described a technique for a direct shunt or anastomosis between the ascending aorta and the right pulmonary artery. The Potts and Waterston shunts are no longer used, as they tend to overperfuse the lungs; in addition, the Waterston shunt tended to distort the pulmonary artery, and the Potts shunt was difficult to take down.

The classic Blalock-Taussig (BT) shunt is a direct end-to-side anastomosis between the subclavian artery and the pulmonary artery, which is performed on the side opposite the aortic arch to avoid kinking of the subclavian artery. The classic BT shunt has the advantage of providing excellent long-term palliation, because the vessels grow with the patient. It rarely is used today, however, except occasionally in older patients when extended palliation is required. The most widely used arterial systemic-to-pulmonary shunt is the modified Blalock-Taussig shunt (Fig. [17-20](#)). The modified BT shunt involves placement of an interposition graft, usually of 4-mm or 5-mm expanded polytetrafluoroethylene (Gore-Tex), between the subclavian artery and the pulmonary artery. The modified BT shunt can be performed more easily in neonates than the classic BT shunt and provides excellent short-term (1 to 2 years) palliation.

Another option for short-term palliation is a central systemic-to-pulmonary shunt, performed through a sternotomy, using a 3.5-mm or 4-mm Gore-Tex graft from the ascending aorta or the innominate artery to the pulmonary artery. The central shunt is preferred by some because it is technically easy to perform and produces minimal distortion of the pulmonary arteries.

The classic Glenn shunt is a direct end-to-side anastomosis between the divided right pulmonary artery and the side of the superior vena cava with ligation of the cavoatrial junction, which provides a passive venous shunt into the isolated right pulmonary artery. This was most widely used for long-term palliation in patients with tetralogy of Fallot who were poor candidates for total repair and for patients with tricuspid atresia or single-ventricle physiology before widespread use of the Fontan procedure. The Glenn shunt is not feasible early in life, until 3 to 6 months of age, because the pulmonary vascular resistance must drop before passive pulmonary flow is possible. This shunt improves oxygenation while minimizing the risks of long-term volume overload on the left ventricle that are inherent in systemic-to-pulmonary shunts. The most significant complications of the Glenn shunt are incomplete correction of cyanosis and the development of significant intrapulmonary arteriovenous fistulas in the right lung, leading to hemoptysis. The classic Glenn shunt is now used infrequently.

A newer, more physiologically appealing venous-to-pulmonary shunt that has been used with increasing frequency in recent years is the bidirectional Glenn shunt or hemi-Fontan procedure. The bidirectional Glenn shunt is an end-to-side anastomosis between the divided superior vena cava and the pulmonary artery (Fig. [17-21](#)), allowing “bidirectional” flow of unoxygenated superior vena caval blood into both lungs. The procedure is performed using cardiopulmonary bypass or with use of a passive temporary cavoatrial shunt. The bidirectional Glenn shunt improves oxygenation without overperfusion of the lungs, minimizing the risks of pulmonary hypertension and left ventricular volume overload. These advantages are similar to those of the classic Glenn shunt, yet both lungs remain perfused and the patient is better prepared for a late modified Fontan procedure. The bidirectional Glenn shunt has become an acceptable intermediate-stage palliative procedure for patients with tricuspid atresia, single-ventricle complex, or hypoplastic left heart syndrome and for patients with pulmonary atresia and intact ventricular septum who are not candidates for biventricular repair. Jonas has reported excellent results using this approach in patients with hypoplastic left heart syndrome after the stage 1 Norwood operation.

Palliative treatment for patients with tetralogy of Fallot with diminutive pulmonary arteries also has changed significantly. Several groups have proposed transcatheter balloon dilation of the pulmonary valve and arteries combined with coil occlusion of large systemic-to-pulmonary collateral vessels. This catheter-based approach has been used instead of early surgical systemic-to-pulmonary shunting in some patients in an attempt rehabilitate the pulmonary arteries and improve the chances of successful total correction. Early results with palliative transcatheter balloon palliation were reported by Kreutzer, demonstrating increased annular size and significant pulmonary artery growth, which simplifies the subsequent total repair.

The use of palliative systemic-to-pulmonary shunts, venous-to-pulmonary shunts, and other forms of palliation such as transcatheter balloon dilation, has had a highly successful history and has an important role in the treatment of patients with cyanotic congenital heart disease. These procedures increase pulmonary blood flow, improve oxygenation, and induce pulmonary artery growth. Use of intermediate palliative procedures in certain patients lowers the risks associated with subsequent total repair. While early total correction is

the norm in the current cardiac surgical era, newer forms of surgical and catheter-based palliation are selectively used in an attempt to modify pulmonary vascular growth and increase the chances of successful biventricular repair. In patients with single-ventricle physiology and decreased pulmonary blood flow, the bidirectional Glenn shunt is used increasingly as an intermediate palliative step before the modified Fontan procedure. Palliative operations have an important role in the surgical treatment of cyanotic congenital heart disease.

Tetralogy of Fallot

Tetralogy of Fallot was described in 1673 by Steno, but it became well known in 1888 when Fallot emphasized the combination of abnormalities that are regularly present; subsequently it was known as the tetralogy. Effective therapy first became possible in 1944 with the Blalock-Taussig shunt previously described. With the development of extracorporeal circulation, total correction of tetralogy of Fallot became possible and was first performed by Lillehei in 1954. The first repair was done using “cross- circulation” with another human being serving as the pump-oxygenator. The first repair using the “heart-lung machine” pump-oxygenator was done in 1955 at the Mayo Clinic. Tetralogy of Fallot is the most common cyanotic malformation, constituting over 50 percent of all cases of cyanotic heart disease. There are no known causes.

Pathology and Pathophysiology

The four features of tetralogy of Fallot are: (1) malalignment ventricular septal defect, (2) dextroposition of the aorta, (3) right ventricular outflow tract obstruction, and (4) right ventricular hypertrophy; all are the result of a specific developmental abnormality, namely, underdevelopment and anterior malalignment of the infundibular septum. In normal embryonic development, the infundibular septum fuses inferiorly with the muscular septum. In tetralogy of Fallot, the infundibular septum deviates anteriorly and cephalad, creating the large ventricular septal defect at the point of nonunion. The anterocephalad deviation of the septum not only narrows the right ventricular outflow tract but also allows the aortic root to “override” the ventricular septum in a rightward direction, producing the malalignment ventricular septal defect (VSD). The nonrestrictive VSD results in systemic pressures in the right ventricle, while concentric right ventricular hypertrophy results from obstruction of the right ventricular outflow tract.

The VSD usually is a large, malalignment-type, perimembranous defect. The aortic cusps often are visible through the defect, depending on the degree of “overriding.” The standard anatomic details regarding the conduction tissue are of concern during repair of the VSD.

The pulmonary valve leaflets are often thickened, with fused commissures and tethering of one or more leaflets with resulting limited motion (Fig. 17- 22). The anatomic location of the right ventricular outflow tract obstruction may vary from isolated infundibular stenosis to diffuse hypoplasia of the right ventricular outflow tract to pulmonary atresia. Discrete infundibular stenosis is present in about 15 percent of the patients and complete pulmonary atresia in about 20 percent. Most patients have a combination of infundibular outflow tract obstruction and pulmonary annular hypoplasia (see Fig. 17-22). Pure valvular stenosis is unusual and may represent a different embryologic lesion, namely, pulmonary stenosis with ventricular septal defect rather than tetralogy of Fallot.

In patients with markedly diminished pulmonary blood flow and diminutive pulmonary arteries there often is striking enlargement of the bronchial arteries, which are referred to as aortopulmonary collateral arteries. These vessels compensate for diminished pulmonary blood flow but may result in significant varicosities throughout the mediastinum and chest wall, which complicates final repair.

Anomalous coronary arteries are found in 5 percent of patients with tetralogy of Fallot. The anomalous origin of the left anterior descending artery from the right coronary artery, which crosses in the outflow tract, is of particular surgical significance. A right aortic arch occurs in about 25 percent of the patients. An atrial septal defect is found in 10 to 15 percent of the patients—a set of anomalies that is sometimes called *pentalogy* of Fallot.

Patients with tetralogy of Fallot have decreased pulmonary blood flow, cyanosis secondary to right-to-left shunting, and systemic right ventricular pressures with right ventricular hypertrophy. The inability to increase pulmonary blood flow results in severe intolerance to exercise.

The severity of right-to-left shunting and hypoxia is dependent on the degree of outflow tract obstruction and the limitation in pulmonary blood flow. Arterial oxygen saturations of 70 to 85 percent are seen in older children with more balanced lesions; in younger children with severe cyanosis, saturations of 20 to 35 percent may be seen. There often is a precipitous fall in oxygen saturation with exercise.

Clinical Manifestations

Polycythemia seldom develops until after 2 years of age, but later in life patients with uncorrected tetralogy of Fallot develop hematocrit levels as high as 60 to 75 percent. The degree of cyanosis increases significantly in the first few years of life, because visible cyanosis is proportional to the amount of unsaturated hemoglobin in the peripheral circulation. Patients who are cyanotic at birth usually have markedly diminished pulmonary blood flow or pulmonary atresia and usually do not survive infancy unless operation is performed. Patients who become cyanotic in the first year of life have a milder course but usually are symptomatic early. A few patients develop only mild cyanosis later in life—the so-called pink tetralogy. These patients have only moderate reduction in pulmonary blood flow because of a lesser degree of outflow tract obstruction. Chronic hypoxia results in polycythemia and eventual clubbing of the digits.

Most untreated patients with tetralogy of Fallot die from progressive hypoxia with repeated cyanotic “spells” resulting in cardiac arrest, cerebral injury, pulmonary thrombosis, or infection. About 25 percent of untreated patients die in the first year of life, 40 percent by 3 years, and 70 percent by 10 years. Brain abscess is a serious, often lethal, late complication, though it is seen infrequently. The most plausible explanation for the increased susceptibility to brain abscess is the right-to-left shunt, which bypasses the lungs, providing bacteria in the venous blood with direct access to the arterial circulation.

Most patients are symptomatic, with exertional dyspnea and cyanosis the outstanding features. Cyanotic spells are the main risk during infancy. When a spell occurs, the infant becomes deeply cyanotic and comatose. Spontaneous recovery usually occurs, but cardiac arrest or neurologic injury may ensue. Total repair or a palliative shunt, depending on the anatomy of the pulmonary outflow tract and pulmonary arteries, should be performed promptly in these infants.

Squatting is a classic clinical manifestation of tetralogy of Fallot. The physiologic benefit from squatting probably derives from an increase in the systemic vascular resistance, which diminishes right-to-left shunting and forces more blood into the lungs. Walking for short distances interrupted by squatting is a common finding in children with tetralogy of Fallot. Hemoptysis is rare, occurring primarily in children with long-standing cyanosis and significant systemic-to-pulmonary collateral vessels.

On physical examination, the obvious features are cyanosis of varying severity and clubbing of the digits. The heart usually has a normal size, rate, and rhythm. A systolic murmur of grade II to III is commonly present along the left sternal border at the third or fourth intercostal space, and in about one-half the patients it is accompanied by a thrill. With severe pulmonic stenosis or pulmonary atresia, the murmur may be faint or absent because of absence of flow through the pulmonic orifice. The second pulmonic sound is weak or absent, and the aortic second sound is increased.

The chest radiograph shows a normal-sized heart with an unusual boot- shaped contour (the so-called *coeur en sabot* form). This appearance results from the combination of diminutive, concave pulmonary artery segment, a horizontal ventricular septum produced by concentric right ventricular hypertrophy, and a small left ventricle. The pulmonary vascular markings are markedly decreased.

The electrocardiogram usually shows right ventricular hypertrophy of varying severity with right axis deviation. The echocardiogram is almost always diagnostic, demonstrating the size of the *VSD*, the degree of aortic overriding, the extent and degree of right ventricular outflow tract obstruction, and the size of the central and peripheral pulmonary arteries. The most important data provided by cardiac catheterization are more precise details of the pulmonary annular size, the branch pulmonary arteries, and the coronary artery anatomy. These are primarily of interest when the anatomy is marginal for total repair.

Treatment

Infants with tetralogy of Fallot who develop early cyanotic spells and are found to have hypoplastic or discontinuous pulmonary arteries and those who have pulmonary atresia frequently require emergent placement of a systemic-to-pulmonary shunt for palliation. An alternate approach in some patients with underdeveloped pulmonary arteries is palliative balloon dilation of the pulmonary valve or pulmonary vessels in an attempt to prepare the pulmonary vascular bed for early total repair. For the vast majority of patients with tetralogy of Fallot, early total correction is the treatment of choice. Corrective operation can be performed at any time if the child's symptoms are worsening and the pulmonary vascular bed is adequate. For children with minimal symptoms, elective repair is recommended at 6 to 12 months of age.

Operative Technique

Repair is performed through a median sternotomy with extracorporeal perfusion and cardioplegic arrest. If a previous palliative shunt operation has been performed, the shunt is isolated before perfusion is begun and occluded once cardiopulmonary bypass is initiated. Once the pericardium has been opened the outflow tract of the right ventricle is examined for anomalous coronary arteries; an approach is chosen that will avoid dividing any such arteries.

After the heart is arrested, the *VSD* is approached through the right atrium or through a right ventriculotomy, depending on the anatomy and the surgeon's preference. The authors generally approach the VSD through the right atrium, primarily to minimize the size of the incision in the right ventricle. The VSD patch is placed through the tricuspid valve, similar to the approach in standard perimembranous VSD repair.

Several potential zones of obstruction in the right ventricular outflow tract should be evaluated. While the areas of obstruction usually are determined by preoperative studies, the important zones should be reevaluated at the time of operation. These include the pulmonary valve, the pulmonic annulus, the infundibular outflow tract, the main pulmonary artery, and the distal pulmonary arteries. It is important that the surgeon evaluate each area separately, because it is critical that all areas of stenosis in the right ventricular outflow tract and pulmonary arteries be corrected.

After opening the right atrium and repairing the *VSD*, the infundibular outflow tract is inspected, and the main pulmonary artery is opened just above the valve. The main pulmonary artery and the pulmonary bifurcation, the left and right pulmonary arteries, the pulmonary valve, and the pulmonary annulus are all inspected and calibrated with Hegar dilators. A transannular incision is made only if the pulmonary annulus is hypoplastic. It is important to decide whether the annulus is of an adequate size by comparing the measured diameter to that of normal subjects using a nomogram. If the annular size is adequate and leaflet mobility can be restored, the valve is preserved. In such cases, a separate small ventriculotomy is made below the valve to correct any infundibular stenosis, or, alternatively, resection of the infundibular obstruction is performed through the tricuspid valve. When the annulus is hypoplastic, the incision is extended across the annulus, through the area of infundibular obstruction (see Fig. 17-22 B). The incision is only carried onto the ventricle far enough to correct the infundibular stenosis, while preserving as much right ventricular function as possible. The small transannular incision is then closed with an appropriately sized pericardial or synthetic patch, enlarging the outflow tract as necessary (see Fig. 17-22 C). Tailoring the transannular patch to appropriate size for the patient usually prevents

late aneurysmal formation in the patch unless a distal obstruction is present. When the distal pulmonary arteries are severely stenosed or hypoplastic, a valved pulmonary artery homograft may be required to reconstruct the outflow tract. The use of a valve in the circuit minimizes regurgitation of blood into the right ventricle and lowers the risk of long-term right ventricular dysfunction.

After cardiopulmonary bypass is discontinued, intracardiac pressure is measured to confirm that right ventricular outflow tract obstruction has been corrected. The right ventricular systolic pressure should be reduced to less than 60 to 70 percent of the left ventricular pressure. If right ventricular pressure is still elevated above this level, more adequate correction of the ventricular obstruction should be considered, or a valved graft or homograft should be inserted. The right ventricular pressure should not be left at systemic levels without further intervention, because progressive right ventricular failure ensues. In most patients the ratio of right ventricular to left ventricular pressure is less than 0.50 to 0.60 after successful repair, with a pulmonary artery systolic pressure of 20 to 25 mmHg.

Results

The operative mortality is 2 to 5 percent in most centers. Operative risk factors include age less than 3 months, older age at repair with severe polycythemia, pulmonary atresia, absent pulmonary valve syndrome, and the presence of severely hypoplastic or discontinuous pulmonary arteries. After successful repair the risk of late death is low throughout life. Late complications include recurrent *VSD*, recurrent branch pulmonary artery stenosis, progressive right ventricular dysfunction resulting from pulmonary insufficiency, aneurysm formation in the outflow tract patch, and arrhythmias.

Pulmonary Stenosis/Pulmonary Atresia with Intact Ventricular Septum

Isolated pulmonary stenosis is a relatively common defect, constituting 10 percent of congenital heart defects. More than one-half of these patients have pure valvular pulmonary stenosis, and the remainder have varying degrees of associated right ventricular outflow tract obstruction. A more severe form of the disease process, pulmonary atresia with intact septum, results in total obstruction of the outflow tract with atresia of the valve, hypoplasia of the annulus, and differing degrees of hypoplasia and maldevelopment of the right ventricle. Pulmonary atresia with intact septum is less common than pulmonary stenosis, accounting for 1 to 3 percent of congenital heart defects.

Pathology and Pathophysiology

Pulmonary stenosis and pulmonary atresia with intact septum represent a spectrum of pathologic findings, ranging from isolated valvular stenosis to total valvular atresia with hypoplasia of the inlet, body, and outflow tract of the right ventricle. With pulmonary stenosis, the degree of valvular obstruction may vary from mild, clinically insignificant stenosis to critical pulmonary stenosis with a "pinhole" opening, which produces symptoms similar to pulmonary atresia. The valve usually is bicuspid or tricuspid and domed, with fused commissures, although more severe valvular dysplasia is present. The annulus may be normal in size or hypoplastic.

In patients with true pulmonary stenosis, the right ventricle usually is relatively normal, except for secondary right ventricular hypertrophy. Approximately 40 percent of these children have some degree of infundibular obstruction, and 10 percent have severe infundibular obstruction.

The more severe lesion of *pulmonary atresia with intact ventricular septum (PAIVS)* presents in the newborn period with an atretic valve and no forward blood flow. In *PAIVS* the right ventricle is underdeveloped to varying degrees, although a small inlet and infundibular portions usually are absent. The body of the ventricle often is totally obliterated and dysfunctional. In *PAIVS* the right ventricle is classified according to its degree of development as unipartite, bipartite, and tripartite ventricle, depending on the adequacy of the inlet, body, and outlet portions of the ventricle. Knowledge of the right ventricular development is essential, a minimum of a bipartite ventricle is necessary, and a tripartite ventricle is optimal if the right heart is to be used in the circulation. Otherwise the patient must be treated as having single-ventricle physiology.

In patients with *PAIVS*, the myocardium in the underdeveloped right ventricle exhibits a wide range of pathologic anomalies. Typically the ventricle is extremely small, thickened, and hypertensive, with myocardial muscular disarray. In approximately 10 percent of the patients with *PAIVS*, sinusoids connect the right ventricle with the coronary circulation, which results in a serious condition termed *right ventricular–dependent coronary circulation*.

In patients with *PAIVS* the tricuspid valve is often regurgitant, but a functioning valve usually is present. If the tricuspid valve is severely hypoplastic or totally atretic, the disease becomes tricuspid atresia and the patient is treated as having single-ventricle physiology.

With moderate degrees of pulmonary stenosis the physiologic consequences are slow to develop, and significant symptoms usually do not develop until adolescence. Significant forward pulmonary blood flow often is present, and cyanosis usually does not occur. The obstruction produces an elevation in right ventricular pressure with subsequent ventricular hypertrophy. The severity is determined by the gradient; a gradient of less than 50 mmHg is considered mild, and a gradient of 80 to 100 mmHg or more is severe. With severe stenosis suprasystemic right ventricular pressures often develop. Some patients tolerate high levels of right ventricular hypertension for years, while others develop progressive tricuspid insufficiency, right-sided heart failure, and right-to-left shunting across the foramen ovale with resulting hypoxemia and cyanosis.

In patients with critical pulmonary stenosis and in patients with *PAIVS* most forward pulmonary blood flow may be absent. Mixing at the atrial level is necessary to maintain life, with right-to-left shunting across the foramen ovale. If the foramen ovale is of inadequate size it must be stretched shortly after birth with a balloon catheter, a procedure termed *balloon septostomy*. This rarely is required.

Clinical Manifestations

Hypoxia, cyanosis, and acidosis usually develop shortly after birth when the ductus arteriosus begins to close. Ductal patency must be maintained with prostaglandins to keep the child alive until the diagnosis can be established and more definitive treatment initiated. The electrocardiogram may develop right atrial P waves shortly after birth, but in *PAIVS* the right ventricular hypertrophy pattern often is absent. The chest radiograph shows decreased pulmonary blood flow and a flat or concave pulmonary artery segment with a normal or enlarged heart size.

The diagnosis usually is made by echocardiography, which demonstrates critical stenosis with a domed, stenotic pulmonary valve and an adequate right ventricle, or *PAIVS* with an atretic valve and a poorly developed right ventricle. Close attention is given to evaluating the body and outlet portions of the right ventricle, the pulmonary annular size, and the adequacy of the tricuspid valve, because these determine whether or not a biventricular repair is feasible.

Cardiac catheterization is used primarily for intervention such as balloon dilation in patients with critical pulmonary stenosis. In patients with *PAIVS*, diagnostic catheterization usually is required to further evaluate the internal anatomy of the right ventricle and to determine the presence or absence of sinusoids feeding the coronary circulation.

Treatment

The initial treatment of choice for most infants with critical pulmonary stenosis and for adolescents with moderate to severe isolated pulmonary stenosis is balloon dilation in the catheterization laboratory. Colli and associates reported excellent early results with balloon dilation in 90 percent of neonates with critical pulmonary stenosis. Similarly, McCrindle reported excellent long-term results after valvuloplasty in 77 percent of 533 patients with pulmonary stenosis, with follow-up out to 8.7 years.

Patients with hypoplastic outflow tract obstruction or with *PAIVS* are not candidates for balloon valvuloplasty and require surgical correction. The operative risk is high, particularly in patients with *PAIVS*. Initial correction requires the use of cardiopulmonary bypass and the combination of surgical valvotomy and a systemic-to-pulmonary shunt, or a transannular patch to relieve outflow tract obstruction plus a systemic-to-pulmonary shunt. Weldon and Cobanouglu each reported that although repair is successful in the majority of patients, the operative mortality exceeds 20 percent. Hanley reported results in 101 patients from 27 institutions, noting that use of valvulotomy without cardiopulmonary bypass and transannular patching without placement of a concomitant shunt were predictors of poor outcome.

Tricuspid Atresia

Systemic-to-pulmonary shunts for tricuspid atresia were first used in the 1940s, with good short-term improvement but poor long-term results. A significant contribution was made with the introduction of the Glenn shunt, which gave better long-term palliation. The major surgical advance came in 1968 when Fontan successfully separated the right and left circulations in a patient with tricuspid atresia for the first time, introducing the concept of totally passive pulmonary blood flow. Since then variations of Fontan's approach have been developed, and some form of cavopulmonary connection remains the procedure of choice for tricuspid atresia.

Pathology and Pathophysiology

Tricuspid atresia is an important form of congenital heart disease, affecting 2 to 5 percent of children with cyanotic heart disease. The basic abnormalities are total atresia of the tricuspid valve and ventricular inlet, varying degrees of hypoplasia of the body and outlet portion of the right ventricle, and an atrial or ventricular septal defect. The mitral valve and the left ventricle usually are normal. Pulmonary atresia or severe pulmonary stenosis is seen in 85 percent of the cases with normally located great vessels, further limiting pulmonary blood flow. Frequently the right side of the heart is totally hypoplastic, without inlet, body, or outlet chambers. Blood must mix at the atrial level, and the typical patient has totally ductus-dependent pulmonary blood flow.

The aorta and pulmonary artery are situated normally, in about 70 percent of patients with tricuspid atresia, and transposition of the great vessels is present in about 30 percent. These two major groups may be further subdivided according to whether pulmonary blood flow is unrestricted or decreased. In the typical patient with tricuspid atresia and normally situated great vessels, pulmonary blood flow is markedly decreased, with most flow provided by the patent ductus arteriosus. When the aorta and pulmonary artery are transposed, about 70 percent of the patients have unrestricted pulmonary blood flow with overperfusion of the lungs rather than hypoxia. This condition also may be termed *single-ventricle complex*. The degree of hypoxia typically is severe because of the absence of blood flow through the atretic tricuspid valve combined with restricted or absent flow across the septal defect and through the rudimentary right ventricle into the pulmonary artery.

Clinical Manifestations

Most patients become markedly hypoxic and dyspneic shortly after birth, when the ductus arteriosus begins to close; most will die unless palliative shunting is performed. The correct diagnosis is made during the first day of life in more than one-half of infants with tricuspid atresia. In a minority of patients with unrestricted pulmonary blood flow, with normal or transposed great arteries, congestive heart failure and pulmonary congestion are present.

The physical examination is not diagnostic because the murmur varies widely, depending on the size of the ventricular septal defect and the anatomic relationship of the great vessels. The chest radiograph usually shows decreased vascularity. The electrocardiogram is strongly suggestive, showing a typical left-axis deviation resulting from the underdevelopment of the right ventricle. The echocardiogram can establish the diagnosis with certainty, and it also can outline any atrial septal defect or ventricular septal defect

and the relationships of the great arteries. Cardiac catheterization usually is not necessary early in life before palliative treatment, but it should be performed later before definitive repair.

Treatment

An emergency systemic-to-pulmonary shunt often is necessary in the first few days or weeks of life to prevent death from hypoxia. A modified Blalock- Taussig Gore-Tex shunt usually is the simplest and most satisfactory. In some patients a small foramen ovale restricts mixing at the atrial level, and a balloon septostomy is necessary to increase the right-to-left shunt. Surgical enlargement of the atrial septal defect (Blalock-Hanlon procedure) can be performed if balloon septostomy is unsuccessful, but this is infrequently required.

After palliative shunting, the ultimate repair goal is a modification of the *Fontan procedure*, in which a direct connection between the systemic venous circulation and the pulmonary arteries is constructed. A *bidirectional Glenn shunt procedure* (see Fig. [17-21](#)) often is performed at 3 to 8 months of age. This is coupled with a takedown of a previously performed systemic-to-pulmonary shunt. The bidirectional Glenn shunt minimizes volume overload on the left ventricle and becomes the first step of the final corrective *modified Fontan procedure*. The most frequently used modification of the Fontan procedure is the *total cavopulmonary connection* (Fig. [17-23](#)), which is performed by connecting the superior vena cava to the pulmonary artery with a bidirectional Glenn shunt and connecting the inferior vena cava to the pulmonary artery. The inferior caval connection may be performed in two ways: with an intraatrial lateral Dacron tunnel or with an extracardiac conduit.

The direct *atriopulmonary connection* is a modification of the Fontan procedure that is used less frequently because of a higher incidence of associated late atrial arrhythmias and pleural effusions. This procedure is relatively simple to perform, but the results in terms of flow dynamics and clinical outcome are better with the total cavopulmonary connection methods.

All modified Fontan procedures are delayed until 8 to 12 months of age or later because the pulmonary vascular resistance must drop to normal levels for a procedure dependent on passive pulmonary blood flow to be effective. High-risk patients may have their Fontan connection *fenestrated*, which technically involves placing a small “punch hole” or fenestration between the systemic venous circulation (right atrium) and the left atrium. This allows the patient to shunt right-to-left if the pulmonary vascular resistance temporarily increases.

Results

Cetta and colleagues, reporting results from the Mayo Clinic in 839 consecutive patients who underwent the Fontan operation, compared outcomes in 339 patients from 1987 to 1992 to outcomes in the previous 500 patients, who were operated on from 1973 to 1986. In the more recent cohort the mortality was 9 percent, as compared to the earlier cohort's 16 percent; similarly, the 5-year survival rate improved, with the recent cohort at 81 percent, as compared to the earlier cohort's 73 percent. Sharma, in reviewing 202 patients, noted that the use of a fenestrated total cavopulmonary connection instead of a procedure without fenestration resulted in a decrease in operative mortality from 15.9 percent to 5 percent. The best outcomes after the Fontan operation are achieved when the postoperative right atrial pressure stays below 15 to 18 mmHg. Higher right atrial pressures result in severe problems, such as low output syndrome, recurrent pleural effusions, and protein-losing enteropathy. These complications may occur early or late. Poor results are more common when the pulmonary vascular resistance is elevated, when the pulmonary arteries are deformed, or when preexisting conditions result in an elevated left ventricular diastolic pressure. Other late complications include atrial flutter and thrombus of the right atrium. The use of the total cavopulmonary connection method as compared with atriopulmonary connection or other Fontan modifications appears to lower the incidence of late atrial flutter.

Since the Fontan operation physiologically depends on using the venous pressure to perfuse the pulmonary vascular bed, the patient's cardiac physiology is markedly abnormal even after repair. Most patients have abnormal exercise capacity and altered hemodynamics. While the concept of passive pulmonary blood flow with separated circulations represents a major breakthrough in the treatment of patients with tricuspid atresia or single-ventricle physiology, the long-term results are limited in terms of survival and cardiac functional recovery.

Fontan and Kirklin, in 1990, reviewed 334 patients who had undergone the Fontan procedure. The variables predicting survival were identified, after which each optimal condition was used to predict outcome after a “perfect” Fontan operation. With optimal conditions in each category, operative mortality was 8 percent, and the 5-year, 10-year, and 15-year survival rates were 86 percent, 81 percent, and 73 percent, respectively. The late hazard or instantaneous risk of death began to increase after 6 years, and the average New York Heart Association (NYHA) functional status progressively decreased after operation. These data suggest that the late outcome after a Fontan operation is imperfect, presumably because of limitations imposed by the abnormal physiology present. Families of patients undergoing this procedure should be counseled appropriately so that long-term expectations are realistic. The procedure still offers the best chance for long-term survival for patients with tricuspid atresia.

Single-Ventricle Complex

Single-ventricle complex or univentricular heart represents a variety of complex defects in which there is only a single functioning ventricular chamber. This category includes patients with single atrioventricular connection and differing ventricular morphology, mitral atresia with unrestricted aortic outlet, unbalanced atrioventricular septal defects, and a variety of other anatomic configurations in which the basic abnormality is the presence of a single functioning ventricle. Other variations include the type of functioning ventricular chamber (morphologically a “left” or a “right” ventricle); the type of hypoplastic ventricular chamber; different abnormalities in the atrioventricular valves; and the origin of the aorta and pulmonary artery from the normal or hypoplastic ventricular chambers.

The most frequent physiologic abnormalities are hypoxia from the mixing of oxygenated and unoxygenated blood in the single ventricle and pulmonary congestion from unrestricted pulmonary blood flow. The degree of cyanosis varies with the pulmonary blood flow, which depends on which chamber the pulmonary artery arises from and on the degree of anatomic obstruction in the outflow tract to the pulmonary artery.

Clinical Manifestations

Some patients with single-ventricle complex have unrestricted pulmonary blood flow. The physiology resembles that of a large ventricular septal defect but with some degree of cyanosis because of the complete mixing of blood. At the other extreme, severe hypoxia and cyanosis are present early after birth in patients with single-ventricle complex and severely restricted pulmonary blood flow. These patients often require an emergency palliative shunt in infancy. In between these two extremes are patients with moderate degrees of restriction in pulmonary blood flow, with a $Q_p:Q_s$ of approximately 2:1. Such patients do reasonably well in the first few years of life because of the balanced circulation, which provides adequate pulmonary blood flow but does not flood the lungs.

Treatment

Two extremes often are present in the neonatal period; unrestricted pulmonary blood flow with severe pulmonary congestion, which requires urgent pulmonary artery banding, or severe cyanosis and hypoxia, which requires emergency placement of a palliative shunt. After palliative treatment early in life, the final planned surgical correction involves the use of some modification of the Fontan procedure, usually total cavopulmonary connection. The treatment strategy for most patients with single-ventricle complex involves the placement of a bidirectional Glenn shunt at 3 to 6 months of age, as an intermediate step, taking down the pulmonary artery band or previously placed palliative shunt. When the patient is 12 to 18 months of age the bidirectional Glenn shunt is converted to a modified Fontan total cavopulmonary connection, which often is fenestrated if the patient remains high risk.

Historically the results of the modified Fontan procedure are worse when done for single-ventricle complex than when done for tricuspid atresia, with an operative mortality of 10 to 15 percent and a 10-year survival rate of 60 to 70 percent. The results may be slightly better with the staged treatment strategy.

Ebstein's Anomaly

Ebstein's anomaly is uncommon, representing 0.5 percent of all cases of congenital heart disease. There is a nearly 400 times increased frequency of Ebstein's malformation when the mother has taken lithium during the pregnancy. This is challenged by recent epidemiologic studies.

Pathology and Pathophysiology

The basic abnormality is a malformation of the septal and posterior leaflets of the tricuspid valve. The origin of the leaflets is displaced downward to a variable degree, creating a third chamber on the right side of the heart. Both the leaflet tissue and its chordae also are abnormal. The anterior tricuspid leaflet usually is normal but may be unusually large and prominent, and described as “sail-like.” The segment of right ventricular wall between the true annulus of the tricuspid valve and the origin of the displaced leaflets becomes functionally part of the right atrium and has been termed the *atrialized ventricle*. There is a varying degree of hypoplasia of this segment, and in some patients it resembles a true aneurysm that bulges paradoxically. In most patients the atrialized segment has some muscle fibers with little paradoxical motion. The distal functioning right ventricle is small. Some investigators have believed that there is a true deficiency in the right ventricular fibers as well, which contributes to the right ventricular dysfunction in this condition. A patent foramen ovale or ostium secundum defect is almost always present, with right-to-left shunting. The right atrium usually is dilated because of the obstructive nature of the valve, often to a huge size in older patients.

The malformation varies widely in severity, ranging from relatively minor valvular abnormalities to virtual atresia of the valve leaflets with severe regurgitation. The valve and right ventricle also are functionally obstructive in many cases, impeding forward blood flow.

The main physiologic disturbances are restricted pulmonary blood flow, inadequate cardiac output from right ventricular dysfunction, tricuspid insufficiency, and right-to-left shunting at the atrial level. The severity of symptoms varies greatly, depending on the severity of the pathologic abnormality. A variety of arrhythmias commonly occur, with supraventricular tachycardias occurring frequently, and Wolff-Parkinson- White syndrome with accessory pathways and preexcitation occurring in approximately 5 percent of patients with Ebstein's anomaly. Cyanosis of moderate degree occurs in at least 50 percent of patients because of a right-to-left atrial shunt. The degree of cyanosis gradually becomes more severe in older patients with progressive right ventricular failure.

Clinical Manifestations

A significant proportion of infants with Ebstein's anomaly present in the first month of life with tachypnea and cyanosis, probably a manifestation of the elevated pulmonary vascular resistance in the neonatal period. About one-half of patients who are severely symptomatic in the first month of life subsequently die. After the first month, the severity of symptoms decreases, often with loss of cyanosis, and hence clinical disability during childhood often is minimal. A mortality of about 15 percent has been estimated between the ages of 1 and 20 years. The onset of clinical symptoms in patients surviving childhood is a gradual one; the average age of diagnosis is in the midteens. Many adults continue to function reasonably well with milder forms of Ebstein's anomaly, depending on the presence of arrhythmias, cyanosis, and cardiac failure. A few patients have lived to beyond 70 years of age, but only about 5 percent of all patients live beyond 50 years.

A variety of systolic and diastolic murmurs are present, though at one time it was stated that auscultatory findings were highly suggestive, emphasizing a slow cardiac rate with a triple or quadruple rhythm, a systolic murmur of tricuspid regurgitation, and often a low-pitched diastolic murmur. The chest radiograph usually shows cardiac enlargement because of the large right atrium and the atrialized right ventricle. The vascularity in the lung fields usually is decreased. Electrocardiographic abnormalities are considered typical, with conduction disturbances, a prolonged P-R interval, and partial right bundle branch block. When Wolff-Parkinson-White syndrome is present the P-R interval is short and a delta wave is present.

The echocardiogram is diagnostic, outlining the different abnormalities with precision. Cardiac catheterization may be considered in order to evaluate the size and function of the right ventricle, but it must be performed carefully, as fatal arrhythmias have occurred. Electrophysiologic studies with radiofrequency ablation of the aberrant pathway are indicated if Wolff- Parkinson-White syndrome is present.

Treatment

Various forms of valve reconstruction have been described, with varying results, depending on the degree of the anatomic abnormality in the valve. Previously, valve replacement was almost always required, and replacement frequently is necessary in patients with severe valvular deformity. Danielson described experience with 72 patients, in whom valve reconstruction was performed in 81 percent and valve replacement in 19 percent. The atrial septal defect also is routinely closed. Plication of the atrialized ventricle is usually unnecessary but may be used when the atrialized segment is extremely thin and contracting paradoxically. The overall operative mortality in the Mayo Clinic series was 7 percent, and 87 percent of the patients improved to New York Heart Association (NYHA) functional class I or II at 5 years.

Transposition of the Great Arteries

The clinical syndrome of *transposition of the great arteries (TGA)* was described by Taussig in 1938. In 1948 Blalock and Hanlon proposed surgical atrial septectomy to create unrestricted mixing at the atrial level. Because of the excellent results now obtained with balloon septostomy, developed by Rashkind in 1969, surgical creation of an atrial septal defect is rarely necessary. Total correction of transposition by redirecting blood flow at the atrial level was based on work by Arbert in 1955. Senning, in 1957, first completely corrected transposition of the great vessels by repositioning the atrial septum to redirect caval blood through the mitral valve and pulmonary venous blood into the tricuspid valve. In 1964 Mustard developed a method of reconstructing the atrial cavity and redirecting vena caval blood into the mitral valve with an intraatrial baffle. The technique significantly improved the results for correction of transposition of the great vessels and was adopted by most centers as the preferred procedure throughout the 1960s and 1970s, although some continued to favor the Senning procedure.

In 1975 Jatene performed the first successful arterial switch procedure for correction of transposition with coronary artery transfer and reimplantation. This technique had potential long-term advantages since it used the anatomic left ventricle in the systemic circulation and should not have been prone to ventricular dysfunction or significant atrial arrhythmias. Results improved significantly with the arterial switch operation in the 1980s. The arterial switch operation is considered the procedure of choice for the treatment of transposition.

TGA is a frequent disorder, representing 5 to 8 percent of all congenital cardiac malformations and accounting for about 25 percent of deaths caused by congenital heart disease in the first year of life. It is about four times more frequent in males than in females. The cause is unknown.

Pathology and Pathophysiology

TGA results from abnormal division of the bulbar trunk in embryologic development, occurring between the fifth and the seventh uterine week. With transposition, the aorta originates from the right ventricle and the pulmonary artery from the left ventricle (Fig. 17-24). As a result, venous blood returning through the vena cava to the right atrium enters the right ventricle and is ejected directly into the aorta. Oxygenated blood returning from the lungs through the pulmonary veins to the left atrium enters the left ventricle and is expelled through the pulmonary artery back to the lungs. This dual, parallel circulatory arrangement is incompatible with life if there is no communication between the pulmonary and systemic circulations. A patent ductus arteriosus, a patent foramen ovale, or a ventricular septal defect must be present for the patient to survive. A patent ductus is present for a few weeks after birth in more than one-half of the patients. A patent foramen ovale also is frequent, and a ventricular septal defect occurs in approximately 50 percent of patients (TGA with VSD).

Other associated anomalies are common. One of the most frequent, left ventricular outflow tract obstruction resulting in pulmonic stenosis, occurs frequently enough to constitute a well-defined variant of the syndrome for which the prognosis is unusually favorable. A wide variety of other anomalies may occur, including coarctation of the aorta, pulmonary atresia, and dextrocardia.

The two basic physiologic derangements in patients with TGA are hypoxia and progressive pulmonary congestion from unrestricted pulmonary blood flow. Severe and rapidly progressive cardiac failure results, mostly from the unrestricted pulmonary blood flow but also because the coronary arteries are filled with unoxygenated blood, which results in myocardial hypoxia. The severity of the hypoxia and the degree of cyanosis vary with pulmonary blood flow and the nature of any VSD present. Patients with TGA and intact ventricular septum and patients with TGA, VSD, and LVOT obstruction develop severe cyanosis. In patients with TGA and intact ventricular septum, a Rashkind balloon septostomy may be required to ensure adequate mixing.

TGA is a lethal condition if untreated. Patients with TGA and intact ventricular septum, depending on the size of the foramen ovale, die most rapidly—30 percent in 1 week, 50 percent in 1 month, and 90 percent by 1 year. Those with TGA and VSD rapidly develop pulmonary vascular changes because of increased pulmonary flow, often within a few months, and 80 percent die within the first year.

Clinical Manifestations

A high proportion of infants with TGA are cyanotic at birth. More than 90 percent are recognized the first day of life. Pulmonary congestion with cardiac failure is similarly frequent. The combination of cyanosis and increased pulmonary blood flow in a newborn is highly suggestive of TGA with VSD. The most prominent symptoms are cyanosis and dyspnea.

Cyanosis usually is an obvious clinical finding. It is often severe. Signs of pulmonary congestion or heart failure are almost always found, along with cardiac enlargement, hepatomegaly, and rales. A systolic murmur usually is present but is variable and not diagnostic. It can result from any of the different intracardiac communications that may be present. Absence of a murmur, which suggests the absence of any intracardiac communication, indicates a particularly unfavorable prognosis.

The chest radiograph often shows three distinctive abnormalities. The contour of the heart has been described as “egg-shaped,” which results from the prominent right ventricle projecting into the left hemithorax and the dilated right atrium bulging into the right side. The base of the cardiac shadow, termed the waist, may be unusually narrow because of the location of the aorta in front of the pulmonary artery, rather than the normal side- by-side relationship. Pulmonary congestion often is marked.

The electrocardiogram consistently shows severe right ventricular hypertrophy. The presence of left ventricular hypertrophy depends on the pulmonary blood flow and the degree of pulmonary valvular stenosis. The echocardiogram usually is diagnostic, outlining the transposed great arteries, the intracardiac communications, and any associated abnormalities. Cardiac catheterization usually is not required.

Treatment

TGA may be classified into four groups: (1) TGA with intact ventricular septum, (2) TGA with VSD, (3) TGA with VSD plus LVOT obstruction, and (4) complex transposition, which includes any of the previous three forms in association with other severe associated cardiac defects.

The most critically ill neonates are those with TGA and intact ventricular septum, for the only communication between the pulmonary and systemic circulations is through the foramen ovale. In these patients, a Rashkind balloon septostomy often is necessary shortly after birth to allow mixing at the atrial level. In patients with TGA and intact ventricular septum, the arterial switch operation should be performed within the first 7 to 14 days of life, before the left ventricle involutes and loses its ability to support the systemic circulation. Patients with TGA and intact ventricular septum who are more than 2 weeks old may require preliminary banding of the pulmonary artery to induce hypertrophy in the left ventricle before it is feasible for the arterial switch operation to be performed. This approach, first proposed by Yacoub, was popularized by Jonas, who reported that only 7 days of preliminary banding was necessary to induce left ventricular hypertrophy before proceeding to a rapid two-stage arterial switch operation.

When TGA with VSD is present, the arterial switch operation often is done within the first 2 weeks of life to avoid progressive pulmonary congestion. If the lungs are not flooded, operation may be safely delayed for 3 to 6 weeks. The left ventricle usually does not have a problem supporting the systemic circulation when a VSD is present even if the operation is done later, because the ventricle has been working against systemic pressure since birth.

Operative Technique

The surgical technique for the arterial switch procedure (Fig. 17-25) was described in detail by Jatene, and modifications have been described by Quaegebeur, Castañeda, and others. The repair often is performed with the patient in deep hypothermia and circulatory arrest, although some groups use “low-flow” continuous perfusion to minimize risk of neurologic dysfunction.

After cardiopulmonary bypass is initiated and the heart is arrested with cardioplegia, the great vessels are totally transected just distal to the sinotubular ridge (Fig. 17-25 A). A small “button” of aorta that includes the ostium of the coronary artery is resected from the posterior sinus of the anterior great vessel and transferred to the corresponding sinus of the posterior great vessel. A similar transfer is made with the “button” of the anterior right coronary artery (Fig. 17-25 B). Microvascular technique is used to suture the coronary buttons into their new locations on the posterior great vessel (Fig. 17-25 C). The anterior aorta is relocated posteriorly and sutured end-to-end to the posterior great vessel just distal to the coronary artery reimplantation site. The pulmonary artery is brought anteriorly and connected to the anterior great vessel, which exits from the right ventricle.

The subgroup of patients with TGA and VSD with concomitant LVOT obstruction may not be suitable for the arterial switch operation or an atrial correction. This group of patients may require initial palliation with a systemic-to-pulmonary shunt, followed by a Rastelli operation at 4 to 5 years of age. The Rastelli procedure involves use of a prosthetic patch to reroute the anterior aorta to connect internally to the posterior left ventricle, across the VSD. The systemic venous ventricle is connected extraanatomically into the pulmonary artery with a valved homograft conduit or a prosthetic valve and graft.

Results

The operative risk for the Rastelli operation in patients with TGA and VSD with LVOT obstruction is relatively low, with mortality at 5 to 10 percent, but higher than the risk

associated with the arterial switch procedure for other forms of transposition. Because of the need for an extraanatomic valved conduit, the risks of late obstruction of the conduit and degeneration of the valve are ongoing, and reoperations for conduit replacement usually are required throughout life.

For patients with *TGA* and intact ventricular septum and for patients with *TGA* and *VSD*, the arterial switch operation achieves excellent long-term results, with a mortality of less than 5 percent. Wernovsky and colleagues reported results in 470 patients who underwent the arterial switch operation between 1983 and 1992, of which 278 were for *TGA* with intact septum and 192 for *TGA* with *VSD*. The overall operative mortality was 7 percent, with 1-year, 5-year, and 8-year survival rates of 92 percent, 91 percent, and 91 percent, respectively. The operative risk was increased when certain unfavorable coronary artery anatomic configurations were present and when augmentation of the aortic arch was required. The risk of subsequent reintervention for pulmonary artery stenosis was approximately 10 percent, with a peak risk at 9 months postoperatively.

Elkins and associates reported similar late results after the arterial switch operation, with excellent survival rates and preservation of ventricular function for up to 8 years. In another study, Colan and associates aggressively tested left ventricular functional parameters after the arterial switch operation, demonstrating that after follow-up periods as long as 10 years, left ventricular size, function, and contractility continued to be normal with no evidence of time-related deterioration. Kramer and Turley reported that the late incidence of cardiac arrhythmias was less in patients undergoing the arterial switch procedure than in patients receiving Senning or Mustard atrial repairs. The overall late results after the arterial switch operation are extremely encouraging in terms of survival, preservation of ventricular function, and low incidence of late arrhythmias. The arterial switch procedure is now the preferred treatment for patients with transposition of the great arteries.

Double-Outlet Right Ventricle

Double-outlet right ventricle is a congenital malformation in which both great arteries arise from the morphologic right ventricle. It occurs in about 5 percent of patients with congenital heart disease. Modern knowledge of the condition emerged from observations by Kirklin, who first recognized the anatomic problem in 1957 in the operating room and performed a surgical correction by creating an intraventricular tunnel, similar to the treatment used today. *Double-outlet right ventricle (DORV)* became established as the appropriate designation after Witham used the term in a report on the disorder.

Pathology and Pathophysiology

Patients with *DORV* were classified by Lev on the basis of the location of the *VSD* in relation to the great vessels: (1) subaortic, (2) subpulmonic, (3) doubly committed, and (4) uncommitted. The aorta rotates anteriorly to varying degrees in all patients with *DORV*, arising from the right ventricle. The great vessels are usually side by side, although in extreme cases the aorta is completely anterior. In *DORV* and subaortic *VSD* with pulmonary stenosis the pathology merges into the spectrum of tetralogy of Fallot. When the aorta rotates to a greater degree anteriorly and the pulmonary artery begins to rotate posteriorly, in some cases overriding the *VSD*, a special condition of *DORV* and subpulmonic *VSD* termed *Taussig-Bing syndrome* occurs. The aorta and the pulmonary artery arise from the right ventricle, allowing some mixing of venous and arterial blood to occur with some degree of bidirectional shunting in all cases. The amount of cyanosis varies with the degree of restriction to pulmonary blood flow.

Clinical Manifestations

Three characteristic types of physiology occur. In patients with *DORV* and a large subaortic *VSD* or doubly committed *VSD*, the clinical presentation is identical to that of a large, isolated *VSD*, with markedly increased pulmonary blood flow. These patients develop congestive heart failure early and have a propensity for pulmonary hypertension and early pulmonary vascular resistance changes. They typically do not have appreciable cyanosis.

The second typical clinical syndrome occurs in patients with *DORV* and subaortic *VSD* with pulmonary stenosis. This condition overlaps with tetralogy of Fallot, and physiologically the patients are identical. The predominant clinical findings are hypoxia and cyanosis from markedly restrictive pulmonary blood flow.

The third clinical variant is the Taussig-Bing syndrome, *DORV* with subpulmonic *VSD*. The aorta usually is anterior and the pulmonary artery posterior, overriding the *VSD*. Anatomically and physiologically these patients resemble patients with transposition of the great arteries and *VSD*. Patients with Taussig-Bing syndrome typically have a combination of increased pulmonary blood flow with pulmonary congestion and moderate hypoxia with cyanosis. *DORV* with subpulmonic *VSD* and pulmonary atresia can occur, which resembles *TGA* with *VSD* and pulmonary atresia.

Echocardiography can almost always determine the diagnosis. Cardiac catheterization is not routinely performed if the patient has straightforward *DORV* and subaortic *VSD* with increased pulmonary blood flow. In patients whose anatomic configuration is less certain, catheterization may be necessary for delineation of the exact anatomic details before repair is performed or to determine the pulmonary vascular resistance.

Treatment

Most forms of *DORV* with subaortic *VSD* or doubly committed *VSD*, with or without pulmonary stenosis, can be corrected satisfactorily with an intracardiac tunnel to channel blood from the left ventricle across the *VSD* to the aorta. The right ventricular outflow tract is then enlarged either with a patch or with a homograft valved conduit as necessary. In patients with Taussig-Bing syndrome (*DORV* with subpulmonic *VSD*) an intracardiac tunnel repair may be difficult. Some surgeons have proposed repair by use of the arterial switch procedure along with *VSD* closure to redirect blood through the neo-aorta.

Repair procedures for *DORV* and simple subaortic *VSD* with or without pulmonic stenosis have an operative mortality of 2 to 5 percent. Repair procedures for more complex forms of *DORV*, including Taussig-Bing syndrome, have an operative mortality of 5 to 10 percent.

Truncus Arteriosus

Truncus arteriosus is a rare malformation accounting for less than 3 percent of congenital cardiac defects. The lesion results from a failure of the fetal truncus to separate into the aorta and pulmonary arteries, resulting in a single arterial trunk. In this condition the entire circulation, including the aortic valve, the aorta, the coronary arteries, and the pulmonary arteries, arise from the common arterial trunk. A single truncal semilunar valve is present, which may be bicuspid (30 percent), tricuspid (50 percent), or quadricuspid (20 percent). There is always an underlying ventricular septal defect.

Pathology and Pathophysiology

Truncus arteriosus is classified according to the origin of the pulmonary arteries (Fig. 17-26): In Type I truncus arteriosus there is a single main pulmonary trunk off the aorta; in Type II the right and left pulmonary arteries originate from the dorsal wall of the truncus arteriosus from separate orifices; in Type III the pulmonary arteries arise as separate ostia; and in Type IV the pulmonary blood supply is provided by systemic aorta- to-pulmonary collateral vessels arising from the descending aorta. Type IV truncus is another way of classifying tetralogy of Fallot and pulmonary atresia with discontinuous pulmonary arteries and major systemic aorta- to-pulmonary collaterals. In most patients the ductus arteriosus is absent; if a ductus is present, usually it is large with a corresponding decrease in size of the aortic isthmus. Interrupted aortic arch Type B is present in 10 percent of patients with truncus arteriosus.

The physiologic abnormality is severe, with 50 percent of patients dying in the first month of life and 90 percent within the first year. Death results from unrestricted pulmonary blood flow with severe pulmonary congestion, congestive heart failure, and progressive pulmonary vascular changes. These patients generally develop early dyspnea and respiratory distress. Severe incompetence of the truncal valve also may occur, which worsens the degree of early heart failure.

Shunting is bidirectional, because the aorta receives blood from the right ventricle and the left ventricle. Arterial oxygen desaturation is always present, with the degree of cyanosis varying with the volume of pulmonary blood flow. In infancy the oxygen saturation usually is above 80 percent, so cyanosis is minimal. Severe pulmonary vascular disease develops rapidly, often before 6 months of age. As this progresses, arterial oxygen saturation decreases and cyanosis becomes progressively more prominent.

Clinical Manifestations

The chest radiograph shows cardiomegaly and pulmonary congestion. Right and left ventricular hypertrophy are evident on the electrocardiogram. The echocardiogram is diagnostic, outlining the single vascular trunk originating from the base of the heart. Cardiac catheterization and angiography define the anatomy precisely, including the origin of the pulmonary arteries and the presence of insufficiency of the truncal valve. The pulmonary vascular resistance also can be determined, which is especially important in older patients.

Treatment

In the procedure developed by McGoon the pulmonary arteries are detached from the truncus, the right ventricle is opened, the ventricular septal defect is closed with a patch, and a homograft valve conduit is used to reconstruct flow into the pulmonary vascular bed. Operation should be performed promptly once the diagnosis is made, preferably within the first month of life.

In 1985 DiDonato described operative results for 167 patients over a 17- year period. There were 48 hospital deaths (29 percent mortality). Of the 119 surviving patients 84 percent were alive after 5 years, and 69 percent after 10 years. Sharma reported experiences with 23 patients, 16 of whom were less than 1 year of age. There were only three operative deaths, two of which occurred in critically ill infants operated on under 1 month of age. Ebert described experiences with 106 infants with 11 operative deaths. Of the 86 long-term survivors, 15 required a later operation for a change of the conduit because of body growth or pseudointimal proliferation in the conduit. There were no mortalities at the time of conduit change.

The operative mortality with this procedure is in the range of 5 to 10 percent. The main causes of perioperative death are ventricular failure, truncal valve insufficiency, and postoperative pulmonary hypertensive crisis. Patients with concomitant interrupted aortic arch have an increased risk. Late complications include the need for reoperation for pulmonary conduit change, which persists throughout life, and the development of progressive insufficiency in the abnormal truncal valve.

OTHER COMPLEX MALFORMATIONS

Hypoplastic Left Heart Syndrome

Lev first described the pathologic findings of aortic hypoplasia in 1952. In 1958 Noonan and Nadas described the clinical syndrome and termed the lesions *hypoplastic left heart syndrome (HLHS)*. *HLHS* encompasses a group of malformations, including aortic hypoplasia or atresia and a poorly developed or absent left ventricle. *HLHS* accounts for 2 to 4 percent of congenital heart defects and for more than 20 percent of congenital cardiac deaths. While the cause is unknown, approximately 25 percent of patients with *HLHS* have an identifiable genetic defect.

Pathology and Pathophysiology

The pathology of *HLHS* includes aortic stenosis, hypoplasia, or atresia, which produces severe left ventricular outflow tract obstruction with almost no forward blood flow. The ascending aorta and proximal aortic arch are diminutive, providing retrograde flow to the coronary arteries. The left ventricle is severely hypoplastic or absent, and the myocardial muscle fibers are in disarray, with severe endocardial fibroelastosis. The mitral valve is hypoplastic or totally atretic in 85 percent of the patients, and approximately 15 percent have a severely malaligned common atrioventricular valve.

Mixing of blood must occur at the atrial level, because no forward blood flow occurs through the hypoplastic left side of the heart. Blood leaves the heart through the large right ventricle and pulmonary artery, which is always enlarged, and proceeds to the lungs and, through the patent ductus arteriosus, to the systemic circulation. Cerebral and cardiac perfusion are retrograde through the atretic aortic arch and ascending aorta.

Shortly after birth oxygenation and perfusion usually are adequate until the ductus arteriosus closes. In some cases the lungs are overperfused, and the patient remains well oxygenated but becomes acidotic because of poor systemic perfusion.

Clinical Manifestations

Most children appear deceptively normal at birth. Typically the child only becomes symptomatic when the ductus arteriosus begins to close, at 24 to 48 h after birth, as they develop cyanosis, tachypnea, and respiratory distress. Severe acidosis develops and the child's color becomes ashen from poor perfusion. Death occurs promptly if the ductus arteriosus is not reopened with prostaglandin infusion. More than 95 percent of patients with *HLHS* die within 2 to 4 weeks, while a small percentage survive because the ductus remains patent.

Physical examination reveals signs of congestive heart failure such as rales and hepatomegaly. Peripheral pulses often are diminished, and perfusion is poor. The chest radiograph shows cardiomegaly with significant pulmonary congestion, and the electrocardiogram shows an absence of left-sided forces, right-axis deviation, and right ventricular hypertrophy. The echocardiogram usually is diagnostic, but it must be done meticulously to outline all of the pathologic anatomy. Color Doppler imaging is used to assess the physiology. Cardiac catheterization is not routinely necessary. In some patients the diagnosis is made prenatally because the lesions are readily identifiable by fetal echocardiography.

Treatment

The initial treatment is to maintain ductal patency with prostaglandin E₁. Acidosis is corrected with sodium bicarbonate. Because the outlook is poor even with surgical therapy, immediate and detailed counseling with the family is done to determine a treatment plan.

The best results have been obtained with a staged palliative treatment plan as described by Norwood. The initial operation is performed in the first week of life with the patient in deep hypothermia and circulatory arrest. The goal of this staged treatment is similar to that for all patients with single-ventricle physiology, that is, to provide passive pulmonary blood flow with a modified Fontan procedure.

The initial palliative stage 1 Norwood procedure involves establishing unobstructed flow from the right ventricle to the systemic circulation by anastomosis of the pulmonary artery to the aortic arch, usually with augmentation of the ascending aorta and arch. Atrial mixing is established by resecting the atrial septum, and controlled pulmonary blood flow is established by a systemic-to-pulmonary modified Blalock-Taussig shunt.

Previously the second stage was a modified Fontan procedure, performed at 12 to 18 months of age after the pulmonary vascular resistance had dropped to normal. A high mortality rate occurred during the interval between operations using this approach. Today the second stage is a bidirectional Glenn shunt, which is performed when the patient is 4 to 6 months of age. Cardiac catheterization is performed before the second stage to evaluate the aortic arch and aortic outflow tract, the pulmonary arteries, and the pulmonary vascular resistance. During the second-stage procedure, the previously placed Blalock-Taussig shunt is taken down, and the pulmonary arteries are augmented as necessary to correct any anatomic obstruction. The final stage is a completion Fontan procedure (total cavopulmonary connection), which is performed approximately 1 year after the second stage.

Results

Survival from the initial stage 1 Norwood procedure exceeds 80 percent at most centers. Subsequent mortalities for stage 2 and stage 3 are 5 to 10 percent each. In 1992 Norwood reported results in 354 patients with *HLHS*, noting an overall survival of 70 percent. Late complications are similar to those associated with the Fontan procedure.

Another viable approach to patients with *HLHS* is cardiac transplantation. This method of treatment is limited by the availability of infant organs. Survival rates for infants treated by transplantation is similar to that for infants treated by the Norwood approach. Aggressive operative treatment of children with HLHS has resulted in late survival rates similar to those obtained in patients treated for pulmonary atresia with intact ventricular septum and for truncus arteriosus.

Total Anomalous Pulmonary Venous Return

Pathology and Pathophysiology

In 1942 Brody published a pathologic study of 100 cases of anomalous pulmonary venous drainage, 35 of which were cases of total anomalous pulmonary venous return (TAPVR). The condition is classified according to the path of the anomalous venous drainage, which is *supracardiac* in 40 to 50 percent of cases, *intracardiac* in 25 percent, *infracardiac* in 25 percent, and *mixed* in about 5 percent. In most cases the anomalous veins enter a common pulmonary venous sinus or channel, which returns the pulmonary venous blood to the right side of the heart. In supracardiac TAPVR, the common venous channel usually is a vertical vein, which enters the innominate vein. Occasionally, the supracardiac vertical vein enters the superior vena cava directly. With intracardiac TAPVR, the anomalous pulmonary veins typically enter the coronary sinus. In patients with infracardiac or infradiaphragmatic drainage the common venous channel traverses the diaphragm and enters the portal system. An atrial septal defect is almost always present and a patent ductus arteriosus is a common associated abnormality.

The pathophysiology commonly found in TAPVR is pulmonary venous obstruction producing severe pulmonary congestion and respiratory distress. Symptoms are often severe, resulting in death in 50 percent of untreated infants within 3 months and in about 80 percent within the first year. Some degree of pulmonary venous obstruction may be present in any of the drainage types described above, but obstruction usually is present when drainage is infracardiac. In infants without significant pulmonary venous obstruction and with an adequate atrial septal defect, the physiology is that of a large left-to-right shunt, similar to a large atrial septal defect. The left ventricle often is small and relatively underdeveloped because of decreased flow, although it develops normally after correction.

Clinical Manifestations

Severe tachypnea is the dominant symptom in a seriously ill infant with TAPVR. Infants with severe obstruction often have severe pulmonary congestion and hypoxemia, requiring intubation and respiratory support. Diminished peripheral perfusion from decreased left-sided cardiac output also may be present. The diagnosis often is unclear initially, and therefore total anomalous drainage should be considered in any severely tachypneic infant. Cardiac murmurs are not diagnostic. Patients with infracardiac drainage may present very early in life with symptoms of pulmonary venous obstruction. Operation is recommended at the time of diagnosis in most patients, because the operative risk increases significantly after critical obstruction develops.

The chest radiograph may show classic abnormalities because of dilation of the common venous channel, innominate vein, and superior vena cava. The well-recognized double contour is termed the “snowman” appearance.

Two-dimensional echocardiography often can establish the diagnosis and outline the abnormal channels, but cardiac catheterization and angiography also are necessary. An MRI study may define the pathway of the abnormal venous drainage. At catheterization, a classic finding is that blood from the right atrium, pulmonary artery, and femoral artery have an identical oxygen content because of mixing in the right atrium.

Treatment

Operation must be performed urgently in critically ill infants and should be scheduled promptly in all other patients. Total repair usually is done with the patient in hypothermia and circulatory arrest. Operative correction includes construction of a large (2.5 to 3.0 cm) side-to-side anastomosis between the common venous trunk and the left atrium, followed by closure of the atrial septal defect and ligation of the left vertical vein. Supracardiac drainage usually is approached transatrially, whereas infracardiac drainage may be repaired by lifting the heart and performing part of the anastomosis from outside of the heart and part from within the atrium. When anomalous veins enter the coronary sinus (intracardiac type), surgical reconstruction is simpler, consisting of creating a large opening between the coronary sinus and the left atrium and performing a patch closure of the atrial septal defect in such a way that coronary sinus flow is directed into the left atrium.

Results

The operative mortality is 5 to 10 percent in infants. After repair the left ventricle grows, and most patients develop normal cardiac function. Late risks include recurrent pulmonary venous obstruction. With current techniques and the use of absorbable suture material for the pulmonary venous anastomosis, the risk of recurrent venous obstruction is approximately 3 percent. The long-term functional results are excellent. In the 15 to 20 percent of patients with a large pulmonary blood flow who are operated on after 1 year of age, operative mortality is less than 1 percent.

Corrected Transposition

The basic characteristics of this unusual malformation were described by Anderson in 1957, with additional contributions by Schiebler in 1961.

Pathology and Pathophysiology

In this malformation the anatomic right ventricle and the anatomic left ventricle are switched or inverted; the lesion also may be called *ventricular inversion*. The defect arises from a malrotation of the embryonic heart tube, which bends to the left (*D*-ventricular loop). The tricuspid and mitral valves “follow” the inverted ventricles, resulting in a right-sided atrioventricular valve that is morphologically a mitral valve and a left-sided atrioventricular valve that is morphologically tricuspid. The aorta arises from a morphologic right ventricle, and the pulmonary artery arises from a morphologic left ventricle. Because the systemic venous blood reaches the pulmonary trunk (by traversing a morphologic mitral valve and left ventricle) and the pulmonary venous blood reaches the aorta (by traversing a morphologic tricuspid valve and right ventricle), the “double discordance” results in a physiologically normal (or “corrected”) circulation.

The significance of the malformation lies primarily in the associated defects and in the propensity of the tricuspid valve to become insufficient. The most frequent associated defect

is a ventricular septal defect, which is present in approximately 80 percent of cases. Another frequent finding is the presence of conduction abnormalities. Normal atrioventricular conduction is present in less than one-half of the patients, and progression to heart block is common. Pulmonic stenosis frequently occurs, which may be severe in some patients. The final associated malformation is left-sided (tricuspid) atrioventricular valvular insufficiency. This gradually develops as a consequence of the tricuspid valve's outflow entering the systemic circulation.

Clinical Manifestations

Conduction defects may cause problems in infancy; 5 to 10 percent of patients are born with a complete heart block, and heart block appears in about 2 percent of the patients each year, with about 30 percent eventually developing complete block.

A large ventricular septal defect usually is present, which produces classic left-to-right shunting and pulmonary congestion. Closure of the VSD is indicated. Severe pulmonary vascular disease develops in most patients with a large VSD unless significant pulmonic stenosis also is present. When pulmonic stenosis is present pulmonary flow is restricted, and the patient is protected from pulmonary overperfusion. In approximately 30 percent of patients, the pulmonic stenosis is severe enough to cause hypoxia and cyanosis, resulting in the need for early total correction or a palliative shunt.

The electrocardiogram usually is abnormal, characteristically indicating conduction disturbances and right ventricular hypertrophy. The echocardiogram usually is diagnostic. Cardiac catheterization may be necessary in older patients.

Treatment

Closure of the VSD is technically difficult because of the uncertainty of the conduction tissue, which tends to course anteriorly and superiorly on the right side of the superior septal tissue. A standard transatrial approach through the right atrium and tricuspid valve is preferred. Heart block is not unusual after operation, with an incidence of 10 to 20 percent. Patients with severe pulmonic stenosis often require placement of an extracardiac valved homograft to the pulmonary artery, because the pathologic anatomy precludes an incision across the stenotic pulmonic valve. When the left-sided tricuspid valve is insufficient, valve repair or replacement is necessary.

Results

The combined series reported by Kirklin and Barratt-Boyes includes a total of almost 100 patients, with an operative mortality of 10 to 15 percent and a 10-year survival rate of approximately 75 percent. In recent years the operative mortality has been less than 10 percent in many centers. Late complications include the ongoing need for conduit change, left ventricular failure, left-sided atrioventricular valvular insufficiency, and progressive conduction disturbances.

OTHER ANOMALIES

Anomalous Origin of the Left Coronary Artery

Anomalous origin of the left coronary artery is a rare malformation, occurring in about 1 of every 300,000 live births and representing about 0.25 percent of patients with congenital heart disease. The clinical features described by Bland in 1933 emphasize a similarity to myocardial infarction in adults. A significant contribution was made in 1959 by Sabiston, who demonstrated that the flow of blood in the anomalous left coronary artery was retrograde into the low-pressure pulmonary artery. Ligation of the anomalous artery subsequently became a form of treatment. Apley, in 1957, recommended ligation of the left coronary followed by bypass with the left subclavian artery. Cooley, in 1966, recommended detaching the coronary artery and connecting it to the aorta with a saphenous vein graft, but late vein graft stenosis has limited the effectiveness of this approach. The creation of an intrapulmonary-artery tunnel for reimplantation of the left coronary artery onto the aorta was proposed by Takeuchi in 1979 and has been quite successful. The most widely used technique is a direct coronary transfer method, proposed by Grace in 1977. Concomitant mitral valve repair frequently is required because of secondary mitral insufficiency.

Clinical Manifestations

Patients become symptomatic early. Most patients develop myocardial infarction and left ventricular failure within 3 months after birth. Only 10 to 20 percent of untreated infants live more than 1 year because of abundant collateral circulation from the right coronary artery. Symptoms usually are mild for the first few weeks after birth because of the elevated pulmonary vascular resistance in the neonatal period. Subsequently, symptoms progress rapidly. Symptoms of tachypnea, sweating, poor feeding, respiratory distress, and heart failure are a result of progressive myocardial ischemia with myocardial infarction, left ventricular failure, and progressive mitral valve insufficiency.

Physical examination may reveal cardiac enlargement without murmur, or with a murmur of mitral insufficiency. The chest radiograph may show extensive enlargement of the left ventricle with pulmonary congestion. The electrocardiogram usually is diagnostic, with inverted T waves or prominent Q waves in the anterolateral leads. Transesophageal echocardiography is useful in confirming the diagnosis and in demonstrating poor contractility, absence of the normal origin of the left coronary artery from the aorta, and, sometimes, mitral insufficiency. Cardiac catheterization and angiography usually are done before repair.

Treatment

Infants with this malformation are at risk for sudden death and progressive myocardial failure, and operative repair is indicated once the diagnosis is made. Because of the excellent experience with coronary artery transfer in the arterial switch operation, the coronary transfer method described by Grace is the treatment of choice. The procedure is performed with the patient on cardiopulmonary bypass and cardioplegic arrest. The left main pulmonary artery is transected just distal to the sinotubular junction, as in the switch operation. Typically the left main coronary arises from the posterior sinus. The ostium is excised from the sinus with a small "button" and reimplanted onto the appropriate point of the lower medial ascending aorta. The posterior pulmonary sinus may be augmented with a small patch of pericardium, and the pulmonary artery is closed. If significant mitral insufficiency is present, mitral valve annuloplasty is performed.

Results

The operative mortality is around 10 percent, depending on the severity of left ventricular dysfunction and the amount of irreversible myocardial damage. Postoperatively, surviving patients often have dramatic improvement in left ventricular function, with more than 95 percent of patients returning to New York Heart Association (NYHA) class I functional status. Late patency of the coronary transfer approaches 100 percent.

Vascular Rings

Vascular rings comprise an uncommon type of congenital defect in which an anomalous arterial formation can result in compression of the esophagus or trachea. Patients are frequently symptomatic, and surgical therapy is effective with little morbidity or mortality. Embryologically, the vascular rings result from some variation in the normal formation of the aorta and pulmonary artery from the embryonic aortic arches. In the normal embryo the first two aortic arches disappear, and the fifth arch never fully develops. The third, fourth, and sixth arches are significant in normal development. The right common carotid artery arises from the third arch, the innominate artery from the right fourth, the transverse aortic arch from the left fourth, and the ductus arteriosus from the sixth.

Pathology

Five types of vascular anomalies have been noted: (1) double aortic arch, (2) right aortic arch with left ligamentum arteriosum, (3) retroesophageal subclavian artery, (4) anomalous origin of innominate artery, and (5) anomalous origin of left common carotid artery. *Pulmonary artery sling*, which is another form of vascular ring that produces severe, symptomatic tracheal compression, is considered separately below.

A double aortic arch, with one limb anterior to the trachea and the other limb posterior to the esophagus, is the most severe of the vascular rings, usually producing symptoms in early infancy. One arch usually is smaller than the other. When the descending aorta is left-sided, the posterior arch is typically the larger one; when the upper descending aorta is right-sided, the anterior arch is likely to be larger. A right aortic arch with a retroesophageal ductus arteriosus or ligamentum arteriosus, or a right arch with a retroesophageal left subclavian artery and a Kommerell (ductal) diverticulum may produce symptomatic tracheal compression. A left-sided aorta with a retroesophageal right subclavian artery is a common anomaly but usually does not cause symptoms. The last two of the five anomalies, anomalous origin of the innominate artery or of the left common carotid artery, are very rare. When these conditions produce symptoms, they are from direct compression, as a true vascular ring often is not present.

Clinical Manifestations

Most symptoms from vascular rings result from compression of the trachea. Difficulty in swallowing from compression of the esophagus is rare. Infants with a double aortic arch often develop difficulty in breathing during the first few months of life and become critically ill. Stridor is the most frequent prominent symptom. Periodic episodes of serious respiratory distress with "crowing" respirations occur. During these attacks the infant lies in hyperextension while gasping for breath. Feeding often precipitates such episodes, perhaps from flexion of the neck or aspiration. Infants quickly become underweight and malnourished.

Most patients with vascular rings requiring surgical treatment are seen in infancy. Those with mild symptoms that develop after 1 year of age may spontaneously recover as they grow older. The most common symptoms are related to intermittent respiratory compression, exacerbated by a respiratory infection or feeding. Recurrent episodes of pneumonia also are relatively common. Difficulty in swallowing, if present, is mild. A mild clinical picture is produced by the retroesophageal subclavian artery, which may cause mild, intermittent dysphagia. Some patients who are mildly symptomatic in infancy spontaneously recover with growth.

The chest radiograph usually is normal unless pneumonia is present. Examination of the esophagus with a barium swallow can establish the diagnosis, demonstrating a typical area of esophageal compression. The diagnosis is established with an MRI scan, or with enhanced MR angiography, which outlines the vascular anatomy and demonstrates any tracheal compression or tracheal stenosis present. Catheterization and arteriography are unnecessary. Bronchoscopy usually is performed to confirm segmental tracheal compression and to rule out diffuse tracheomalacia.

Treatment

If no symptoms are present, no treatment is needed. If symptoms are mild and of uncertain cause, a period of observation may be prudent to ensure that other medical problems are not responsible. If clear symptoms are present, operation should be performed promptly, because death from airway obstruction can easily occur, particularly in patients with complete vascular rings.

Operative Technique

If the child is in respiratory distress, the airway is secured before the child is fully asleep. A rigid pediatric bronchoscope should be in the operating room. Although the optimal incision varies with the type of pathology, an incision through the fourth intercostal space on the appropriate side usually is selected. An important principle is to dissect the aorta and the aortic arch completely, identifying the innominate artery, the left common carotid artery, the subclavian artery, the ligamentum arteriosus, and, in the case of double aortic

arch, the anterior and posterior arches. Opening the pericardium can facilitate identification of these vessels. The vagus and phrenic nerves should be identified and preserved.

With a double aortic arch, the smaller of the two arches should be divided. Usually, with a left descending aorta, the anterior arch is smaller and can be divided between the left common carotid and left subclavian artery. In some cases the divided anterior arch may produce some mild compression; if so, it should be sutured to the posterior surface of the anterior chest wall. If the posterior arch is smaller, it can be divided behind the esophagus. With a right descending thoracic aorta, the posterior arch usually is the smaller of the two, and it is divided.

With a right aortic arch and a retroesophageal ligamentum or ductus arteriosus, division of the ligamentum or ductus is all that is necessary. In patients with a retroesophageal left subclavian artery and a Kommerell diverticulum, the diverticulum should be resected to relieve compression.

Results

The operative risk is primarily related to the age of the patient and the severity of compression of the trachea. The operative mortality should be less than 1 to 2 percent. After the patient recovers from operation, most symptoms should quickly resolve, and the long-term results are excellent.

Pulmonary Artery Sling

Pulmonary artery sling is a rare congenital malformation in which the left pulmonary artery arises from the right pulmonary artery, coursing to the left between the trachea and the esophagus to reach the left pulmonary hilus, thus forming a sling or ring around the trachea. The trachea often is segmentally narrowed at the site of compression. In approximately 50 percent of the patients severe tracheal stenosis is present with complete cartilaginous rings. This occurs from abnormal intrauterine development of the trachea, possibly from vascular compromise. The presence of severe tracheal stenosis with complete tracheal rings greatly worsens the prognosis and increases the difficulty of the repair. The most severe cases have long, funnel-shaped stenosis or diffuse tracheal hypoplasia with complete rings for the entire length. Other cardiac anomalies also are present in nearly one-half of reported patients.

Clinical Manifestations

Most infants with pulmonary artery sling develop severe symptoms in the first few months of life, with feeding difficulty, wheezing, stridor, and severe respiratory distress. The diagnosis may be suspected from abnormalities visible on the chest radiograph, with a density separating the trachea from the esophagus on the lateral view. An esophageal barium swallow usually is diagnostic, showing anterior indentation of the esophagus just above the carina tracheae. MRI is the diagnostic imaging technique of choice. It outlines the vascular anatomy and the severity and length of the tracheal stenosis. If the diagnosis remains uncertain, catheterization and angiography are performed.

Bronchoscopy should be performed routinely to evaluate the severity of the tracheal malformations. This is probably the most important determinant of operative strategy and prognosis. Older patients are occasionally seen with minimal or no symptoms. Such patients often require no specific treatment.

Treatment

The operative procedure is relatively simple in the absence of severe tracheal stenosis. The repair, performed through the left lateral thoracotomy, entails dividing the anomalous pulmonary artery at its origin and reanastomosing it to the main pulmonary artery anteriorly. The ligamentum arteriosum also is divided at this time. This procedure also may be done through a median sternotomy.

When significant tracheal stenosis is present the procedure becomes more complicated. Segmental tracheal stenosis should be resected at the time of sling repair, reanastomosing the trachea end-to-end. Jonas reported that the use of absorbable polydioxanone suture (PDS) allowed growth of the tracheal anastomosis with excellent long-term results. In patients with more diffuse tracheal stenosis from complete rings, the operative repair is difficult. Several methods of repair have been proposed. Idris, Bando, and de Lorimier all reported successful repairs using an anterior pericardial tracheoplasty, which involves making an incision along the length of the trachea, followed by tracheal augmentation. Additional buttressing and stenting with cartilage was suggested by de Lorimier, and Murphy proposed the use of vascularized pericardial grafts. Grillo and Tsang both reported use of a slide tracheoplasty technique for long-segment stenosis with excellent success. Any technique for repair of pulmonary artery sling with long-segment tracheal stenosis requires a sternotomy and cardiopulmonary bypass.

Results

The operative risk and the long-term prognosis are primarily determined by the nature and degree of disease present in the trachea. When tracheal stenosis is absent or segmental, the success rate of repair is approximately 98 percent, with most patients remaining asymptomatic. Late occlusion of the pulmonary artery has been reported in some patients. In patients with long tracheal stenosis with complete rings, the operative results are less predictable. Operative mortality is approximately 10 percent in these patients, and there is a small but significant incidence of late recurrent tracheal problems.

Acknowledgment

Original medical illustrations for figures: 17-2–17-7; 17-9; 17-10; 17-13; 17-17–17-19; 17-21; and 17-23 by Joel Herring, Oceanside, New York.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 18 - Acquired Heart Disease

Aubrey C. Galloway
Richard V. Anderson
Eugene A. Grossi
Frank C. Spencer
Stephen B. Colvin

CLINICAL EVALUATION

The importance of the history and physical examination cannot be overemphasized in the evaluation of a patient with acquired heart disease for potential surgery. It is imperative that the surgeon be well aware of the patient's functional status and the clinical relevance of each symptom, because many operative decisions depend on the accurate assessment of the significance of a particular pathologic finding. With the number of available diagnostic tests rapidly increasing, appropriate sequencing of the diagnostic work-up requires clinical perspective and diagnostic acumen that are obtained through the history and physical examination. The associated risk factors or concomitant diseases that can significantly influence the patient's operative risk for cardiac or noncardiac surgery must be accurately identified. Specific physical findings, such as prior saphenous vein stripping, previous thoracic surgery, or peripheral vascular disease, must be accurately identified preoperatively.

History

The classic symptoms of heart disease are fatigue, angina, dyspnea, edema, cough or hemoptysis, palpitations, and syncope. When a patient describes or complains of any of these symptoms, the clinical scenario leading to the symptom must be explored in detail, including symptom intensity and duration, provocation, associated symptoms, and conditions that lead to relief. The initial goal is to determine whether a symptom is cardiac or noncardiac in origin as well as to determine the clinical significance of the complaint. An important feature of cardiac disease is that myocardial function or coronary blood supply that may be adequate at rest may be completely inadequate with exercise or exertion. Chest pain or dyspnea that occurs primarily during exertion frequently is cardiac in origin, while symptoms that occur at rest but not during exercise frequently are not.

In addition to evaluating the patient's primary symptoms, the history should include family history, past medical history (prior surgery, prior myocardial infarction, or concomitant hypertension, diabetes, and other associated diseases), personal habits (smoking, alcohol or drug use), functional capacity, and a detailed review of systems. After the patient's symptoms have been carefully assessed, appropriate diagnostic studies may be ordered and interpreted. The classic symptoms are outlined in detail below.

Fatigue

Easy fatigability is a frequent but nonspecific symptom of cardiac disease that can arise from many causes. In some patients easy fatigability probably reflects a generalized decrease in cardiac output or low-grade heart failure. Subjective easy fatigability is a vague and nonspecific symptom.

Angina

Angina pectoris is the hallmark of myocardial ischemia secondary to coronary artery disease, although a variety of other conditions can produce chest pain. Classic angina is precordial pain described as squeezing, heavy, or burning in nature, lasting from 2 to 10 min. The pain usually is substernal, often radiating into the left shoulder and arm but occasionally occurring in or radiating into the midepigastrium, jaw, right arm, or interscapular region of the back. True angina typically is provoked by exercise, emotion, sexual activity, or eating and is relieved by rest or nitroglycerine. Angina is present in its classic form in 75 percent of the patients with coronary disease, while 25 percent of patients with coronary disease have atypical symptoms, with a small but significant number of patients developing "silent" ischemia.

Angina also is a typical symptom of aortic stenosis, occurring secondary to the combination of left ventricular hypertrophy, increased intracardiac pressure, increased ventricular wall tension (leading to higher oxygen requirements), and decreased cardiac output. This combination results in a myocardial oxygen supply–demand mismatch with resultant ischemia and angina.

Noncardiac causes of chest pain that may be confused with angina include esophageal reflux or spasm, costochondritis (Tietze's syndrome), musculoskeletal pain, ulcer disease, biliary disease, pleuritis, pulmonary embolus, pulmonary hypertension, pericarditis, and aortic dissection.

Dyspnea

The physiologic change in most patients with heart failure is a rise in intracardiac pressure followed by cardiac enlargement, usually a combination of dilatation and hypertrophy. This is a manifestation of Starling's law of the heart: increased work can be achieved in response to an increase in diastolic fiber length. Symptoms develop subsequently as different compensatory mechanisms fail. This concept is important because abundant data indicate that operation should be considered for many diseases on the basis of physiologic abnormalities, such as a progressive drop in left ventricular ejection fraction or a reduction of cross-sectional area of an aortic or mitral valve below 1.0 cm², rather than from the presence of symptoms. Delaying operation until symptoms are severe often results in irreversible ventricular injury, which in turn can be a major cause of death in the first few years after a technically successful operation.

The normal left ventricular end-diastolic pressure is less than 12 mmHg. Pressures in the range of 12 to 20 mmHg represent moderate disease, while pressures of 20 to 25 mmHg or higher represent severe disease. Because the oncotic pressure of plasma is approximately 25 mmHg, as left atrial pressure rises, pulmonary congestion develops when the left atrial pressure approaches the oncotic pressure of plasma. The tolerance for pulmonary congestion depends on several factors, including the capacity of the pulmonary lymphatics to resorb fluid. Dyspnea is one of the cardinal symptoms of left heart failure. It can be graded with the degree of exertion required to initiate dyspnea, as opposed to dyspnea at rest that represents a severe form of heart disease. With mitral stenosis, dyspnea appears as an early sign because of restriction of flow from the left atrium into the left ventricle. With other forms of heart disease, however, dyspnea is a late sign because it develops only after the left ventricle has failed and the end-diastolic pressure rises significantly. Dyspnea with mitral insufficiency, aortic valvular disease, or coronary disease represents relatively advanced pathophysiology, in contrast to mitral stenosis, in which dyspnea occurs relatively early in the disease process.

A number of other respiratory symptoms represent different degrees of pulmonary congestion. These include orthopnea, paroxysmal nocturnal dyspnea, cough, hemoptysis and pulmonary edema. Dyspnea occasionally represents an "angina equivalent," occurring secondary to ischemia-related ventricular dysfunction.

Edema

Left-sided heart failure can result in fluid retention and pulmonary congestion, subsequently leading to pulmonary hypertension and progressive right-sided heart failure. A history of exertional dyspnea with associated edema frequently is from heart failure. Primary right heart failure occasionally results from right ventricular injury and dysfunction or from primary tricuspid valve disease. Right atrial pressure, normally less than 5 to 8 mmHg, may be elevated up to the range of 15 to 30 mmHg and sometimes higher. Retention of more than 7 to 10 pounds of fluid results in visible edema of the lower extremities, which is almost always symmetric bilaterally. Jugulovenous distention and hepatomegaly develop with severe right heart failure. With chronic severe failure, generalized fluid retention may be acute, with marked deformities from accumulation of 20 or more pounds of edema fluid, with ascites and massive hepatomegaly.

Palpitations

Palpitations are secondary to rapid, forceful, ectopic or irregular heartbeats. Palpitations often are innocuous, but they should not be ignored, because they might represent significant or potentially life-threatening arrhythmias. The underlying cardiac arrhythmia ranges from premature atrial or ventricular contractions to atrial fibrillation, atrial flutter, paroxysmal atrial or junctional tachycardia, or sustained ventricular tachycardia. Atrial fibrillation is one of the most common causes of palpitations, occurring frequently in patients with mitral stenosis, and results from left atrial hypertrophy evolving from the sustained elevation in left atrial pressure. With other forms of heart disease, arrhythmias are less common, occurring sporadically. They are more frequent in older patients, probably from intrinsic disease in the atrioventricular conducting mechanism, resulting in "sick sinus syndrome" or intermittent heart block, and in patients with ischemic cardiomyopathy, probably as a manifestation of diffuse myocardial scarring. Severe, life-threatening forms of ventricular tachycardia or ventricular fibrillation may occur in any ischemic patient from ongoing ischemia or scarring. Palpitations caused by a slow heart rate often are from complete or intermittent atrioventricular nodal block.

Syncope

Syncope, or sudden loss of consciousness, usually is a result of sudden decreased perfusion of the brain. The differential diagnosis includes: (1) transient heart block with asystole or ventricular fibrillation (Adams-Stokes syncope or syndrome), (2) malignant tachyarrhythmias, (3) ventricular fibrillation, (4) aortic stenosis, (5) hypertrophic cardiomyopathy, (6) atrial myxoma with transient mitral valve obstruction, (7) primary pulmonary hypertension, (8) seizure disorder, (9) carotid artery or cerebrovascular disease, and (10) vasovagal reaction (fainting). Any episode of syncope must be worked up thoroughly, because these events can result in sudden death.

Functional Classification

An important part of the history is the assessment of the patient's overall functional status or the degree of cardiac disability. This generally is a good approximation of the severity of the patient's underlying cardiac disease. The *New York Heart Association (NYHA)* has developed a classification of patients with heart disease based on symptoms that has been useful in evaluating clinical course and operative risk. A more detailed grading system for patients with angina pectoris was developed by the Canadian Cardiovascular Society (CCS). These classification systems are adequate for the majority of patients. When a more precise functional analysis is necessary, the specific activity scale proposed by Goldman and based on the estimated metabolic cost of various activities is used. The NYHA and CCS classifications are outlined in Table 18-1.

Physical Examination

A few basic physical abnormalities are discussed briefly here, because abnormal physical findings are best discussed in conjunction with the specific disease causing them. In some cardiac diseases, physical abnormalities are virtually diagnostic of the disease and the severity of the problem, while in others, such as coronary disease, the absence of any physical abnormality can be seriously misleading. Conducting a thorough physical examination before performing any surgical procedure is critical if the surgeon is to deliver optimal care. Previously overlooked physical findings can lead the surgeon to suspect a concomitant disease process, in which case additional work-up or altered management can lessen the risk of the proposed operation.

The important parts of the preoperative general and cardiovascular examinations are as follows:

General

The patient must be assessed for overall body size and phenotype, level of body fat, musculoskeletal deformities, and prior surgical scars, because abnormal anatomy or prior

surgery may influence the operative approach.

Generalized muscular wasting can occur in patients with chronic, severe, congestive heart failure from the combination of low cardiac output and an inability to eat, resulting in malnutrition from lack of calories and protein. These severely ill patients are especially susceptible to infection after operation.

Vascular

The peripheral vascular examination is important for all surgical patients, but particularly for patients undergoing cardiac surgery. The vascular examination involves assessment of pulses, an evaluation of peripheral perfusion, and assessment of skin quality and hair loss. The venous system is evaluated, and venous insufficiency or varicose veins are noted. If varicosities are present in the greater saphenous system, the lesser saphenous veins and cephalic veins should be evaluated and preserved.

If the surgeon is considering the use of a radial artery for a cardiac procedure, Allen tests should be performed bilaterally to assess the contribution of the ulnar artery to the palmar arch. In questionable cases, digital plethysmography of the fingers with and without radial artery occlusion should be performed.

Auscultation for carotid and abdominal bruits should be performed in all patients, and the abdomen should be palpated for aneurysmal masses. Significant abnormalities are evaluated by vascular Doppler studies, ultrasonography, or scanning modalities.

Cardiac

A cardiac murmur heard during auscultation frequently establishes the diagnosis of valvular heart disease. An apical diastolic rumble is diagnostic of mitral stenosis, while a parasternal diastolic murmur suggests aortic insufficiency. Systolic murmurs, including the basal systolic murmur of aortic stenosis and the apical systolic murmur of mitral insufficiency, are classic findings. Underlying valvular conditions may be completely unsuspected and the patient referred for coronary bypass or an unrelated general surgical procedure. Whenever an unsuspected murmur is found by the surgeon before operation, additional diagnostic work-up is indicated, usually with echocardiography.

The initial adaptation to underlying valvular heart disease usually is cardiac enlargement. This often is detected in the physical examination. The finding of a forceful apical impulse in the anterior axillary line indicates long-standing disease. Less obvious signs of cardiac enlargement may be seen on the chest x-ray, the electrocardiogram (ECG), or, most precisely, the echocardiogram, which can define the exact size of the cardiac chambers and the thickness of the heart muscle.

DIAGNOSTIC STUDIES

Electrocardiography and Radiology

The electrocardiogram and the chest x-ray are the two classic diagnostic studies. The electrocardiogram is used to detect rhythm disturbances, heart block, atrial or ventricular hypertrophy, ventricular strain, myocardial ischemia and myocardial infarction. The chest x-ray is excellent for determining cardiac enlargement and pulmonary congestion. A cardiothoracic ratio above 0.5 represents severe cardiomegaly. Analysis of the pulmonary circulation may show several abnormalities. Pulmonary venous congestion develops when left atrial pressure is chronically elevated above the upper normal limit of 12 mmHg, seen typically with severe mitral stenosis. The signs of pulmonary congestion include engorged pulmonary veins and congestion of pulmonary alveoli. Fluid accumulating in the interlobar planes forms transverse linear opacities perpendicular to the surface of the pleura (Kerley lines). Their presence usually indicates a left atrial pressure exceeding 20 mmHg.

Marked enlargement of the pulmonary arteries may occur from an increase in pulmonary blood flow or an increase in pulmonary vascular resistance with pulmonary hypertension. Normally the central pulmonary arteries are three to four times larger than the peripheral arteries. With an increase in pulmonary blood flow, as with an atrial septal defect, central and peripheral arteries are symmetrically enlarged. With pulmonary hypertension, the central pulmonary arteries may become strikingly enlarged while the peripheral arteries do not.

Echocardiography

Echocardiography, which has become the most widely used cardiac diagnostic study, incorporates the use of ultrasound and reflected acoustic waves for cardiac imaging. Two-dimensional (2-D) color Doppler echocardiography is now widely available. These studies give accurate dynamic imaging of the cardiac chambers and assessment of valvular configuration or mobility. Doppler flow velocity and frequency are assigned colors, allowing a visual evaluation of direction and velocity of intracardiac blood flow. Intracardiac pressures, valvular insufficiency, and transvalvular gradients can be estimated from Doppler measurements. Color Doppler information can be superimposed onto the 2-D image to provide a graphic illustration of the directional intracardiac flow pattern and an assessment of valvular insufficiency.

Transthoracic 2-D color Doppler echocardiography has become an excellent noninvasive screening test for evaluating myocardial thickness or hypertrophy, cardiac chamber size, cardiac wall motion, intracardiac and pulmonary artery pressures, internal cardiac anatomy, and the degree of valvular stenosis or insufficiency. Corrective operation for valvular disease frequently is performed on younger patients on the basis of these studies.

Transesophageal echocardiography, which is done by placement of the 2-D transducer in a flexible endoscope, improves the image quality by minimizing scatter from the chest wall and is particularly valuable in evaluation of the left atrium, the mitral valve, and the aortic arch. Transesophageal echocardiographic studies are used when more precise imaging is required or when the diagnosis is uncertain after the transthoracic study.

Dobutamine stress echocardiography has evolved as an important noninvasive provocative study. This study is used to assess cardiac wall motion in response to inotropic stimulation, as wall motion abnormalities reflect underlying ischemia. Several reports have documented the accuracy of dobutamine stress echocardiography in identifying patients with significant coronary artery disease. The predictive value of a positive test for myocardial infarction or death after noncardiac surgery is approximately 10 percent, while 20 to 40 percent will have some cardiac event. A negative test is 93 to 100 percent predictive that no cardiac event will occur.

Radionuclide Studies

The most widely used myocardial perfusion screening study is the thallium scan, which uses the nuclide thallium-201. Initial uptake of thallium-201 into myocardial cells is dependent on myocardial perfusion, while delayed uptake depends on myocardial viability. Reversible defects occur in underperfused, ischemic, but viable zones, while fixed defects occur in areas of infarction. Fixed defects on the thallium scan suggest nonviable myocardium and may be of prognostic value.

The *exercise thallium test* is used widely to identify inducible areas of ischemia and is 95 percent sensitive in detecting multivessel coronary disease. This is the best overall test to detect myocardial ischemia, but it requires that the patient be able to exercise on the treadmill. The study also gives excellent, specific information about the patient's cardiac functional status.

The *dipyridamole thallium* study is a provocative study using intravenous dipyridamole, which induces vasodilation and consequently unmasks myocardial ischemia in response to stress. For patients who cannot exercise, this is probably the most widely used provocative study for risk stratification. In patients undergoing noncardiac surgery the predictive value of a positive dipyridamole thallium study is 5 to 20 percent for myocardial infarction or death, while a negative study is 99 to 100 percent predictive that a cardiac event will not occur. This is a very effective screening study for moderate- to high-risk patients who are undergoing a general surgery procedure.

Global myocardial function is evaluated by the *gated blood pool scan* (equilibrium radionuclide angiocardiology) using technetium-99m. This study can detect areas of hypokinesis and measure left ventricular ejection fraction, end-systolic volume, and end-diastolic volume. An exercise-gated blood pool scan is an excellent way to assess a patient's global cardiac response to stress. The ejection fraction increases with exercise, but with significant coronary artery disease or valvular disease the ejection fraction may stay unchanged or even drop. The resting gated blood pool scan is an excellent way to determine the degree of prior cardiac injury and to assess baseline cardiac function, and the exercise gated blood pool scan assesses the functional response to stress.

Positron-Emission Tomography (PET) Scan

The PET scan is a special radionuclide imaging technique used to assess myocardial viability in underperfused areas of the heart. The technique may be more sensitive than the thallium scan for this purpose. The PET scan is based on the myocardial metabolism of glucose or other compounds tagged with positron-emitting isotopes. The PET scan is most useful in determining whether an area of apparently infarcted myocardium may be "hibernating" and capable of responding to revascularization. These data can be used to determine whether patients with congestive heart failure might improve with operative revascularization.

Cardiac Catheterization

The cardiac catheterization study remains the gold standard for cardiac diagnosis. Complete cardiac catheterization includes the measurement of intracardiac pressures, measurement of cardiac output, localization and quantification of intracardiac shunts, determination of internal cardiac anatomy and ventricular wall motion by cineradiography, and determination of coronary anatomy by coronary angiography (Fig. 18-1).

The cardiac output usually is calculated using the Fick oxygen method, in which cardiac index is calculated as follows:

Cardiac index (L/min/m²) = oxygen consumption (mL/min/m²) / arteriovenous oxygen content difference (mL/min)

For determining the arteriovenous oxygen difference, the oxygen content is calculated separately in the arterial and venous circulations by the formula:

Oxygen content (mL oxygen/L blood) = hemoglobin (g/100 mL) × percent hemoglobin saturation × 1.36 (mL oxygen/g hemoglobin) × 10.

Calculation of systemic vascular resistance (SVR) is by the formula:

SVR = (mean systemic arterial pressure - mean right atrial pressure) × 80 / systemic blood flow (cardiac output)

The normal SVR is 1,200 dynes·sec·cm·sec·cm⁻⁵ The pulmonary vascular resistance (PVR) is calculated by the formula:

PVR = (mean pulmonary artery pressure - mean left atrial pressure) × 80 / pulmonary blood flow (equal to the cardiac output when no shunt is present)

The normal PVR is 70 to 80 dynes·sec·cm⁻⁵.

The area of a cardiac valve can be determined from measured cardiac output and intracardiac pressures using Gorlin's formula that relates the valve area to the flow across the valve divided by the square root of the transvalvular pressure gradient ($A = f/k \times \text{square root of pressure gradient}$). The Gorlin formula indicates that a small valve area might be manifested as a small transvalvular pressure gradient when the cardiac output is low, demonstrating the danger of basing decisions on transvalvular gradient alone. Estimations of the significance of an obstructing valvular lesion should be based on the calculated valve area obtained from precise measurements in a cardiac catheterization laboratory. In an adult the normal mitral valve area is 4 to 6 cm² and the normal aortic valve area is 2.5 to 3.5 cm².

Coronary angiography is essential for the diagnosis of coronary artery disease and is routinely performed before a coronary artery bypass operation (see Fig. 18-1). The posterior descending artery and the atrioventricular (AV) nodal artery arise from the right coronary artery in 80 to 85 percent of patients, and the right coronary artery is termed dominant in these cases. The left coronary system supplies the major portion of left ventricular myocardium in all cases. In 15 to 20 percent of cases the circumflex branch of the left coronary system also supplies the posterior descending branch and the AV nodal artery and is termed dominant, while 5 percent are codominant.

Preoperative Assessment of General Surgical Patients

Preoperative cardiac risk stratification for patients undergoing noncardiac surgery has been identified as an important part of the preoperative evaluation of the general surgery patient. The joint American College of Cardiology/American Heart Association task force has produced guidelines and recommendations that are summarized here. The preoperative cardiovascular evaluation involves an assessment of clinical markers, the patient's underlying functional capacity, and various surgery-specific risk factors.

The *clinical markers* that predict an increased risk of a cardiac event during noncardiac surgery are divided into three grades. *Major* predictors include unstable angina (Canadian class III or IV), recent myocardial infarction with ongoing ischemic risk, decompensated heart failure, significant arrhythmias, and severe valvular disease. *Intermediate* predictors are mild angina (Canadian class I or II), old myocardial infarction, compensated heart failure, and diabetes. *Mild* predictors are advanced age, hypertension, prior stroke, abnormal ECG, and poor functional status.

Various *surgical risk factors or procedures* expose the patient to greater or lesser risk of a cardiovascular event. *High-risk* procedures include emergent major procedures in the elderly, major vascular procedures (i.e., thoracic or abdominal aortic or peripheral vascular) and long procedures with large anticipated fluid shifts (e.g., pancreatectomy, hepatic resection, abdominal-perineal resection, etc.). *Intermediate-risk* procedures include any intraabdominal or intrathoracic operation, carotid endarterectomy, orthopaedic, prostate, and head and neck procedures. *Low-risk* includes endoscopic, breast, cataract, and superficial operations.

On the basis of the clinical markers, the patient's functional class, and the proposed surgical procedure, the patient is assigned a high, intermediate, or low cardiac risk and managed appropriately. Sometimes further risk stratification is required, such as in patients with intermediate-level cardiac risk factors who are undergoing a high-risk surgical procedure. These patients should probably undergo exercise stress testing or provocative testing (dipyridamole thallium or dobutamine echocardiography) before operation. In patients who are considered to be at high cardiac risk because of clinical markers or noninvasive testing, coronary angiography may be recommended before surgery. Coronary artery disease is then managed according to the classic indications. In patients who are thought to be at low or moderate cardiac risk, medical management alone is sufficient.

Because of the common atherosclerotic etiology and the close association between clinically relevant coronary artery disease and peripheral vascular disease, all patients undergoing major vascular surgery should be screened closely, either by history or by provocative testing. Those with a positive clinical history, a decreased ejection fraction from previous myocardial infarction, or with exercise or provocative tests suggestive of ischemia should undergo coronary angiography before the vascular surgery, as outlined by Pasternack and others. The underlying coronary disease should be appropriately treated, medically or with revascularization, using classic indications. This aggressive screening approach followed by appropriate intervention in patients with significant coronary artery disease has greatly lowered the operative risk of patients undergoing major vascular surgery.

MEDICAL THERAPY

Coronary Artery Disease

Atherosclerosis is a multifactorial disease resulting in accumulation of lipids, smooth muscle cells, and connective tissue in the vessel wall. This process results in the formation of obstructive lesions in the aorta, the peripheral vessels or the coronary arteries. Atherosclerosis is the leading cause of death in the Western world and acute myocardial infarction accounts for 25 percent of the deaths in the United States each year. Risk factors for coronary artery disease include smoking, obesity, hypertension, diabetes, hypercholesterolemia, hyperlipidemia, sedentary life-style, Type A personality, and male gender. The most important factor in the long-term treatment of coronary disease is the modification of risk factors, including the immediate cessation of smoking, control of hypertension, weight loss and reduction of total serum cholesterol to less than 175 to 200 mg/dL and low-density lipoprotein (LDL) cholesterol to less than 100 mg/dL. If dietary control of cholesterol cannot be achieved in patients with coronary disease, evidence suggests that the use of medications (such as HMG-CoA reductase inhibitors) to lower cholesterol can slow disease progression significantly and lower the risk of subsequent cardiac events.

Angina is the principal manifestation of coronary artery disease. Mild stable angina sometimes is treated successfully with sublingual nitroglycerin for pain relief. More significant angina usually is treated with a combination of aspirin and beta-blockers. Aspirin presumably works through its antiplatelet affect, while beta-blockers limit ischemia through a reduction in myocardial oxygen consumption and tension-time index. Both agents limit the number of cardiac events and improve survival after myocardial infarction. If angina symptoms are unresponsive to beta blockers, second-line therapy is begun with a calcium antagonist or long-acting nitrate. For severe angina "triple therapy" with beta blockers, a calcium antagonist, and a nitrate can be used. A progression of angina despite good medical therapy usually is an indication for cardiac catheterization, as is a change in the frequency or severity of angina. Some cardiologists use a stress test evaluation in patients with angina for risk stratification, and cardiac catheterization is recommended if significant areas of the heart become ischemic at a low work load.

Unstable angina represents a medical-surgical emergency; this condition can progress to acute myocardial infarction and death. Patients with unstable rest angina are admitted to the hospital and treated with aspirin, beta blockers, and intravenous heparin, with close monitoring for cardiac arrhythmias or recurrent pain. Myocardial oxygen demands should be minimized by obtaining a heart rate of 50 to 60 beats/min and a systolic blood pressure of 100 to 110 mmHg. If angina is not controlled, intravenous nitroglycerin and calcium antagonists may be added. When hypotension or congestive heart failure complicate unstable angina, further monitoring with a Swan-Ganz catheter and an arterial pressure line can be helpful in maximizing therapy. Urgent cardiac catheterization is indicated for unstable angina to assess the underlying coronary pathology and to plan revascularization when appropriate.

Acute myocardial infarction, resulting from acute thrombosis of a coronary artery, occurs in 900,000 people in the United States each year, with 225,000 deaths. The American College of Cardiology/American Heart Association task force management guidelines were reported by Ryan and associates and are summarized here. Patients with acute myocardial infarction should be given oxygen, sublingual nitroglycerin, pain relievers, and aspirin on arrival in the emergency department. Early treatment with thrombolytic therapy (streptokinase, tissue plasminogen activator, anistreplase) or transluminal angioplasty within the first 4 to 6 h of an acute infarction decreases mortality rate by 20 to 25 percent, and more if initiated within 1 to 2 h. Immediate revascularization is indicated whenever feasible. Long-term survival after acute infarction is improved by treatment with beta blockers and aspirin, which are begun within 2 to 4 h after admission to the hospital. Contraindications to thrombolytic therapy are cerebrovascular disease, advanced age, or recent surgery; contraindications to beta blockade are heart failure, heart block, and bronchospasm.

After admission to the hospital the patient should be monitored for 24 h. Arrhythmias may be from electrical instability, pump failure, or conduction disturbance. While routine arrhythmia prophylaxis is not indicated, ventricular arrhythmias should be promptly treated with lidocaine, and bradycardia or complete heart block should be treated with atropine and a transvenous pacemaker. Most patients receive intravenous nitroglycerine for 24 h along with aspirin and beta blockers. If the patient does not have hypotension or other contraindications, an angiotensin-converting enzyme (ACE) inhibitor should be started within hours of admission to lower myocardial oxygen requirements by unloading the left ventricle.

Postinfarction angina or recurrent ischemia should be treated with intravenous heparin, nitroglycerin, beta blockade, and urgent cardiac catheterization. Emergent revascularization by catheter-based therapy or coronary bypass surgery is indicated for postinfarction angina or for recurrent ischemia after thrombolytic therapy. This carries a low risk, around 1 to 2 percent. Postinfarction patients with heart failure or with poor ventricular function should undergo catheterization before discharge. While there is variability in the routine use of catheterization in patients with uncomplicated infarction and good left ventricular function, available data suggest that routine catheterization does not reduce the incidence of reinfarction or death. Patients with uncomplicated infarction and good ventricular function should have a modified stress test before discharge or a complete stress test in 2 to 4 weeks. If the patient develops spontaneous ischemia or if the stress test is positive for reversible ischemia, cardiac catheterization is indicated; otherwise, medical therapy is continued.

Cardiogenic shock develops in 10 to 15 percent of patients with acute myocardial infarction, but is less frequent after successful thrombolytic therapy. Patients with cardiogenic shock are monitored closely with a Swan-Ganz catheter and an arterial pressure line and evaluated for mechanical abnormalities such as papillary muscle rupture, cardiac rupture, or ventricular septal defect. Inotropic support frequently is necessary and mechanical support with an intraaortic balloon pump is used if hypotension persists or if the cardiac index remains less than 1.5 L/min/m². Immediate cardiac catheterization is indicated and revascularization should be performed if feasible in areas of myocardium that are not irreversibly damaged. Patients with cardiogenic shock who undergo early revascularization by angioplasty or surgery have a survival rate of 60 to 70 percent, compared to a mortality rate of over 75 percent with medical therapy alone.

Congestive Heart Failure

Valve replacement is indicated when congestive heart failure occurs in patients with valvular heart disease. Medical therapy often is necessary while preparations are made for operation. In such cases digitalis and diuretics are used to increase myocardial contractility and to control volume overload. Afterload reduction, which increases forward blood flow and relieves pulmonary congestion, has been found to be even more important, often allowing the heart to shrink in size and to remodel. ACE inhibitors, such as captopril or enalapril, or intravenous vasodilators, such as sodium nitroprusside, are the agents of choice for afterload reduction. Acutely ill patients in congestive heart failure awaiting valve replacement should be monitored in an intensive care unit while heart failure is treated with a combination of diuretics, afterload reduction, and gentle inotropic support.

When chronic congestive heart failure occurs from presumed cardiomyopathy, a search for reversible causes, such as ischemia, "hibernating myocardium," valvular disease, or ventricular aneurysm is mandatory. When reversible causes are not found, the cornerstone of medical therapy for chronic congestive heart failure has become afterload reduction with ACE inhibitors. Afterload reduction minimizes cardiac work, increases forward flow, diminishes pulmonary congestion and improves long-term survival, resulting in beneficial myocardial remodeling and a decrease in cardiac size. Diuretics are used to control fluid and salt retention, and digitalis is used to increase the force of cardiac contraction.

EXTRACORPOREAL CIRCULATION—MYOCARDIAL PROTECTION

Historical Background

The pioneering imagination and efforts of Gibbon were largely responsible for the development of extracorporeal circulation (cardiopulmonary bypass with pump-oxygenators). In 1932 Gibbon initiated laboratory investigations that continued for over 20 years until the first successful open heart operation in human beings was performed by him in 1953. A bubble oxygenator using a blood-gas interface was later developed, and for the past decade the membrane oxygenator has been used almost routinely.

Pumps

The initial heart-lung machine used a simple roller pump, originally developed by DeBakey. A variety of other pumps have been used since, such as the centrifugal pump (Fig. 18-2) that was developed to minimize trauma to blood elements. In recent years the focus of perfusion technology has been on the development of devices that improve end-organ perfusion while producing less trauma to the patient's blood. A significant advance was the introduction of biocompatible circuits to perfusion technology. This concept involves coating the plastic circuits with heparin or other biocompatible materials that minimize the need for anticoagulation and lessen the activation of complement and other inflammatory cytokines during extracorporeal circulation.

Oxygenators

Since 1985 the disposable membrane oxygenator has been the most widely used type. The hollow-fiber membrane oxygenators improved the efficiency of gas exchange while minimizing trauma to the blood elements. Many newer oxygenators are coated with biocompatible materials such as heparin to further diminish activation of inflammatory mediators. Pump oxygenators usually require a priming volume in the range of 1,000 to 1,800 mL. Usually the use of blood is avoided because of the risk of transmission of hepatitis or acquired immunodeficiency syndrome (AIDS), so the pump-oxygenator and extracorporeal circuit usually are filled with a crystalloid solution. A hemodiluted hematocrit level of 20 to 22 percent probably is optimal during hypothermic cardiopulmonary bypass because of viscosity and shear stress, with improved microvascular perfusion. If the hematocrit level decreases below 18 to 20 percent, blood is added to the pump oxygenator.

Technique of Perfusion

Sufficient heparin is given to elevate *activated clotting time (ACT)* well above 600 s, starting with a heparin dose of 3 to 4 mg/kg. Heparin requirements may be less, however, when heparin-bonded systems are used. Venous blood is aspirated by gravity drainage through large cannulae through the right atrium. Oxygenated blood is returned to the arterial circulation, usually through a cannula in the ascending aorta. Initial perfusion is done at a flow rate of about 2.4 to 3.0 L/min/m², which is the normal cardiac index. Because oxygen consumption decreases with hypothermia, flow rates may be diminished as the patient is cooled. Safe bypass flow rates for 30°C are 1.8 to 2.3 L/min/m², for 25°C 1.5 to 1.8 L/min/m², and for 20°C 1.2 to 1.5 L/min/m². Oxygen flow through the oxygenator is adjusted to produce an arterial oxygen tension above 100 mmHg. Systemic temperature is controlled with a heat exchanger in the circuit; the temperature usually is lowered to 20 to 32°C, although lower temperatures occasionally are necessary for some complicated procedures. Spilled intrapericardial or intracardiac blood is aspirated with a suction apparatus, filtered, and returned to the oxygenator. A cell-saving device is used routinely to aspirate spilled blood before and after bypass. Aspirated blood is washed and reinfused in order to avoid blood transfusion.

During perfusion a number of parameters are monitored. Arterial pressure and central venous pressure are monitored through intravascular catheters. A Swan-Ganz catheter may be inserted to monitor pulmonary artery pressures and cardiac output after bypass.

The blood pressure varies widely among patients during perfusion. It usually decreases sharply with the onset of perfusion, apparently from vasodilatation, and then subsequently rises to above 60 mmHg. The importance of the actual level of mean arterial pressure when the flow rate is adequate is uncertain as long as the patient does not have significant carotid or cerebral vascular disease. Because cerebral autoregulation of blood flow becomes ineffective below a mean pressure of 50 to 60 mmHg, perfusion pressure usually is maintained above 50 mmHg, though with moderate hypothermia (25 to 30°C) an arterial pressure of 45 to 50 mmHg has no harmful physiologic effects. In elderly patients with less autoregulation and in patients with carotid disease the perfusion pressure is maintained at higher levels, usually more than 65 mmHg. After 15 to 30 min of perfusion, perfusion pressure may gradually rise from progressive vasoconstriction.

Oxygen and carbon dioxide tensions are periodically measured in the venous blood returned to the oxygenator and the oxygenated blood returned to the patient. Preferably, the arterial oxygen tension should be above 100 mmHg and the carbon dioxide tension 30 to 35 mmHg. Venous blood returning to the heart-lung machine with the described flow rate usually will have an oxygen saturation of more than 50 percent. With flow rates and oxygen saturations in this range, metabolic acidosis of a significant degree does not occur. A drop in venous oxygen saturation or systemic acidosis suggests underperfusion, and bypass flow rates should be increased accordingly.

Heparin is metabolized gradually by the body and so additional heparin is given each hour of perfusion, as necessary, to keep the ACT above 600 s, usually 1 mg/kg of body weight. During perfusion the lungs are kept stationary in a partially inflated position.

Termination of Perfusion

When the operation is completed and the patient is systemically rewarmed to normothermic levels, perfusion is slowed and stopped. Before discontinuing bypass, the surgeon should check several important variables: ECG (for rate, rhythm, and ST segment changes), potassium level, hematocrit level, contractility of the heart, and hemostasis of the suture lines. As the perfusion flow rate is slowed, blood is infused from the pump to restore normal intracardiac pressures and to maintain an adequate blood pressure and cardiac output. Close observation of the heart for arrhythmias and adequate contractility is essential at this phase. Atrial pressure or pulmonary artery pressure is monitored. Intraoperative transesophageal echocardiography is useful in assessing myocardial wall motion and cardiac function. If a Swan-Ganz catheter is in place in the pulmonary artery, cardiac output is measured and attempts are made to maintain the cardiac index at a normal level. In many patients the pulmonary capillary wedge pressure (PCWP) provides a reasonable guide to left atrial pressure or preload. The preload should be optimized with fluids. If cardiac output and systemic blood pressure are inadequate despite an optimal preload, inotropic support is begun, usually with amrinone, dobutamine, or dopamine. Epinephrine and norepinephrine are reserved as second- or third-line agents. A decreased peripheral resistance because of vasodilation is treated with alpha agonists, such as phenylephrine hydrochloride (Neo-Synephrine).

Heparin is neutralized with protamine, giving sufficient protamine to return the activated clotting time as closely as possible to prebypass levels. Usually this requires 4 to 5 mg/kg of protamine, given in divided doses. If a coagulopathy is present, the activated clotting time may not return to prebypass levels, indicating the need for infusion of coagulation products, such as fresh frozen plasma, cryoprecipitate, or platelets. This occurs infrequently in routine cases but is more common as the complexity of the procedure and the length of the pump time increases. In longer procedures when a coagulopathy is noted and fibrinolysis may be present, aminocaproic acid is administered, but this is not used routinely.

Trauma from Perfusion

Extracorporeal circulation (cardiopulmonary bypass) inevitably produces some trauma to the blood, primarily from exposure of blood to plastics in the oxygenator and the extracorporeal circuitry, as well as from the use of suction to aspirate intracardiac blood. Minimizing the injury to blood during oxygenation is the rationale for using a membrane oxygenator rather than a bubble oxygenator. Trauma to blood from the pump itself has been minimized by the newer, less traumatic, centrifugal pumps. Tolerance for longer periods of perfusion, up to 4 h, is surprisingly good. Beyond this period, the number of complications from perfusion exponentially increases.

Significant changes in bodily functions occur during extracorporeal perfusion. These changes mainly involve a generalized systemic inflammatory response and probably are from the activation of complement and other acute-phase inflammatory components by the extracorporeal circuits. The severity of the inflammatory response and the level of subsequent end-organ dysfunction are related to the length of the pump time, with complement and cytokine activation leading to an up-regulation of white blood cell adhesion molecules and in the ability of white blood cells to release superoxide. White blood cell up-regulation or "priming" produced by extracorporeal circulation results in increased capillary permeability throughout the body, with the "primed" white blood cells placing the patient in a potentially vulnerable state for 24 to 48 h, during which any secondary insult may result in various levels of multiorgan dysfunction. Other effects may include confusion, renal insufficiency, decreased oxygen exchange, low-grade hepatic dysfunction, and hyperamylasemia. Current research is focused on minimizing the body's systemic inflammatory response during extracorporeal circulation by coating circuits with biocompatible materials or by blocking specific cytokines. Steroids and protease inhibitors, such as aprotinin, also minimize the inflammatory response to bypass.

A low-grade coagulopathy is not uncommon after cardiopulmonary bypass, but it is reduced with the newer materials. This coagulopathy also is related to the length of the pump time. Platelet dysfunction may occur because of the activation of platelets by artificial surfaces during bypass. Low levels of consumptive coagulopathy and hyperfibrinolysis from plasmin activation may be present. These findings generally do not become clinically relevant in the routine case but may become significant during longer, more complex procedures. The use in recent years of biocompatible surfaces has minimized the number of complications related to blood trauma and increased the safety of cardiopulmonary bypass. Similarly, the use of aminocaproic acid to decrease fibrinolysis and the protease inhibitor aprotinin may significantly lessen coagulopathic bleeding after prolonged extracorporeal circulation.

Myocardial Protection

The development of cold, hyperkalemic solutions for inducing cardioplegia, intraoperative cardiac arrest with myocardial preservation, was a major advance in cardiac surgery. This is based on the concept that a cold, potassium-arrested heart has a slow metabolic rate, allowing the surgeon to work precisely on the heart with a return of good cardiac function after completion of the procedure. The ability to protect the heart, combined with the increased ease of performing complex cardiac procedures in a dry, quiet field, greatly augments the safety and effectiveness of virtually all cardiac operations. This factor alone is probably responsible for the greatly improved results and wide applicability of cardiac surgery. Many of the complications in earlier years were probably because of myocardial injury and infarction. With current cardioplegia techniques the heart can be stopped and protected for 2 to 3 hours, allowing time for complicated procedures and nearly total recovery of cardiac function.

Crystalloid and blood cardioplegic solutions are widely used, with the exact components of the cardioplegic mixture varying among institutions. With periods of cardiac arrest for 60 to 90 min, there seems to be little measurable difference in the two techniques, but with the cold blood cardioplegia technique the heart can be safely arrested for longer periods if good distribution of the cardioplegic solution and uniform myocardial cooling are achieved.

The most widely used blood cardioplegic solution was developed by Buckberg and associates. This cardioplegic solution has a hematocrit level of 15 to 18 percent, providing red blood cells for oxygen delivery and a pH buffering effect. The cardioplegic solution is cooled to a temperature of 5 to 8°C to slow the basal metabolic rate of the myocardium and to lower oxygen consumption. A potassium concentration of 20 to 30 mEq/L in the initial cardioplegia injection induces mechanical arrest and ECG silence and also slows oxygen use because of a lack of wall tension. Subsequent maintenance injections given throughout the procedure have a potassium concentration of 8 to 10 mEq/L, which maintains hyperkalemic arrest. Other additives include CPD to lower calcium concentration and prevent calcium-induced reperfusion injury, buffer (tromethamine [THAM]) to scavenge hydrogen ions and prevent damage from tissue acidosis, and glucose, which provides metabolic substrate and makes the solution hyperosmolar to minimize edema.

In the setting of acute infarction or when ongoing myocardial ischemia is present Buckberg recommends that a form of resuscitative cardioplegia be used in which amino acids are added to the cardioplegic solution to provide substrate enhancement and that the final injection of cardioplegia be delivered at 37°C, allowing the heart to "repay" oxygen debt and repair cellular damage before resuming full contractile function. In experimental studies this technique has been found to minimize reperfusion damage of ischemic myocardium and

to enhance myocardial functional recovery. Similarly, mechanical unloading of the ischemic heart during the initial reperfusion interval has minimized reperfusion damage. Recent work has focused on the role of pharmacologic agents in providing “preischemic conditioning” of the heart, a technique that may additionally reduce the level of intraoperative myocardial injury.

The method used for intraoperative myocardial protection at New York University (NYU) is as follows. After the aorta is clamped, blood cardioplegic solution is infused at a rate sufficient to produce an aortic root pressure of 60 to 90 mmHg (200 to 400 mL/min), which should rapidly produce hyperkalemic arrest and ECG silence. Alternatively, retrograde delivery of cardioplegia through the coronary sinus of the heart is used, achieving a coronary sinus pressure of 20 to 30 mmHg. Retrograde administration of cardioplegic solution through the coronary sinus venous system has been a major advance in patients with severe disease, as this technique avoids maldistribution of cardioplegic solution from coronary artery stenosis. Cardioplegic solution distribution and myocardial cooling are confirmed by measuring regional myocardial temperatures in four regions of the myocardium (right ventricle and anterior, lateral, and posterior walls of the left ventricle), continuing infusion of cold blood until the “warmest” zone has cooled below 10 to 15°C. Topical hypothermia with cold saline irrigation is then used to prevent cardiac rewarming. Cardioplegic solution is reinfused after each anastomosis or after every 20 to 30 min. The left heart is kept in an empty, decompressed state by emptying the right heart with a large-lumen cannula or with a vent. Crooke and associates reported complete recovery of cardiac function after 2 h of protected cardiac arrest using the blood cardioplegia technique. With this and other, similar methods of myocardial protection longer, more complicated cardiac surgical repairs are now routine, and postoperative low cardiac output syndrome is exceedingly rare.

POSTOPERATIVE CARE AND COMPLICATIONS

General Considerations

Postoperative cardiac surgical care is a prime example of applied principles of cardiovascular physiology. Five key areas are involved: (1) hemodynamic evaluation, (2) electrocardiographic assessment, (3) blood loss, (4) ventilation and pulmonary care, and (5) general care, including fluid and electrolytes, nutrition, wound care, renal function, and rehabilitation.

After heart surgery, patients are observed in a specialized recovery room or intensive care unit. The ECG is monitored continuously, as are the arterial and intracardiac pressures. The Swan-Ganz pulmonary artery catheter is equipped with a thermistor for thermodilution measurement of cardiac output. Accurate hemodynamic profiles can be established on each patient as outlined below in the section on hemodynamic evaluation.

Initial laboratory information should include levels of hemoglobin, hematocrit, electrolytes, blood urea nitrogen, creatinine, and creatine phosphokinase isoenzyme, platelet count, *prothrombin time (PT)*, *partial thromboplastin time (PTT)*, and arterial blood gas and mixed venous blood gas pressures. The blood gases, hematocrit, and potassium determinations are repeated serially for 6 h, depending on the stability of the patient. Flow sheets are used to chart hourly laboratory reports, blood pressure, pulse, hemodynamic data, fluid input and output, and blood loss. Chest x-rays are periodically made to evaluate the mediastinal shadow and the lung fields.

Early postoperative complications include bleeding, tamponade, arrhythmias, myocardial infarction, graft occlusion, coronary spasm, low cardiac output syndrome, cardiac arrest, and stroke. Other complications include delayed bleeding, postpericardiotomy syndrome with pericardial effusion, tamponade, arrhythmias, renal dysfunction, ileus, ischemic bowel, gastrointestinal hemorrhage, pneumothorax, respiratory insufficiency, pneumonia, wound infection, wound dehiscence, and chronic cardiac dysfunction. While the incidence of serious complications is relatively low (3 to 6 percent), depending on patient and operative variables, every complication can be potentially life-threatening and is associated with significant morbidity.

While cardiologists, anesthesiologists, intensivists, and others are important members of the team and essential to the long-term care of the patient, the surgeon is responsible for the patient's overall care. The surgeon is best positioned to understand the subtleties of the patient's intraoperative physiology, technical problems encountered during a procedure, the adequacy of the myocardial protection, and other events that have a profound effect on the patient's clinical course. A strong commitment to the patient is an essential part of medical ethics and good patient care and should guide all actions.

Hemodynamic Evaluation

The adequacy of cardiac function is the key question in any patient after a cardiac operation. Adequacy of cardiac output is reflected in the blood pressure and the urine output, but exact measurement of cardiac output is far more precise. This can be done with a thermodilution technique if a Swan-Ganz catheter has been inserted, and this is the mainstay of treatment in any seriously ill patient. A normal cardiac index is 2.5 to 3.0 L/min/m². A cardiac index below 1.7 to 1.8 L/min/m² is an ominous finding, often resulting in death from inadequate perfusion of peripheral organs unless cardiac output can be increased. The classic clinical findings of low cardiac output with inadequate oxygen transport are the familiar ones of hypotension, vasoconstriction, oliguria, and metabolic acidosis. Untreated low cardiac output is ultimately fatal from progressive renal failure or arrhythmias.

When evaluating low cardiac output, the first consideration is to exclude cardiac tamponade or hypovolemia from intrathoracic bleeding. When these two factors have been excluded, the physiologic causes of low output should be reviewed in terms of preload, afterload, and intrinsic contractility of the heart. If the patient is hypovolemic, therapy consists of infusion of sufficient fluids to elevate left atrial pressure or pulmonary capillary wedge pressure to an appropriate level. As defined by the Starling principle, cardiac stroke volume rises with a rise in left atrial pressure. The cardiac output can be plotted against preload, assuming pulse and afterload are constant, to determine the “optimal” filling pressure of the heart.

Afterload reduction consists of reduction in peripheral vascular resistance with specific drugs to cause vasodilatation. If peripheral vascular resistance is elevated above the normal 1,200 dynes·sec·cm⁻⁵, afterload reduction should be one of the initial forms of therapy. The most commonly used drugs for intravenous infusion are sodium nitroprusside (Nipride) or nitroglycerin. Vasodilation, or decreased peripheral resistance, should be treated with vasoconstrictors to maintain an adequate perfusion pressure. Afterload should be controlled to keep the systolic blood pressure above 100 mmHg but below 150 mmHg and the mean arterial blood pressure above 70 mmHg but below 90 to 95 mmHg with a nearly normal vascular resistance (1,200 dynes·sec·cm⁻⁵).

Once bleeding and tamponade are excluded and preload and afterload have been optimized, a wide variety of inotropic agents may be used to augment myocardial contractility. Our first preference usually is amrinone or dobutamine, augmented if necessary with small amounts of norepinephrine, metaraminol, or phenylephrine when the peripheral vascular resistance is low. The alpha-adrenergic agents frequently are administered directly into the left atrium to minimize pulmonary hypertension. Dopamine and epinephrine are useful second-line inotropic agents.

If cardiac rhythm is not satisfactory, cardiac pacing should be used to maintain an adequate rate and rhythm. If a sinus mechanism is too slow or absent, atrial pacing or atrial-ventricular pacing is valuable for augmenting cardiac output. An optimal heart rate to maximize cardiac output without unduly increasing myocardial oxygen consumption is 80 to 90 beats/min.

If low cardiac output persists despite optimizing preload, afterload, and inotropic support, an intraaortic balloon pump may be necessary. The balloon pump (see Intraaortic Balloon Pump below): (1) decreases afterload, which improves the cardiac output while lowering wall tension and oxygen requirements; (2) decreases preload in the failing heart, lowering wall tension and oxygen demand; and (3) provides diastolic counterpulsation to augment systemic blood pressure and diastolic coronary perfusion. The balloon pump lowers oxygen consumption by approximately 20 percent while augmenting cardiac output by about 700 mL/min/m². The need for a balloon pump can be determined in the operating room by performing serial measurements of cardiac output after the cessation of cardiopulmonary bypass. If the cardiac index remains below 1.5 to 1.7 L/min/m² despite significant administration of inotropes, an intraaortic balloon usually is recommended.

An unexpected fall in cardiac output in the recovery room is frequently because of a correctable mechanical problem. A thorough search for various causes of low output syndrome is essential before resorting to mechanical support. If correctable problems are not present and the low output syndrome does not respond to a combination of inotropes and a balloon pump, temporary placement of a left ventricular assist device may be necessary.

Electrocardiographic Assessment

The postoperative ECG is important for determining heart block, bundle branch block, infarction (Q waves), ischemia (ST-segment elevation, T-wave inversion), or other signs of intraoperative injury. The ECG should be repeated periodically, and patients with signs of ischemia or injury should be monitored in the intensive care unit. Acute ischemia or evolving infarction should be treated initially with nitrates or calcium blockers. If ECG changes do not resolve promptly, a return to the operating room to rule out graft occlusion should be considered, particularly if the patient is unstable or hemodynamically compromised. In addition to standard 12-lead ECGs, continuous monitoring of the cardiac rhythm for at least 2 days after operation is important. The patient is monitored in the recovery room or intensive care unit for 24 h. Once transferred to the ward, some form of telemetry with an appropriate alarm mechanism usually is used for an additional 24 h. Life-threatening arrhythmias may develop unexpectedly despite the presence of a normal cardiac output and without any other signs of circulatory failure. Delayed detection of a significant arrhythmia is a major cause of unexpected death after cardiac operations. Bradyarrhythmias and heart block are not uncommon postoperatively, and for this reason temporary cardiac pacing wires are routinely left in the right ventricle and right atrium for several days.

Ventricular extrasystoles and ventricular tachycardia are more serious because their appearance may herald the development of ventricular fibrillation. Hypokalemia should always be considered because patients in cardiac failure preoperatively may have significant depletion of body stores of potassium from chronic diuretic therapy. The serum potassium level should be kept well above 4.0 mEq/L. Continuous intravenous lidocaine, 1 to 4 mg/min, is a valuable form of therapy for temporary control of ventricular arrhythmias. Procainamide generally is used as a second-line agent and bretylium as a third-line agent for serious ventricular arrhythmias.

The treatment of atrial fibrillation, the most common postoperative arrhythmia, begins with heart rate control and includes the use of intravenous digitalis in conjunction with a beta blocker or a calcium-channel blocker. Once rate control has been achieved, a class IA antiarrhythmic agent, such as procainamide or quinidine, can be used to attempt “chemical cardioversion,” or the patient can be given anticoagulation therapy with warfarin and discharged. A significant number of patients with persistent postoperative atrial fibrillation spontaneously convert to normal sinus rhythm within 6 weeks; for those who do not, electrical cardioversion can be used. Atrial fibrillation resulting in hemodynamic instability should immediately be treated with cardioversion at 50 to 100 J. Virtually all antiarrhythmic agents cause serious side effects in a small proportion of patients, particularly when administered intravenously, so these patients require careful periodic monitoring.

Cardiac Arrest

Complete circulatory collapse in the postoperative cardiac surgical patient can occur without warning. If effective circulation is not restored within minutes, the brain and myocardium suffers anoxic injury that may become irreversible. Determining the cause of the circulatory arrest may indicate a specific therapy; however, treatment should not be delayed if the exact cause is unknown. A high incidence of survival without any permanent disability is possible when cardiac arrest is recognized promptly, circulation is quickly restored, and the underlying cause is treatable. Frequent causes of cardiac arrest after cardiac surgery include cardiac tamponade, arrhythmias, ischemia or graft occlusion, hypoxia, and drug toxicity.

Cardiac tamponade is a serious complication after cardiac surgery. It may occur early in the postoperative period from the accumulation of intrapericardial blood. Later in the postoperative course, tamponade can occur from a pericardial effusion. The classic findings of tamponade include: (1) elevation of central venous pressure; (2) equalization of central venous pressure, pulmonary artery diastolic pressure, and left atrial pressure; and (3) a pulsus paradoxus of more than 10 mmHg during inspiration. Extreme cases of tamponade may present with life-threatening hypotension or sudden circulatory collapse; the diagnosis should be strongly considered in any patient with hypotension and a low cardiac output. A widening of the mediastinal shadow on chest x-ray or detection of significant pericardial effusion by echocardiography is suggestive of the diagnosis. No single test can exclude tamponade short of surgical exploration. Because of this, any patient with suspected tamponade should be returned promptly to the operating room for definitive diagnosis and treatment. Any patient in extremis with suspected tamponade should have his or her chest opened or subxiphoid drainage performed at the bedside in order to rule out tamponade, which is treated effectively by clot and fluid removal.

Ongoing ischemia and myocardial injury secondary to poor intraoperative myocardial protection are probably the two most common causes of postoperative ventricular fibrillation. Ischemia and ventricular irritability may be from graft occlusion, preoperative ischemia or injury, intraoperative injury, coronary spasm, or a nonbypassed area of myocardium.

An electrolyte imbalance, such as a deficiency of potassium or magnesium, also can cause serious arrhythmias leading to cardiac arrest or ventricular fibrillation. A serum potassium level below 3.0 mEq/L can produce severe cardiac irritability postoperatively, though the precise influence is determined by the coexisting concentration of calcium ions and the presence of acidosis or alkalosis along with the patient's body temperature.

Difficulties with ventilation and hypoxia can lead to cardiac arrhythmias from low arterial oxygen tension and progressive metabolic acidosis. Inadequate ventilation from pneumothorax, from dislodgment of the endotracheal tube, or from plugging of the airway with secretions are common causes.

Drugs may induce bradycardia, heart block, ventricular fibrillation, or cardiac arrest, from toxicity or an idiosyncratic reaction. Digitalis is a common offender because of its widespread usage. The sensitivity of the myocardium to digitalis varies with a number of factors, one of the most important of which is the concentration of potassium. Procainamide and quinidine are examples of drugs with known proarrhythmic effects.

Profound bradycardia from any cause (heart rate < 60 beats/min) may result in escape beats leading to ventricular fibrillation and cardiac arrest. Ventricular arrhythmias may progress to bigeminy, ventricular tachycardia, and ventricular fibrillation. This well-known sequence is the reason for constant monitoring of the ECG postoperatively.

General Considerations for Treatment

Through the efforts of the American Heart Association, Advanced Cardiac Life Support (ACLS) has been taught to almost every health care worker in the United States. ACLS guidelines include specific algorithms for the treatment of cardiac arrest and other causes of circulatory arrest. Although the treatment of cardiac arrest described in the remainder of this section apply mainly to postoperative cardiac surgery patients, the principles discussed follow ACLS guidelines and also are applicable to noncardiac surgical patients.

Cerebral anoxia associated with circulatory arrest produces brain injury within 3 to 4 min, so the diagnosis must be made and treatment begun rapidly to avoid serious brain injury. Periods of anoxia for 6 to 8 min may produce extensive but reversible brain injury, whereas longer periods regularly cause irreversible injury. Closed-chest massage and ventilation should be started promptly. The ABCs of cardiopulmonary resuscitation (CPR) are: airway, breathing, and circulation.

The immediate first steps in CPR are to secure an adequate airway and provide prompt ventilation (*breathing*). Cardiac massage for more than a few seconds without securing the airway and providing adequate ventilation (breathing) is futile. The mouth and throat are cleared of secretions, and ventilation is quickly accomplished by mask inflation with an oral airway. This can be begun immediately and continued until endotracheal intubation is achieved. With a laryngoscope, an endotracheal tube can easily be inserted by a physician or other trained personnel. In some patients, such as those with a short, thick neck, the anatomy is such that intubation is difficult even for highly experienced staff. Unless intubation can be accomplished quickly and with certainty, oral insufflation should be continued until a cricothyroidotomy has been performed.

An infrequent but serious error can occur when the endotracheal tube is inadvertently placed in the esophagus. Because of this possibility, immediate auscultation of the lungs for breath sounds is essential after attempted intubation. If any uncertainty remains after auscultation, the tube should be removed and intubation should be repeated. A tightly fitting face mask can provide a method of temporary ventilation. If endotracheal intubation is difficult, requiring repeated attempts, or simply cannot be performed, a cricothyroidotomy should be performed promptly. The patient cannot recover unless the airway is rapidly controlled and ventilation established.

The third step is to provide effective perfusion (*circulation*) with closed-chest massage or with inotropes or pressors. Effective CPR depends on adequate intermittent compression of the heart between the sternum and the vertebral column. The patient must be on a firm surface, which can be accomplished by placing a board under the back. The heel of the hand should be applied over the lower third of the sternum with the other hand above it to depress the sternum 3 to 4 cm intermittently. Massage should be at a rate of about 60/min. The sternal compression should be brisk, depressing the sternum sharply and then releasing it to permit cardiac filling. Mechanical ventilation must be synchronized with massage.

In the postoperative cardiac surgery patient with cardiac arrest closed-chest massage may be initiated immediately as part of a resuscitative effort, which also may include defibrillation and the administration of pharmacologic agents. If cardiac activity is not quickly restored or if the cardiac arrest is thought to be a result of a mechanical cause, such as tamponade or hemorrhage, the sternotomy incision is reopened immediately at the bedside and internal cardiac massage is instituted. Open cardiac massage is highly effective if instituted promptly. For this reason, postoperative resuscitative efforts employing closed cardiac massage and external defibrillation are quickly aborted unless immediately successful. Rarely, open-chest massage may be effectively continued for more than an hour until the basic condition causing refractory arrhythmias or cardiac arrest is corrected. More commonly, the patient is returned to the operating room once the chest is opened and cardiopulmonary bypass is initiated for support. After the patient is resuscitated the surgeon must determine if further therapy such as mechanical support or additional bypass grafting is indicated.

Drugs and Fluids

Epinephrine, sodium bicarbonate, and calcium are the most useful pharmacologic agents. The protocols suggested by the ACLS program are followed. Epinephrine 1mg may be given intravenously or by direct intracardiac injection. Calcium chloride (3 to 4 mL of 10% solution) is another powerful stimulant of myocardial contraction. Calcium administration in cardiac arrest may be valuable in some circumstances but harmful in others and should not be used indiscriminately. Lidocaine is given primarily for ventricular arrhythmias, followed by bretylium or procainamide for refractory ventricular tachycardia. Medications usually are ineffective if severe myocardial anoxia or significant acidosis is present. Anoxia can be corrected only by the combination of effective cardiac massage and ventilation. Large amounts of sodium bicarbonate may be required to treat the acidosis. Excessive use of drugs before the hypoxia and acidosis have been treated is probably futile.

Small amounts of fluid should be rapidly infused because vasodilatation usually is present. An intravenous infusion of a vasoconstrictor, norepinephrine or phenylephrine (Neo-Synephrine), is often helpful in order to maintain perfusion pressure. Blood transfusion may be required in the postoperative patient for volume resuscitation and improved oxygen-carrying capacity.

Defibrillation

Ventricular fibrillation can be differentiated from asystole only by the ECG or by direct inspection of the myocardium. Initial treatment of cardiac arrest should include a precordial thump, because it is often effective in converting ventricular tachycardia or ventricular fibrillation. Intravenous lidocaine is promptly administered. If this is not effective quickly and an ECG is not available, empiric defibrillation can be tried briefly, because most resuscitations are effective when defibrillation is done promptly. Asystole is treated by pacing the heart with a transcutaneous pacemaker, a transvenous pacemaker if access is readily available, or a temporary pacing lead that can be placed directly on the myocardium if the chest is open.

Closed-chest defibrillation usually is done by applying electrodes over the base and apex of the heart. Defibrillation is best done with a direct current of 200 to 360 J. When open-chest defibrillation is used, the electrodes are applied directly to the heart and an impulse of 20 to 40 J is delivered. Vigorous cardiac massage should precede defibrillation to oxygenate the myocardium sufficiently. Perfusion pressure must be maintained to obtain successful defibrillation because ongoing subendocardial ischemia occurs when the perfusion pressure is inadequate. Correction of acidosis with bicarbonate can be confirmed by blood gas determinations. Intramyocardial injection of epinephrine may stimulate myocardial tone and improve the chances of successful defibrillation. Unless a massive myocardial infarction has occurred, it should be possible to defibrillate the majority of fibrillating hearts, though the ensuing cardiac arrest may be refractory to therapy. The most significant factor influencing survival is the institution of defibrillation within 1 min of onset of fibrillation or the ability to defibrillate with less than five shocks.

After restoration of an adequate heart beat and blood pressure, the critical question is the extent of injury to the heart and to the central nervous system. A thorough search for reversible causes of the arrhythmia is undertaken immediately. If ischemia or graft occlusion is suspected, the patient is taken for cardiac catheterization or back to the operating room.

A detailed neurologic evaluation to elicit specific normal and abnormal reflexes should be done promptly. Permanent brain injury is frequent after more than 5 min of cardiac arrest, even though experimentally cerebral neurons can tolerate nearly 20 min of normothermic ischemic anoxia. Intravenous steroid therapy usually is given for 24 to 48 h to minimize cerebral edema, although the efficacy is difficult to measure. If serious cardiac or cerebral injury is not present after resuscitation, the prognosis is excellent. This fact is the basis for the enthusiastic development of widespread training in cardiopulmonary resuscitation by all physicians, paramedical personnel, and lay people, for most effective resuscitation is accomplished when begun within 1 to 2 min after onset of cardiac arrest.

Blood Loss

Blood conservation and minimization of bleeding associated with cardiac surgery begin preoperatively. Preoperative work-up routinely includes a PT, PTT, and platelet count. Patients with a history of abnormal bleeding or with chronic passive congestion of the liver receive full coagulation profiles and evaluation by a hematologist. Patients taking warfarin are instructed to discontinue the drug 3 to 4 days before operation so that the PT can return to normal.

Because blood conservation is highly desirable, many patients operated on electively are able to donate autologous blood 1 to 3 weeks before surgery. Often erythropoietin is used after autologous donation to enhance the patient's red blood cell mass. Donor-directed blood units are solicited from family and friends. Intraoperatively, patients with a large blood volume and an adequate hematocrit level may have 1 unit of fresh whole blood removed before bypass. The unit is saved and reinfused after bypass in order to take advantage of the fresh plasma and uninjured platelets. All blood in the operative field is collected during the procedure, either by suction into the cardiopulmonary bypass pump or into a cell-saver device. Similarly, mediastinal blood shed postoperatively through the chest tubes is collected sterilely and reinfused. By reinfusion of blood from the pump, cell saver, and

chest tubes, the need for blood transfusion associated with heart surgery has been diminished significantly, and many patients avoid transfusion altogether.

Postoperative Bleeding

For most open-heart cases coagulopathy is nonexistent and the postoperative blood loss is low. On completion of cardiopulmonary bypass, heparinization is reversed with protamine and chest closure is initiated only after hemostasis has been achieved. The normal postoperative blood loss should range from a total of 300 to 800 mL. Blood loss in excess of 300 to 500 mL/h or over 1 to 1.5 L total usually indicates active surgical bleeding and is associated with a high incidence of hemodynamic compromise or cardiac tamponade. For this reason it usually is mandatory that patients with active bleeding or with a total blood loss greater than 1 to 1.5 L be returned to the operating room to control any active bleeding and avoid tamponade. If surgical bleeding is not found and a coagulopathy is present, appropriate treatment is initiated.

In some patients the effects of cardiopulmonary bypass can be damaging to the blood coagulation system, resulting in abnormal bleeding. The incidence of coagulopathy is related most strongly to the duration of cardiopulmonary bypass, but coagulopathy also is more frequent when hemodilution is excessive, after severe hypothermia, in patients who received aspirin or recent thrombolytic therapy, and in reoperations in which ongoing blood loss from scar tissue results in progressive loss of coagulation factors.

The diagnosis of coagulopathy is made by the operating surgeon with the observation of abnormal bleeding from the operative field in the absence of a surgical source. Laboratory tests can confirm the diagnosis, but treatment should not be delayed until test results return because this might worsen the coagulation deficit with fatal consequences. Nonsurgical causes of abnormal bleeding after heart surgery include: (1) inadequate neutralization of heparin from insufficient protamine; (2) a functional platelet deficit from aspirin effect or from the activation of platelets by the cardiopulmonary bypass circuit; (3) dilutional coagulopathy from the combination of crystalloid priming volume and transfusions; (4) a consumptive coagulopathy from a low-grade activation of the clotting factors by the bypass oxygenator and circuit; and (5) abnormal fibrinolysis.

Treatment of coagulopathy is urgent. Hypothermia should be corrected and extra protamine should be given until the activated clotting time returns to normal or until no further drop is seen in the activated clotting time. If abnormal bleeding persists, transfusion of platelets, fresh frozen plasma, and cryoprecipitate are given until the clotting deficit is corrected. Antifibrinolytic agents such as *e*-aminocaproic acid may be given to correct fibrinolysis. Heparin-bonded circuits and oxygenators are used to minimize coagulopathy in high-risk cases. The pharmacologic prevention of coagulopathy is more effective if the protease inhibitor aprotinin or the antifibrinolytic agent *e*-aminocaproic acid is administered preoperatively.

Ventilatory and Pulmonary Insufficiency

Nearly all cardiac surgical patients return to the recovery room intubated. While ventilating a patient after heart surgery, the physician should periodically assess the breath sounds, particularly in the posterior bases of the lungs, to determine the adequacy of ventilation. Tidal volumes on the respirator usually are set at 10 mL/kg, but occasionally up to 15 mL/kg is necessary in patients with chronic obstructive disease. For most cases, auscultation of the bases and determination of lung compliance (static lung compliance = tidal volume/end inspiratory pressure) are the best ways to determine the optimal tidal volume, with the maximum lung compliance considered optimal compliance for most cases. The adequacy of ventilation and oxygenation is then checked by periodic blood gas determinations. For patients with marginal oxygenation the optimal tidal volume and positive end-expiratory pressure (PEEP) may be better determined on the basis of oxygen delivery, with oxygen delivery equal to cardiac output times arterial oxygen content.

Weaning the patient is done via the intermittent mandatory ventilation (IMV) mode or via progressive continuous positive airway pressure (CPAP) trials. Criteria for extubation include adequate blood gases on CPAP, with an inspired oxygen of less than 50 percent, a respiratory rate of less than 15 to 20 breaths/min, a spontaneous tidal volume of 5 mL/kg, a forced vital capacity of 10 mL/kg, and a negative inspiratory pressure of 15 mL of water as determined by bedside pulmonary function tests. The patient should be awake enough to control the airway. Most patients can be extubated within 6 to 12 h of surgery.

Postextubation pulmonary care involves use of the incentive spirometer, coughing, and deep breathing. Control of secretions is essential. Difficult cases might require the use of inhaled bronchodilators, such as the β_2 -adrenergic agonist albuterol, or the use of intravenous aminophylline. Rarely, a short course of steroids is beneficial. Tracheobronchitis should be treated promptly with antibiotics. Observation of daily sputum specimens is important in infected patients, because clearing of the sputum and a drop in the white blood cell count indicate effective therapy. Clearing of secretions by coughing or suctioning is essential if the patient with tracheobronchitis is to recover. Flexible bronchoscopy has been effective in assessing the adequacy of suction and for evaluating the clearance of secretions in the refractory ventilator-dependent patient with tracheobronchitis.

Operation on the patient with a long history of smoking and with chronic obstructive pulmonary disease poses a special problem. These patients should be strongly counseled on the need to stop smoking preoperatively with the aid of transdermal nicotine patches if necessary. Evidence suggests that discontinuing smoking for even 1 week is markedly beneficial. Poor pulmonary function, however, usually does not prohibit cardiac surgery. For example, if a patient can function, albeit poorly, with substandard pulmonary reserve and critical aortic stenosis, the patient's overall cardiorespiratory function should significantly improve after correction of the valvular pathology. Sternotomy has relatively little deleterious effect on pulmonary reserve. A study by Bevelaqua and colleagues noted that heart patients with poor pulmonary function had more short-term postoperative pulmonary complications than other patients but did not have a higher overall hospital mortality rate.

With current preoperative preparation and intraoperative management, significant postoperative impairment of pulmonary function is uncommon, except in patients with preexisting pulmonary disease or advanced cardiac failure. Most patients are now extubated within 8 to 12 h, although mechanical ventilation for 24 to 72 h is well tolerated by most patients. If periods of ventilation longer than 3 or 4 days are anticipated, a cricothyroidotomy or tracheostomy should be considered. In addition to providing patient comfort and safety and eliminating dead space, pulmonary toilet is more effective when done through a tracheostomy or cricothyroidotomy. Ventilatory support for more than a short time seldom is necessary except in chronically ill or elderly patients in whom simple physical weakness may significantly impair the effectiveness of breathing and coughing. These patients may require ventilatory support for days or even weeks.

General Care

Nutrition

The need for adequate postoperative nutrition cannot be overemphasized, particularly in elderly or chronically ill patients. The sick postoperative heart patient may require 25 to 35 kcal/kg/day. Care should be taken to give adequate protein in patients with normal renal and hepatic function, and special formulas are available for patients with kidney or liver failure. Ill patients who remain intubated in the intensive care unit should have nutritional support started early as tube feeding or as intravenous hyperalimentation.

Wound Care

Early postoperative care of the surgical wound consists of the use of a sterile occlusive dressing for the first 24 h. Wounds may be painted twice a day with a povidone-iodine solution or dressed sterilely and examined daily for redness, drainage, or sternal instability. Prophylactic antibiotics are started preoperatively and continued for 24 to 48 h postoperatively, usually until indwelling catheters and chest tubes have been removed. The prophylactic antibiotic used should be chosen on the basis of the organisms commonly causing infection in that hospital.

All fever should be recorded and the patient examined for signs of infection. A moderate fever of 100 to 101°F is common in the first 1 to 2 days, usually resulting from the systemic inflammatory response induced by the extracorporeal circulation and the stress of major surgery or from atelectasis. A fever occurring from 3 to 7 days postoperatively with a normal white blood cell count is frequently from postpericardiotomy syndrome, which results from a combination of a generalized inflammatory response and localized irritation of the pericardium from surgical trauma. Postpericardiotomy syndrome may be associated with a friction rub and pericardial or pleural effusions. The syndrome can be treated with nonsteroidal anti-inflammatory agents or, occasionally, with a short course of steroids. Significant fevers should be evaluated with a chest x-ray and blood and urine cultures.

A serious sternal wound infection occurs in 1 to 2 percent of all open heart operations. Risk factors include diabetes, obesity, impaired nutrition, chronic obstructive pulmonary disease, prolonged ventilatory support, harvest of bilateral internal mammary arteries, older age, low cardiac output, excessive mediastinal bleeding, reexploration for bleeding, multiple transfusions, and prolonged cardiopulmonary bypass. The incidence is highest in diabetics receiving bilateral internal mammary artery grafts. Signs of infection include redness or pain in the wound, purulent drainage, sternal instability, fever, and an elevated white blood cell count. The diagnosis is made easily by sternal aspiration. Early diagnostic sternal aspiration is strongly recommended for any patient with persistent localized sternal pain or unexplained sepsis. *Staphylococcus aureus* and *Staphylococcus epidermidis* are the most common organisms isolated. Treatment requires prompt operation with debridement and closure over antibiotic irrigation/drainage catheters or debridement followed by immediate or delayed closure with muscle flaps. The mortality rate from a sternal infection after cardiac surgery is 10 to 20 percent.

Renal Function

Close attention to renal function is necessary. The urine output for most adults should be 0.5 to 1 mL/kg/h. The blood urea nitrogen and creatinine levels are followed for several days postoperatively. Patients with a marginal urine output or with borderline renal function may benefit from low-dose (3mg/kg/min) dopamine. Some degree of salt and fluid retention often is present during postoperative days 3 to 5, even in patients with normal renal function, and diuretics may be necessary until the patient returns to preoperative weight. In patients with progressive renal dysfunction early dialysis is indicated to control volume overload and to minimize the risk of arrhythmias and infection. Permitting the blood urea nitrogen level to rise above 90 to 100 mg/dL may result in serious cardiac arrhythmias and immunocompromise. The mortality rate for patients with anuric renal failure after heart surgery exceeds 25 percent.

Rehabilitation

Most patients recover uneventfully after heart surgery and do not require a rigorous rehabilitation program. Recommendations include daily walking in the halls and a progressive walking schedule of up to 2 miles/day after discharge. The patient should follow normal moderate activities for the first 6 to 8 weeks. More vigorous exercise should be preceded by an exercise stress test, which is recommended after 2 to 3 months; this amount of time is required for the patient to recover physically and intellectually. Elderly and debilitated patients should receive a more regimented rehabilitation effort, beginning with physical therapy during the hospital stay and often continuing with a formalized rehabilitation program for the first 2 to 3 weeks after discharge.

CORONARY ARTERY DISEASE

Historical Data

Starting in the late 1930s different investigators attempted to increase the blood supply of the ischemic heart by developing collateral circulation with vascular adhesions. Beck was the leading investigator, trying different methods for many years, but ultimately all failed. A separate ingenious concept arose in 1946 when Vineberg developed implantation of the internal mammary artery into a tunnel in the myocardium. Vineberg applied this clinically in 1950 and continued for many years. The artery remained patent in over 90 percent of patients, but the amount of flow through the artery was small, and the procedure was eventually abandoned. Coronary endarterectomy without bypass grafting was attempted by Longmire in 1956 with short-term success but a high reocclusion rate from progressive fibrosis in the arterial wall. Patch graft reconstruction of coronary arteries, as proposed by Senning in 1961, was attempted subsequently by Effler and colleagues with disappointing results.

In the 1960s the development of the coronary bypass operation was a dramatic medical milestone. In the United States the principal credit for developing coronary artery bypass belongs to Favalaro, Effler, and associates from the Cleveland Clinic who did the first series of coronary bypass grafts beginning in 1967 using cardiopulmonary bypass and vein grafts to bypass the right coronary artery. Johnson and associates showed that similar grafts could be effectively used for the left coronary system. These were quantum achievements that launched the modern era of coronary bypass grafting.

An additional breakthrough, which was not recognized initially as being particularly significant, came in 1968 when Green performed the first left internal mammary artery to the left anterior descending artery bypass using cardiopulmonary bypass and an operative microscope. This followed earlier experimental work by Spencer, who had demonstrated the feasibility of this type of anastomosis in 1964 in a canine model. Kolessov had independently performed internal mammary artery to left coronary artery bypass grafting on the beating heart in Russia in 1967, although his work was largely unknown for many years. The internal mammary artery was not widely used at first, but since the early 1980s, after 10-year angiographic follow-up studies reported long-term patency rates exceeding 90 percent, it has been widely adopted.

Since 1975 coronary artery bypass surgery has become one of the most widely applied surgical procedures in the United States. The bypass operation has had a monumental impact on the treatment of coronary artery disease and came to symbolize the technical advances of Western medicine. Over 350,000 coronary bypass procedures are performed yearly in the United States alone.

Etiology and Pathogenesis

Atherosclerosis is the fundamental cause of coronary artery disease. The disease is multifactorial, involving serum lipids and cholesterol metabolism, local vascular mitogens such as platelet-derived growth factor, basic fibroblast growth factor, or transforming growth factor beta and other risk factors, such as smoking, hypertension, and diabetes, which lead to injury of the vessel wall. Atherosclerosis is more common in males than in females, with a 4:1 ratio. The frequency of coronary artery disease varies widely throughout the world, and it is less common in populations with an average blood cholesterol below 200 mg/dL. The frequency is lowest in Japan, where the average blood cholesterol is near 160 mg/dL. The United States has the second highest frequency in the world.

The basic lesion is a segmental plaque within the coronary artery. This segmental localization makes bypass grafts feasible. Involvement of small distal vessels usually is less extensive, while arterioles and intramyocardial vessels usually are free of disease. The proximal anterior descending artery frequently is stenosed or occluded, with the distal half of the artery remaining patent; the right coronary artery often is stenotic or occluded throughout its course, but the posterior descending artery and left atrial-ventricular groove branches are almost always patent. The circumflex artery often is diseased proximally, but one or more distal marginal branches usually are patent.

The popular terminology of single-, double-, or triple-vessel disease refers to the number of coronary arteries involved. In over 50 percent of patients, “triple-vessel” disease is present.

Clinical Manifestations

Myocardial ischemia from coronary artery disease can produce several serious events: angina pectoris, myocardial infarction, congestive heart failure, or sudden death. Angina is the most frequent symptom, but myocardial infarction or sudden death can appear without warning. Congestive heart failure usually results as a sequela of myocardial infarction, with significant muscular injury resulting in ischemic myopathy.

Angina pectoris, the most common symptom, is manifested by periodic discomfort, usually substernal and typically appearing with exertion, after eating, or with extreme emotion. Characteristically these symptoms subside within 3 to 5 min or are dramatically relieved by sublingual nitroglycerin. In about 25 percent of patients the symptoms are not typical and may radiate to bizarre areas, such as the teeth, the shoulder, or the epigastrium. Establishing a diagnosis of angina in these patients is difficult and perhaps impossible without diagnostic studies. Physical examination usually is normal. Differential diagnosis includes anxiety states, musculoskeletal disorders, and reflux esophagitis or esophageal spasm.

The risk of sudden death varies with the extent of disease and the degree of impairment of ventricular function, ranging from 2 to 10 percent. Death apparently results from ventricular fibrillation in many cases, or from myocardial infarction with acute decompensation.

Myocardial infarction is the most common serious complication, with 900,000 cases occurring in the United States annually. With modern therapy the mortality rate is 5 to 8 percent. Most deaths occur in the first 1 to 2 h, often before the patient ever reaches a hospital. With modern treatment in coronary care units and early reperfusion the fatality rate is small.

In a small proportion of patients congestive heart failure eventually develops, resulting from multiple infarctions that ultimately destroy over 40 percent of the left ventricular muscle mass. When areas of salvageable myocardium are still present, often manifested as angina associated with the heart failure, significant improvement can be obtained with revascularization. In contrast, with late-stage chronic congestive failure, because of diffuse myocardial scarring, the outlook is ominous. Bypass grafting may be futile in these circumstances unless reversible ischemia can be demonstrated on PET scan or thallium scan or unless there is a mechanical defect that can be corrected, such as mitral valve insufficiency or ventricular aneurysm. Cardiac transplantation is required in this group of patients with severe ischemic cardiomyopathy and nonreversible ischemia.

Preoperative Evaluation

A general history and physical examination should be performed in every patient along with a chest x-ray and an ECG. The resting ECG is normal in over 50 percent of patients. The other commonly used diagnostic tests—radionuclide studies, stress or provocative stress tests, and echocardiography—were described earlier (see Diagnostic Studies). The usual purpose of these studies is to assess the level of ischemia in response to stress, the functional significance of the coronary disease, and the level of cardiac function. These studies are often used to decide whether or not cardiac catheterization and subsequent revascularization is indicated.

For patients with coronary artery disease cardiac catheterization remains the cornerstone of evaluation, for it outlines the location and the severity of the coronary disease, accurately assesses ventricular function, and determines the presence of associated valvular pathology. The three most important prognostic factors determined angiographically are the number of vessels involved, the location and severity of the stenoses, and the ventricular function as measured by the left ventricular ejection fraction.

“Angiographically significant” stenosis is considered present when the diameter is reduced by more than 50 percent, corresponding to a reduction in cross-sectional area greater than 75 percent. The number, location, and severity of the coronary stenoses are used to determine the subsequent mode of therapy, i.e., bypass surgery, angioplasty, or medical therapy.

The long-term course of coronary disease is determined by the balance between two opposing factors, the rate of progression of the atherosclerotic stenoses versus the rate of development of the collateral circulation. Ventricular function probably reflects the ability of the heart to develop sufficient collateral circulation to compensate for the arterial stenoses present. Collateral vessels develop in response to an “ischemic gradient,” probably from chemical mediators or ischemia-induced growth factors that stimulate collateral vessel growth, but this physiologic ability varies widely; some patients with extensive triple-vessel disease have normal ventricular function with occluded coronary arteries and well-developed collaterals, while others with less severe disease have marked impairment in cardiac function and no collaterals.

Ventricular function usually is expressed as the left ventricular ejection fraction, with 0.55 to 0.70 considered normal; 0.35 to 0.55, moderately depressed; and below 0.35, especially below 0.20, severely depressed. An ejection fraction below 0.35 is often associated with intermittent congestive heart failure, and below 0.20 to 0.25, with severe heart failure. When a ventriculogram is evaluated, the contraction of individual segments of ventricular wall—i.e., regional wall motion—is separately analyzed. Segmental wall motion is classified as normal, hypokinetic (impaired), akinetic (little or no visible contraction), or dyskinetic (paradoxical contraction, as with a left ventricular aneurysm). Regional wall motion abnormalities are used to help determine zones of injury, while the global ejection is used to determine the overall cardiac function, operative risk, and the long-term prognosis.

Although angiography and ventriculography are the most precise methods for evaluating coronary disease, several limitations of the technique also should be emphasized, for erroneous decisions can be made easily. An angiogram indicates the severity and complexity of the disease but is not an absolute guide to “operability.” Other clinical and diagnostic data must enter into this decision. It is often a serious error to conclude from the ventriculogram that a diseased artery supplying an akinetic zone should not be bypassed because that segment of the ventricle is “scar.” Other studies, such as the PET scan or thallium scan, may reveal reversible ischemia in these areas that would then benefit from revascularization. A patient is not “inoperable” because the global ventricular function is severely depressed. If a significant amount of myocardium has signs of reversible ischemia and is thought to be viable but “hibernating,” the patient usually will benefit from operative revascularization.

Coronary Bypass Indications

The indications for coronary artery bypass usually involve one of three general categories: (1) chronic angina of varying severity, (2) unstable angina, and (3) complications of acute myocardial infarction or postinfarction angina.

Chronic Angina of Varying Severity

There is general agreement that patients with severe angina not responding to medical therapy should be electively operated on. The proper therapeutic choice is less clear when patients present with significant coronary disease but with less severe symptoms. Usually the coronary artery disease was detected because of mild angina or because of a positive stress test. Since angina is not severe, revascularization to improve symptoms or quality of life is not an issue. The key question in these patients is: “Will survival be improved by revascularization?” Factors such as the severity and distribution of the coronary stenoses and degree of myocardial dysfunction are used to determine which patients will benefit from operative revascularization, since patients with extensive disease and poor ventricular function do not do well with medical therapy. Other tests that evaluate the functional significance, the amount of myocardium at risk and the level of activity that induces ischemia are useful for assigning risk levels to patients to determine potential benefit from surgical therapy.

Three historical randomized studies attempted to answer the question of which patients with less severe angina would obtain a survival benefit from coronary bypass.

(1) *The Veterans Administration Cooperative Study.* This study involved 668 males treated between 1972 and 1974. The study demonstrated improved long-term survival in patients with left main disease treated with surgical therapy. Since this trial most patients with significant left main disease have been treated with coronary bypass revascularization regardless of their symptoms because of the survival advantage shown.

(2) *The European Coronary Surgery Study Group.* This study (1973–1976) randomized men under 65 years of age, 57 percent with mild angina and 42 percent with moderate angina, into medical or surgical therapy. Equal survival rates were found in patients with single-vessel disease. In patients with double-vessel disease in which a proximal anterior descending artery lesion was present and in patients with triple-vessel disease, a long-term survival advantage was noted in patients treated surgically. The study recommended

initial surgery for patients with double- vessel disease and a proximal anterior descending artery lesion and for patients with triple-vessel coronary disease.

(3) Coronary Artery Surgery Study (CASS) . This multicenter study was done in the United States from 1975 to 1979, randomizing initial treatment for patients with mild angina and operable coronary disease. Patients treated medically could be crossed over to surgery at the discretion of the cardiologist for worsening angina. After 10 years, results suggested that survival was improved by surgery in patients with triple-vessel disease when cardiac function was depressed. No survival advantage was demonstrated from early surgery in patients with single-, double-, or triple-vessel disease when ventricular function was normal. Survival was calculated on initial intent for treatment, and a significant number of patients initially treated medically eventually had surgery. For example, 38 percent of those with triple-vessel disease treated medically eventually required operation. Initial surgical treatment was recommended for patients with triple-vessel disease and depressed ventricular function regardless of symptoms. On the basis of these data, patients with minimal symptoms and good ventricular function can be safely treated medically, as long as surgery is recommended promptly when symptoms progress.

Other, nonrandomized trials and studies involving the overall CASS registry (of both randomized and nonrandomized patients) have suggested that surgical intervention might improve survival and event-free interval in other patients with triple-vessel disease. Similarly, studies have suggested that patients with double- or triple-vessel disease who become ischemic at a low cardiac workload during stress testing have a poor prognosis with medical therapy and might benefit from early surgical intervention.

A study by Jones and associates (see Angioplasty Versus Bypass Surgery below) evaluated the long-term benefits of bypass surgery and angioplasty versus medical therapy in 9,263 patients with documented coronary artery disease treated between 1984 and 1990. Treatment was nonrandomized, with 2,449 patients receiving medical therapy, 3,890 patients receiving bypass surgery, and 2,924 patients receiving angioplasty. Both interventional treatment groups had better long-term survival than the medical therapy group for all levels of disease severity.

While most patients with mild angina do well with initial medical therapy, coronary artery bypass results in improved survival in patients with more severe symptoms of ischemia, in patients with left main disease, and in patients with triple-vessel disease with depressed left ventricular function or a poor response to stress testing. Bypass operation is recommended in these patients. Some patients with triple-vessel disease and good ventricular function and some patients with double-vessel disease and a tight (>95 percent) proximal anterior descending artery stenosis do better with surgical therapy, especially when ischemia is easily inducible at a low workload.

Unstable Angina

“Acute coronary insufficiency” exists when angina is persistent or rapidly progressive despite optimal medical therapy. This occurs from an acute physiologic state in which the blood flow to a segment of myocardium is seriously jeopardized, but necrosis has not yet occurred. It probably represents instability of a coronary artery plaque, with elevation of the plaque, local thrombus, and spasm, resulting in a sudden decrease in regional blood flow. All agree that a change in angina pattern or unstable angina is a medical emergency. The patient should be promptly hospitalized in a coronary care unit and managed as outlined in the section Medical Therapy. If myocardial oxygen demand is lowered and collateral blood flow increased adequately to compensate for the ischemia, angina subsides and recovery is prompt. These patients should receive aspirin and heparin in addition to standard medical therapy. Cardiac catheterization should be performed promptly. Most patients with unstable angina require revascularization with angioplasty or coronary bypass, with the form of therapy based on the anatomy. Untreated patients with unstable angina have a high incidence of myocardial infarction or death—hence the term “preinfarction angina.”

Operation After Acute Infarction or for Postinfarction Angina

The treatment of myocardial infarction was described in the section Medical Therapy. The indications for surgical therapy after acute infarction include a subendocardial infarction with multivessel coronary disease and postinfarction angina, hemodynamic instability, or heart failure after acute infarction with associated left main disease or multivessel disease unsuitable for angioplasty. Cardiogenic shock after myocardial infarction is an indication for emergency cardiac catheterization and angioplasty or coronary bypass, because complete revascularization of the noninfarcted zones of the heart is necessary for survival. Other indications for operative therapy after myocardial infarction include the mechanical complications of myocardial infarction, such as ventricular septal defect, ruptured papillary muscle with acute mitral insufficiency, and myocardial rupture. Most stable patients with a myocardial infarction undergo a stress test before discharge or at 6 weeks. Cardiac catheterization is done if the stress test is positive. Coronary bypass may be recommended if left main or triple-vessel disease is present and in some cases when proximal anterior artery descending lesions are found and angioplasty is not advisable.

Congestive Heart Failure as a Contraindication

The only absolute contraindication to coronary bypass operation is chronic congestive failure and ischemic cardiomyopathy with no signs of angina or reversible ischemia on PET or thallium scanning. In this situation the majority of the left ventricular muscle is already necrotic, so cardiac transplantation is the only therapy likely to help. In contrast, intermittent ischemia-induced congestive failure, manifested by intermittent episodes of pulmonary edema with angina or ECG changes, is not a contraindication to operation but actually a strong indication for immediate operation. This indicates a serious degree of myocardial ischemia that can easily progress to an irreversible stage or death. The intermittent episodes probably evolve from an acute ischemic episode that elevates end-diastolic pressure sufficiently to produce pulmonary edema.

While a severely depressed ejection fraction of 0.20 to 0.25 or less results in an increased operative risk, it is not a contraindication to surgery. This erroneous concept probably arose from earlier experiences with ineffective myocardial preservation that produced some degree of infarction during operation. Jones and associates reported experiences with 188 patients with an ejection fraction below 0.35, about 24 percent of whom had ejection fraction lower than 0.20. The operative mortality rate was only 2.1 percent. Pigott and Kouchoukos reported results for 192 patients with an ejection fraction less than .35, of whom 77 were operated on and 115 were treated medically. Seven-year actuarial survival was 63 percent in the surgical group versus 34 percent in the medical group. Recurrent infarction developed in 19 percent of the medical group, compared to 7 percent of the surgical group.

Similar results were found in an NYU evaluation of 157 patients with ejection fractions of less than .30 who underwent coronary bypass grafting. Angina was present in 55 percent, but severe congestive heart failure was the predominant symptom in 24 percent. The overall operative mortality rate was 7 percent, and the cumulative 4-year survival rate was 67 percent. While congestive heart failure and pulmonary hypertension had a negative impact on long-term survival, 4-year survival was approximately 50 percent in patients with severe congestive heart failure and pulmonary hypertension.

Angioplasty Versus Bypass Surgery

Transluminal coronary angioplasty, developed by Gruentsig in 1978, significantly changed the treatment of patients with coronary artery disease. Over 400,000 angioplasties are performed each year in the United States. Angioplasty is used primarily for patients with single- or double- vessel involvement and discrete coronary stenoses, although the indications for angioplasty (catheter-based therapy) have continually expanded as the technology has advanced. Directional atherectomy and coronary stenting have improved the results over those achievable with traditional balloon angioplasty, lowering the restenosis rate to approximately 20 to 30 percent. Angioplasty has become a widely used alternative therapy for many patients with coronary artery disease, including many patients with multivessel disease.

Several large studies have compared results after bypass surgery and angioplasty in an attempt to outline the most appropriate therapy for various subgroups of patients. Two significant studies and a metaanalysis of three additional large randomized trials are summarized here:

(1) The BARI Trial (Bypass Angioplasty Revascularization Investigation). This NIH-sponsored trial randomized patients with multivessel coronary disease to either coronary artery bypass graft (CABG, N = 914) or percutaneous transluminal coronary angioplasty (PTCA, N = 915). The patients all had multivessel disease that was suitable for angioplasty in the opinion of the invasive cardiology team, excluding patients with diffuse multivessel disease who were thought to be unsuitable. The average follow-up time was 5.4 years. In-hospital event rates for CABG and PTCA, respectively, were 1.3 and 1.1 percent for mortality, 4.6 and 2.1 percent for myocardial infarction, and 0.8 and 0.2 percent for stroke. The 5-year survival rate was 89.3 percent for CABG and 86.3 percent for PTCA. Within 5 years, 8 percent of those assigned CABG required additional revascularization procedures, 54 percent of those assigned to PTCA required additional revascularization, and 31 percent required surgery. For diabetic patients, 5-year survival was 80.6 percent after CABG and 65.5 percent after PTCA ($p = .003$).

(2) Duke University Nonrandomized Trial. This report by Jones and associates was a long-term survival study involving 9,263 patients treated at Duke University between 1984 and 1990, comparing medical therapy, coronary artery bypass surgery, and angioplasty. CABG and PTCA had better survival rates than medical therapy in all anatomic subgroups, but the survival benefit was stronger for PTCA in patients with single-vessel disease and in patients with double-vessel disease without a tight (<95 percent) left anterior descending artery stenosis. CABG provided better survival rates than medical therapy or PTCA in patients with triple-vessel disease and in those with double-vessel disease and a tight (≥95 percent) left anterior descending artery stenosis.

(3) Metaanalysis of Randomized Trials. Sim and associates reported a metaanalysis based on 3,077 patients from randomized trials other than BARI that compared results of PTCA with CABG. The initial procedure risk was found to be 15 percent higher after CABG than after PTCA, and the periprocedural risk of myocardial infarction was 38 percent higher after CABG than after PTCA. For patients who survived the initial procedure the subsequent risks of both cardiac death and nonfatal myocardial infarction were only two-thirds as high (odds ratio = 0.67, $p = .02$) for CABG patients as for PTCA patients.

Summary

The overall results in comparisons of CABG and PTCA for the treatment of patients with coronary artery disease demonstrate that PTCA and CABG are equally safe and both provide excellent relief of angina. Patients undergoing PTCA have less procedure-related pain, a shorter hospital stay, fewer periprocedural infarctions, and equal immediate relief of angina compared with CABG patients. PTCA may offer an early survival advantage over CABG in patients with single-vessel disease and in some patients with double- vessel disease with less than 95 percent stenosis of the left anterior descending artery. Revascularization is less complete with PTCA, and restenosis is a significant problem, leading to a higher need for subsequent procedures. Approximately one-half of the patients undergoing PTCA require a subsequent intervention within 5 years, and approximately one-third need bypass surgery.

CABG provides more complete revascularization and better long-term relief of angina. After CABG less than 10 percent of patients need subsequent revascularization within 5 to 7 years. CABG results in improved survival in diabetic patients, in some patients with triple-vessel disease, and, possibly, in patients with double-vessel disease and a tight (≥95 percent) left anterior descending artery stenosis.

Silent Ischemia

Better therapy is needed to prevent the development of extensive ventricular injury from coronary disease, the so-called bad left ventricle with an ejection fraction of 0.20 to 0.25, or even lower. It is disconcerting that these patients often have had good medical management because their coronary disease was first recognized years earlier, and they have had neither recurrent major infarctions nor severe angina. Despite periodic medical observation and therapy, the disease has silently progressed. Methods other than severity of angina need to be used to identify these patients and to allow bypass operation to be performed earlier. It is likely that these patients have had ongoing asymptomatic silent ischemia.

Periodic stress testing in patients with known coronary disease is indicated. The stress thallium study or 24-h digital ECG ischemic monitoring may be helpful. Patients with significant silent ischemia or a positive stress test should undergo cardiac catheterization and revascularization if a significant amount of myocardium is at risk.

Ischemic Cardiomyopathy

Patients with ischemic cardiomyopathy and intractable congestive heart failure have been treated by cardiac transplantation with 1-year and 5-year survival rates of approximately 90 and 60 percent, respectively. When reversible ischemia is demonstrable in patients with congestive heart failure, standard bypass operation should be used, which results in a similar 5-year survival rate.

Carpentier, Magovern, and others have tried a procedure termed cardiomyoplasty to augment left ventricular function in patients with ischemic cardiomyopathy. This procedure relies on programmed stimulation of the latissimus dorsi muscle for biomechanical cardiac assistance, subsequently wrapping the muscle around the heart and stimulating synchronized muscular contraction to augment the stroke volume of the heart. Cardiomyoplasty has mixed results with little demonstrable long-term benefit.

Operative Technique of Coronary Artery Bypass

The standard coronary artery bypass operation is performed through a median sternotomy incision (Fig. 18-3), using cardiopulmonary bypass, moderate systemic hypothermia (28 to 32°C), and cold cardioplegia for intraoperative myocardial protection (see Myocardial Protection above).

Historically, *vein grafts* were used in the majority of cases (Fig. 18-4). The saphenous vein graft is still used as the conduit for several of the distal bypasses in most patients with multivessel disease. When using vein grafts, the greater saphenous vein is harvested from the leg, and ligatures or hemoclips are applied to the side branches. The optimal diameter for a vein graft is approximately 4.0 mm. Care is taken to avoid significant trauma to the vein, which is gently flushed with a dilute heparin-crystalloid solution, but not overdiluted. If the greater saphenous vein is not available or if the vein quality is poor, the lesser saphenous vein and the cephalic vein are acceptable alternatives. For patients requiring several grafts the incision can extend over the entire length of the leg, depending on the number of grafts needed.

The proximal saphenous–aortic anastomosis connects each vein graft to the ascending aorta. This is done after completion of the distal grafts or before bypass is started, with subsequent serial construction of the distal anastomoses. In some patients, particularly those with significant atherosclerotic disease in the ascending aorta, all distal and proximal anastomoses are done during the single cross-clamp time to avoid placement of the partial occlusion clamp on the diseased aorta.

The number of distal coronary anastomoses usually varies from two to six. With the heart stopped and cold, a small, 5- to 6-mm arteriotomy is performed in the coronary artery, entering a soft area of the vessel beyond the obstruction. An end-to-side saphenous–coronary anastomosis is performed with continuous 7-0 monofilament cardiovascular suture, using a microvascular suturing technique. Optic magnification at 3 to 4 power is routinely used, allowing for a precise anastomosis on the still, arrested heart.

The *left internal mammary artery* (internal thoracic artery) is used as the primary conduit in the vast majority of patients, usually to graft the left anterior descending artery (Fig. 18-5 A). The mammary artery is taken down from the internal surface of the left anterior chest wall as an in situ pedicle graft. This is a living conduit that remains connected proximally to the native subclavian artery. The mammary artery remains metabolically active and enlarges with time in response to demand and runoff. The mammary artery is resistant to intrinsic atherosclerotic disease and rarely becomes stenotic secondary to atherosclerosis.

Because of these favorable characteristics expanded use of the mammary artery as a conduit has been advocated in recent years, and bilateral mammary grafting has been employed with increasing frequency in nondiabetic patients. After attaching the distal vein grafts, the mammary bypass grafts are performed end-to-side with continuous 7-0 or 8-0 suture. Occasionally, sequential mammary bypass grafts are performed, constructing side-to-side anastomoses to multiple vessels.

Since Grondin reported in 1984 the alarming deterioration of vein grafts between 5 and 10 years after operation but showed that the internal mammary artery grafts remaining patent without late deterioration in 90 to 95 percent of patients, a wider application of left internal mammary grafting has become the standard of care. Certain facts regarding internal mammary grafting were established: (1) The 10-year patency rate is reproducibly 95 to 98 percent (Fig. 18-5 B) with no signs of deterioration or late stenosis after the first few postoperative months. The internal mammary artery seems to be relatively immune to atherosclerosis. (2) A patent mammary artery can enlarge substantially over a number of years, perhaps responding to a decrease in peripheral resistance in the coronary vascular bed as atherosclerosis occludes adjacent vessels. This striking ability to enlarge with time indicates the possibilities with multiple anastomoses constructed from a single mammary artery. (3) Data indicate that bilateral or sequential mammary grafts can be performed in good-risk patients without significant morbidity. In several significant reports, Lytle and associates described experiences with bilateral mammary grafting in 500 patients with little increase in morbidity.

Loop and associates compared longevity in patients who received one mammary graft with those in whom only vein grafts were used. The series included 2,306 internal mammary grafts and 3,625 vein grafts. There was a statistically significant difference in survival at 10 years, gradually becoming apparent between 5 and 10 years after operation. Results suggest that use of a single mammary bypass can lower the incidence of recurrent symptoms at 10 to 15 years, and these results might be improved even more by using double mammary grafts in appropriately selected patients.

Available data concerning “free” mammary artery grafts, in which the artery is divided and reimplanted into the aorta, identical to the method used with a saphenous vein graft, suggest reasonably good results over 10 years. Patency rates for the free mammary graft of approximately 70 to 80 percent at 10 years is slightly better than vein graft patency, but not as good as the patency of in situ mammary pedicle grafts. Free mammary grafts are used when the mammary artery is not of adequate length to reach its destination as an in situ graft.

Alternative Conduits

Encouraged by the success of the internal mammary graft, surgeons have explored the use of other arterial conduits in an attempt to achieve better long-term patency rates than those obtainable with saphenous vein grafts. The most widely used has been the *radial artery graft*. Use of the radial artery was revived by Acar and Carpentier in 1992 when late results showed excellent 15-year patency in several patients. An Allen test and Doppler studies are performed routinely beforehand to assure blood flow to the hand, with the radial artery harvested as a free graft. Because the radial artery has a tendency to spasm, intravenous calcium-channel blockers are used intraoperatively. Reports by Acar, Chen, and others have demonstrated early patency rates of 93 to 96 percent.

The *right gastroepiploic artery* has been used as an in situ pedicle graft and as a free graft. Similar to the radial artery, the gastroepiploic artery has a tendency to spasm, and the early patency rate has been reported to be approximately 95 percent. The *inferior epigastric artery* also has been used as a free graft with similar patency rates. The inferior epigastric artery is only 10 to 12 cm in length and 1.5 to 2.0 mm in diameter and therefore is suitable for use only into a diagonal or high marginal branch.

Cardiac surgery has expanded the use of arterial conduits in an attempt to improve the long-term patency of bypass grafts. While this concept is appealing, late angiographic results with alternative arterial conduits are necessary before it can be accurately determined whether this approach results in better late patency rates than vein grafts.

Coronary endarterectomy, as advocated by Johnson and associates, is possible as an adjunct to bypass grafting in patients with diffuse coronary disease but rarely is necessary, because segmental areas of the coronary arteries usually are adequate for bypass grafting. Most patients with diffuse disease are better served with multiple grafts than with endarterectomy, even if only a part of the vessel has runoff, because endarterectomy has an increased risk of late occlusion from progressive fibrosis. In selected patients a long endarterectomy is performed, which allows subsequent grafting of an otherwise nonbypassable vessel.

After discontinuing cardiopulmonary bypass, flow rates are measured in the grafts with a flow probe to assess graft patency, usually finding a mean flow rate of 50 to 150 mL/min if the vessel has an adequate runoff. The flow rate may be used for comparison later if the patient's clinical status changes.

Postoperative Management

Care initially is given in the recovery room or intensive care unit, with the patient progressing to a monitored intermediate care unit and then to a standard room. Most patients without complications are discharged from the hospital in 5 to 7 days. Activity is increased gradually over 6 to 8 weeks, usually with a progressive walking program over the first 2 to 3 weeks at home. Older patients, however, often need a more formalized cardiac rehabilitation program. Full activity is resumed after 6 to 8 weeks. If vigorous exercise is planned, a stress test is recommended. Long-term management should include antiplatelet therapy with aspirin and aggressive risk factor management. If necessary, cholesterol-lowering agents should be used to achieve the targeted serum cholesterol levels. Hypertension and diabetes are treated, and smoking should be avoided.

Early Results

Operative Mortality

The operative mortality rate for coronary artery bypass is 1 to 3 percent, with the risk varying with the number of risk factors present. The Society of Thoracic Surgeons has a National Cardiac Surgery Database to determine nationwide results and establish preoperative risk variables. These include female gender, age over 65 years, diabetes, renal failure, systemic or pulmonary hypertension, peripheral vascular disease or prior stroke, cardiomegaly, chronic obstructive pulmonary disease, immunosuppressive therapy, prior cardiac surgery, prior noncardiac surgery, recent thrombolysis or percutaneous transluminal coronary angioplasty, prior myocardial infarction, unstable angina, arrhythmias, cardiogenic shock, preoperative inotropes, preoperative intraaortic balloon pump, preoperative heart failure or NYHA functional class IV status, emergency surgery, triple-vessel or left main disease, concomitant valvular heart disease, low ejection fraction (< 0.50), and an elevated left ventricular end-diastolic pressure (Table 18-2). Numerous attempts have been made to develop equations based on statistical methods to predict the “expected risk” for patients or groups of patients according to the number of risk factors present. These approaches have had some success but have not been uniformly reliable.

Operative Morbidity

The most frequent perioperative complications after coronary bypass surgery are myocardial infarction, bleeding, and stroke, which occur in approximately 2 to 4 percent of patients. Other complications include sternal infection, tamponade, aortic dissection, reoperation for graft occlusion, cardiac arrhythmia, cardiac arrest, renal failure, gastrointestinal complications, respiratory failure, and multiorgan failure.

Stroke is a serious complication that occurs with an overall frequency of 2 percent, but with an incidence as high as 10 percent in patients in their seventies and eighties. The most common cause of stroke probably is embolization of atherosclerotic debris from the ascending aorta or aortic arch, although carotid artery disease and emboli from the heart are other potential causes.

Asymptomatic carotid disease is seldom a significant cause of perioperative stroke. Carotid artery disease is surgically treated primarily in the patients with acute ischemic symptoms. If the patient has cerebral symptoms from carotid disease and significant coronary symptoms, the procedures may be done concomitantly. If the coronary symptoms are mild and carotid disease severe, carotid endarterectomy is done first with awake anesthesia and good monitoring. Unless carotid symptoms are present, there are sparse data showing a benefit from performing carotid endarterectomy concurrently with coronary bypass. Exceptions may be patients with unilateral total carotid occlusion and subtotal stenosis of the contralateral carotid and some patients with stenoses with more than 95 percent occlusion in the dominant hemisphere. These patients may benefit from a combined procedure despite the absence of cerebral symptoms.

Considerable attention has been focused on previously unrecognized atherosclerotic disease in the aorta as a source of emboli and cause of perioperative stroke, especially in older patients. Use of intraoperative echocardiography has been very helpful in this regard because this technique can readily detect intraluminal atheroma that are removed or avoided at the time of surgery. In patients over 70 years of age a special aortic arch cannula often is used, with the tip placed beyond the orifice of the left subclavian artery to avoid dislodgment of atherosclerotic material from the transverse arch by the jet of blood emerging from the perfusion cannula. In patients with palpable disease of the aorta or demonstrable intraluminal disease by transesophageal echocardiographic evaluation, the aorta may be surgically explored using the technique of circulatory arrest. The aorta can be readily explored through an appropriately placed 3- to 5- cm aortotomy and loose atherosclerotic debris removed. The aortic cross-clamp then can be placed in an appropriately selected site that is free of debris. Other maneuvers to avoid embolization include placement of the aortic cannula into a nondiseased area on the aorta using intraoperative palpation and transesophageal echocardiography for guidance. Partial-occlusion clamps should not be placed on diseased areas; in some cases all proximal and distal anastomoses are done while the cross-clamp is in place to avoid undue manipulation of the aorta. In rare patients with extensive disease, the aorta is replaced or the entire operation is performed during circulatory arrest. Awareness of the risks inherent in the diseased aorta allows the use of various techniques that lower the risk of perioperative stroke.

Late Results

Relief of Angina

Relief of angina is the most dramatic aspect of coronary bypass surgery. Angina is completely relieved or markedly decreased in over 90 to 95 percent of patients. Persistent angina almost always indicates that a graft has become occluded or a significantly diseased artery was not bypassed.

Early recurrent angina is rare after bypass surgery; reintervention is required in less than 10 percent of patients within 5 years. Recurrent angina occurs more frequently between 5 and 15 years, paralleling the development of late graft occlusion. If significant angina recurs, angiography should be done promptly. In the majority of patients, recurrent angina is the result of one of two causes—stenosis or occlusion of a bypass graft, or progression of disease in the native circulation. Either situation can be treated with repeat bypass grafting or angioplasty. Close control of risk factors postoperatively minimizes the risk of recurrent angina, as does the use of an internal mammary graft.

Graft Patency

With a proper operative technique, combined with the preoperative use of antiplatelet therapy, early vein graft patency rates exceed 95 percent. Early graft patency is determined mainly by operative technique and by the size of the runoff vessel, with small (< 1 mm) vessels occluding more frequently. In the first 5 years after operation, the vein graft patency remains relatively stable, decreasing 1 to 2 percent per year because of intimal hyperplasia in the vein, with a 5-year patency rate of approximately 80 to 85 percent. Intimal hyperplasia can be worsened by trauma to the vein or overdistention at the time of operation, a process that might be related to liberation of growth factors locally or to an enhanced genetic expression of mitogenic signals.

In the period from 5 to 15 years after operation there is a significant increase in atherosclerotic disease in the vein grafts, resulting in a 10-year vein graft patency rate of approximately 60 percent, with significant atherosclerosis occurring in many of the remaining patent grafts. Late graft patency is improved with long-term antiplatelet therapy (aspirin) and by controlling cardiac risk factors. Cessation of smoking and control of cholesterol are particularly important, because late graft occlusion is five to seven times higher in smokers and patients with persistent hypercholesterolemia.

In the past few years the use of the internal mammary artery has increased markedly, including the more frequent use of bilateral and sequential mammary grafts. The 10-year patency rate is over 95 percent for a left internal mammary artery graft and 80 to 90 percent for a right internal mammary artery graft. Failures are primarily technical, because the mammary artery develops little late graft atherosclerosis. The impact of alternative arterial conduits on late graft patency remains to be determined, but early results suggest that arterial grafts may remain patent longer than vein grafts.

Progression of Atherosclerosis

Late angiographic studies have reported a serious progression of atherosclerosis in the native coronary arteries in a certain proportion of patients, ranging from 5 to 10 percent per year and ultimately affecting over one-half of the patients. The need to control cardiac risk factors cannot be overemphasized. Diabetes must be tightly controlled, and the discontinuation of smoking is essential. Weight loss is strongly encouraged in obese patients. Moderate exercise is recommended as long as revascularization is complete and the cardiac function is adequate.

Large studies have demonstrated that reduction in total plasma cholesterol and low-density lipoprotein (LDL) cholesterol levels results in a significant reduction in the progression of atherosclerosis, demonstrating a 2 percent reduction in coronary risk for each 1 percent reduction in plasma cholesterol level. Postoperatively patients are placed on a low-fat, low-cholesterol diet, and ingestion of meat or foods high in saturated fats is minimized. Attention is given to total cholesterol and to the ratio of high-density lipoprotein (HDL) to LDL cholesterol levels. The total cholesterol level should be lower than 200 mg/dL or the LDL cholesterol less than 100 mg/dL; lipid-lowering drugs should be used as necessary to achieve these goals.

Improvement in Ventricular Function

With earlier techniques of cardiac surgery, improvement in impaired ventricular function could not be demonstrated after surgery, probably because of inadequate myocardial protection. With the today's techniques of myocardial preservation the ejection fraction often improves in response to exercise after surgery. Similarly, the regional wall motion frequently improves in ischemic zones. Patients undergoing coronary bypass have a markedly improved functional response to exercise compared with patients treated medically. This improvement lasts 10 years and longer in patients receiving mammary grafts.

Arrhythmias and Sudden Death

Arrhythmias usually are little improved in patients with significant preoperative myocardial scarring. In contrast, in patients with ongoing ischemia and good ventricular function arrhythmias may be significantly improved, because ischemia was the basic cause. Effective medical treatment of arrhythmias is significantly improved with widespread use of beta-adrenergic blockers and other drugs such as amiodarone. The more frequent use of 24-h Holter monitoring and electrophysiology studies has improved the effectiveness of antiarrhythmic therapy. Patients with recurrent ventricular tachycardia have been effectively treated with electrical ablation. In refractory cases the implantable defibrillator may be required. Late sudden death primarily from arrhythmias remains a possibility in any patient with coronary artery disease.

The influence of medical and surgical treatment on “sudden death” was reported from results with over 13,000 patients in the CASS study. Over a period of 4 to 5 years sudden death occurred in 452 patients (3.4 percent of the group). This occurred in 6 percent of the medically treated patients but only 2 percent of the surgically treated patients.

Longevity

The influence of coronary bypass on longevity depends on the anatomic severity of the coronary lesions, the degree of reversible ischemia, the underlying cardiac function, and the degree of revascularization. Studies have demonstrated improved survival with surgery compared with medical therapy in patients with double-vessel disease and a tight proximal left anterior descending artery stenosis, in patients with triple-vessel disease and easily inducible ischemia or depressed left ventricular function, and in patients with left main disease. Coronary bypass surgery has had a significant impact on longevity in symptomatic patients and in the anatomic groups listed.

Survival for the first 5 to 7 years after bypass surgery is similar to age- and-sex-matched controls in patients with good left ventricular function. Loop and associates noted 5-year survival rates among different groups of 90 to 93 percent. Kirklin and associates reported a 5-year survival rate after triple bypass of 89 percent. A 10-year survival rate of 75 to 88 percent in several thousand patients was reported by Cosgrove and associates. The 10-year survival rate varied between 84 percent with a normal ventricle and 54 percent with severe cardiac impairment. Survival beyond 10 years may be significantly improved by use of the internal mammary artery, with a 10-year survival rate of nearly 90 percent in those receiving mammary grafts. Late death, like recurrent angina, usually is found to result from stenosis of a previously inserted bypass graft or from progressive disease in an ungrafted vessel.

New Developments

Minimally Invasive Coronary Bypass Surgery

One of the most significant developments in cardiac surgery in the past 20 years has been the introduction of minimally invasive direct coronary artery bypass (MIDCAB). This procedure also has been referred to as the limited anterior small thoracotomy (LAST) bypass operation and “keyhole” bypass surgery. An expanded approach to minimally invasive cardiac surgery has been developed with the use of endocardiopulmonary bypass and balloon aortic occlusion with cardioplegic arrest. This technique has expanded the applicability of minimally invasive bypass surgery, allowing multivessel bypass procedures and improved accuracy.

The initial experience with minimally invasive beating-heart techniques was in Europe, introduced by Calafiore. Initial results were encouraging and the method quickly spread to the United States, with significant work from Subramanian, Fonger, Mack, Magovern, and others, each exploring variations of the beating-heart technique.

MIDCAB Technique

The operative approach uses a small (5 to 7 cm) left anterior thoracotomy incision. The skin incision is beneath and medial to the left breast, and the fourth costal cartilage often is excised for direct exposure to the left internal mammary artery and the anterior surface of the heart. The left internal mammary artery is harvested by direct vision or with a thoracoscope for visual assistance. After the internal mammary artery is harvested the pericardium is opened. The heart rate is slowed with intravenous beta-adrenergic blockers or intermittent asystole is induced with a bolus of intravenous adenosine. Mechanical stabilizers have been developed to decrease the movement of the coronary artery during the anastomosis. Similarly, the coronary artery is temporarily snared or occluded to provide a relatively dry field for the anastomosis.

The worldwide experience with minimally invasive beating-heart bypass procedures exceeds 1,000 patients, with good early clinical results. The early event-free survival rate approaches 90 percent, with an operative mortality rate of 2 percent or less. The advantages of the technique include less pain and a shorter recovery time than with conventional surgery. This approach has numerous potential benefits to the patient.

A significant disadvantage of the beating-heart technique is the need to perform an anastomosis on the moving heart, which results in decreased anastomotic accuracy and a lower graft patency rate than in operations performed on the arrested heart. Early anastomotic failure rates with the beating-heart approach have been as low as 7 percent in experienced hands to a high range of 15 to 20 percent in several series. The overall value of the minimally invasive nature of the procedure may be partially offset by these anastomotic inaccuracies. The beating-heart approach also is only applicable to single-vessel disease, which accounts for only 4 percent of the patients requiring coronary bypass surgery.

Port-Access Technique

An exciting development proposed by Stevens and coworkers is the endovascular cardiopulmonary bypass and cardioplegia delivery system that introduced the concept of minimally invasive cardiac surgery using peripheral cardiopulmonary bypass and endaortic occlusion (Fig. 18-6) for cardioplegic arrest. With newly developed technology for endovascular aortic occlusion and cardioplegia delivery the heart can be stopped, decompressed, and protected without opening the chest. This allows improved anastomotic accuracy and expands the applicability of minimally invasive bypass surgery.

The endovascular cardiopulmonary bypass–cardioplegic arrest system was tested extensively in the experimental laboratory. Reports from Stephens, Schwartz, and associates demonstrated that this method is safe, that it provides effective myocardial protection, and that a high level of anastomotic accuracy could be achieved. These studies provided the basis for expanded clinical applicability of the minimally invasive coronary bypass.

The initial experiences with the port-access system for coronary artery bypass grafting have been positive, and clinical use of the technique has spread rapidly over a relatively short period. Within 4 months of the technique's introduction more than 350 port-access cardiac procedures had been performed, with encouraging results. The technique has been used in patients with multivessel disease, patients with mitral valve disease, and patients requiring combined valve–coronary artery surgery.

The chest incision and mammary artery takedown are identical to those of the beating-heart approach. The mammary artery may be harvested by direct vision or with the aid of a thoracoscope. If multivessel bypass is necessary the radial artery or saphenous vein is harvested, and a end-to- side Y graft is constructed off of the internal mammary artery. This provides inflow for subsequent grafting of the lateral or posterior walls of the heart when the heart is arrested. Peripheral cardiopulmonary bypass is established through the femoral vessels, and the balloon endovascular aortic occluder is positioned and inflated in the ascending aorta. The heart is arrested with cardioplegic solution, which may be delivered into the aortic root (antegrade) or through a specially designed percutaneous coronary sinus catheter (retrograde). Once the heart is stopped and protected, precise distal anastomoses are possible, with the surgeon readily visualizing most areas of the heart for multiple grafts. This approach maintains many of the fundamental principles of conventional coronary bypass surgery, including myocardial protection and anastomotic accuracy, while offering the advantages of minimally invasive surgery.

Transmyocardial Laser Revascularization (TMLR)

Experimental investigations of laser revascularization of the myocardium have been under way since the late 1980s, but clinical feasibility was demonstrated only recently. The technique uses a high-powered carbon dioxide laser to drill multiple holes (30 to 40 per procedure) through the myocardium into the ventricular cavity. The procedure is performed on the beating heart, and the laser pulses are gated to the R-wave on the ECG. The TMLR procedure is used primarily for patients with refractory angina or hibernating myocardium who have documented reversible ischemia but who are unsuitable candidates for standard coronary artery bypass graft, angioplasty, or heart transplantation.

The results of a large multicenter trial with TMLR were reported by Horvath and associates, who analyzed 200 patients from eight centers. Eighty percent of the patients had undergone a prior coronary artery bypass graft operation, and 36 percent had undergone prior angioplasty. The operative mortality rate was 8 percent. Results at 1 year demonstrated a significant improvement in mean angina class from 3.8 to 1.3 and an improvement in the level of perfusion defect measured by nuclear perfusion scan from 5.1 to 2.8. One center demonstrated a significant improvement in subendocardial flow using the PET scan. Hospital admissions for angina decreased from 2.5 to 0.3 per year.

The mechanism of benefit of TMLR remains uncertain. Theoretically, enhanced subendocardial flow comes directly from the ventricle through the laser channel, but evidence on the long-term patency of the subendocardial channels remains equivocal. The TMLR technique may result in the opening of microvascular coronary-coronary collaterals or enhance coronary angiogenesis, resulting in improved blood flow to the ischemic area of myocardium.

Gene Therapy

Significant progress has been made in understanding the cellular and molecular mechanisms of cardiovascular disease, leading to new forms of molecular-genetic therapy. Much of this work has focused on vascular mitogens, on intracellular signaling pathways, and on the underlying genetic functions that control the cellular response. Since the early 1990s new technology has allowed effective gene transfer, opening up a previously impossible field of biomolecular gene therapy.

In the area of coronary artery disease, several gene therapy protocols are now under active investigation. These include: (1) genetic engineering of vein grafts that are resistant to atherosclerosis or restenosis; (2) gene therapy for coronary arteries to prevent restenosis after angioplasty or surgical endarterectomy; (3) genetic manipulation of the coronary and peripheral vascular system in an attempt to prevent lipid accumulation and progression of atherosclerosis; and (4) the biomolecular delivery of growth factors or genetic material into the coronary system in an attempt to enhance coronary collateral growth through therapeutic angiogenesis.

LEFT VENTRICULAR ANEURYSM

Historical Data

Safe excision of a ventricular aneurysm was not possible on a routine basis until the development of a pump oxygenator, although Bailey and colleagues reported successful repair of five cases between 1954 and 1958. Cooley, in 1958, presented one of the first reports of successful excision of an aneurysm with cardiopulmonary bypass. In the next few years excision became a standard procedure in most cardiac clinics. The procedure is common in most cardiac surgical centers, and numerous series of more than 100 aneurysm repairs have been reported. Cooley and colleagues at the Texas Heart Institute reported more than 4,000 ventricular aneurysm repairs over 30 years.

Etiology, Pathology, and Pathogenesis

A left ventricular aneurysm develops over a period of 4 to 8 weeks or longer in 10 to 15 percent of patients who have had a myocardial infarction. It results when a severe transmural infarction destroys virtually all muscular fibers in the area of the infarction and they are replaced by fibrous tissue. It probably does not occur more frequently after transmural infarctions because collateral circulation often is sufficient to maintain the viability of a variable number of muscle fibers in the zone of the infarct.

The classic aneurysm is an avascular thin scar, 4 to 6 mm thick, that bulges outward when the remaining left ventricular muscle contracts in systole (Fig. 18-7 A); hence the term *paradoxical contraction*, more commonly termed *dyskinesis*, is applied rather than *hypokinesis* (impaired contractility) or *akinesis* (absence of contractility). Mural thrombi are found attached to the ventricular surface of the scar in over one-half of patients, but arterial emboli are rare. The aneurysm usually enlarges to a moderate degree and then stabilizes; progressive enlargement and rupture, which usually occurs with atherosclerotic aortic aneurysms, is rare. Spotty calcification eventually develops in the aneurysmal wall in chronic cases. Over 80 percent of aneurysms are in the anteroapical portion of the left ventricle, resulting from occlusion of the left anterior descending artery. This usually involves significant aneurysmal dilatation of the upper part of the ventricular septum. Posterior ventricular aneurysms are less common (15 to 20 percent), and lateral aneurysms are rare. Posterior aneurysms can distort the mitral valve, resulting in concomitant mitral insufficiency.

A “false aneurysm” is a rare finding that differs from a true ventricular aneurysm. A false aneurysm is a contained rupture of the heart resulting from a hematoma that is formed after a rupture from a myocardial infarction has been temporarily supported by adjacent fibrous tissue. Excision should be done promptly because false aneurysms soon expand and rupture.

In 30 to 40 percent of patients with anteroapical aneurysms coronary disease is limited to the anterior descending artery, which is completely occluded or severely stenosed. Multivessel disease is present in the other 60 to 70 percent of patients. Small aneurysms (less than 5 cm diameter) have negligible physiologic significance, except, possibly, as a site of arrhythmias. Larger aneurysms decrease ventricular function by dissipating the energy of ventricular contraction with the ineffective paradoxical expansion of the wall of the aneurysm during systole. With larger aneurysms the left ventricular volume is increased and left ventricular hypertrophy develops. The left ventricular end-diastolic pressure often is markedly elevated. The combination of elevated intracavitary pressure and radius results in a significant elevation in wall tension (by Laplace's law: tension = pressure × radius). The decrease in effective ventricular contraction eventually results in cardiac failure and angina, though angina may be a result of the accompanying coronary disease. Arrhythmias are prominent in 15 to 20 percent of patients. Areas of scar tissue around the edge of the aneurysm frequently serve as foci for the origin of ventricular arrhythmias, and patients with scar tissue often suffer sustained ventricular tachycardia, syncope, or sudden death.

Because the physiologic burden from an aneurysm is related to its size, the degree of improvement after operation to some extent is related to the size of the aneurysm. For this reason it is doubtful that an aneurysm is ever large enough to be truly inoperable because of size. Improvement depends on the amount of functioning myocardium left, and the best results obtain when the other walls of the heart contract normally.

The development of an aneurysm depends on almost total destruction of muscle fibers. A true aneurysm, composed of dyskinetic scar, must be distinguished from a scar that results from an infarction and that may be akinetic but whose wall is composed of varying proportions of fibrous tissue and viable muscle fibers. Excision of akinetic scars has been investigated in detail but has not been determined to be clinically beneficial.

Because of the wide spectrum between a ventricular “scar” and a true “aneurysm,” accurate data to define the course of the disease are almost impossible to obtain. The outcome of a patient with a ventricular aneurysm will be determined by at least four factors: the size of the aneurysm, the residual coronary disease, the function of the remaining viable muscle, and the presence of severe arrhythmias. Five-year survival rates with an untreated aneurysm have ranged from as low as 10 percent to as high as 70 percent in different reports. Brusche and associates reported 5-year survival rates of 70 percent in patients with an akinetic segment, 54 percent in patients with a dyskinetic segment and good residual ventricular function, and 36 percent in patients with a dyskinetic segment and poor ventricular function.

Clinical Manifestations

Dyspnea and angina, alone or in combination, are the two most common symptoms. Arrhythmias are prominent in 15 to 20 percent of patients. Abnormalities on physical examination usually are not diagnostic. The apical impulse may be forceful and diffuse with a “double impulse.”

Diagnostic Studies

The chest x-ray may show a localized enlargement in the anteroapical area of the left ventricle. ECG changes usually show only the signs of the previous infarction. Persistent ST segment elevation in the precordial leads occasionally occurs, but this also can be present with a large infarction without aneurysmal disease. A paradoxical area from the

aneurysm may be demonstrable by echocardiography or gated blood pool scan, and the echocardiogram is useful in detecting intracavitary clot. Most diagnostic information comes from the left ventricular angiogram, outlining an akinetic area bulging paradoxically during systole. Often no clear differentiation can be made between an akinetic scar and a true aneurysm, with the final decision being made at the time of operation. If a discrete scar is found at operation containing few or no muscle fibers, resection is indicated. If a diffuse bulging is present without discrete borders and obviously containing a moderate amount of muscle tissue, resection probably is contraindicated. If symptoms suggestive of sustained ventricular tachycardia are present, electrophysiologic studies should be done. If no known arrhythmias exist, a 24-h Holter study or a signal-averaged ECG should be done to screen for arrhythmias.

Treatment

Operation is indicated for symptomatic aneurysms, and most asymptomatic aneurysms may be simply observed. The operative procedure includes excision of the aneurysm and bypass grafting of the diseased coronary arteries. Grafting of the diseased anterior descending artery, which often supplies principally ventricular scar, is of questionable value, although at one center it is grafted when feasible because of the possibility that the improved blood supply to the ventricular septum might benefit ventricular arrhythmias.

The aneurysm usually is not manipulated until the aorta has been cross-clamped and the heart arrested to prevent dislodgment of mural thrombi. Once the heart has been arrested, the aneurysm is mobilized and freed from the pericardial adhesions. In the classic ventricular aneurysm repair a subtotal excision of the aneurysm is performed, dividing the wall of the aneurysm about 2 cm from its junction with left ventricular muscle, followed by linear closure. The subtotal concept is a crucial one, because the suture line closing the ventriculotomy is inserted through scar rather than through viable muscle surrounding the aneurysm. This precaution also avoids any excessive reduction in size of the ventricular cavity. Many operative deaths from excision of huge aneurysms probably have resulted from excessive excision of the wall of the aneurysm with injury of the surrounding viable ventricular muscle. The classic ventricular aneurysm repair can have a geometrically deforming effect on the remaining left ventricular cavity because of the linear cardiac closure required after aneurysmal excision. Symptoms often improve after standard repair, although an increase in the ejection fraction does not always occur.

A technique of *reconstructive intracavitary aneurysm repair* has been proposed by Jatene and Cooley. Operative repair involves placement of a Dacron patch to obliterate aneurysmal ventricle and septum, reestablishing a Dacron “roof” on the left ventricular cavity (Fig. 18-7 *B, C, D*). This type of repair remodels the ventricular cavity and obliterates the septal component of the aneurysm, neither of which is achievable with classic aneurysm repair. The Dacron patch is sutured circumferentially to the junction of viable and nonviable myocardium and can be combined with subendocardial resection to obliterate conduction of arrhythmias. This technique results in lowering the diastolic volume without deforming the other walls of the heart and provides better improvement in cardiac function.

The wall of the aneurysm usually includes the area of the anterior descending artery with the scar extending into the ventricular septum. Bypass grafting of the anterior descending artery is of uncertain significance, because the muscle supplied by that artery has been infarcted. If significant tributaries to the ventricular septum are patent, the artery should be preserved and a bypass graft attached. Bypass grafts are routinely placed into other areas of viable myocardium if obstructive coronary lesions are present.

The extent of the endocardial scar usually is greater than that of the external aneurysm. Because of the hazard of malignant ventricular arrhythmias, an extensive excision of the subendocardial scar for at least 1 cm around the periphery of the aneurysm is important when arrhythmias are present; this is done easily without significant injury to functioning ventricular muscle. When combined with intracavitary repair and septal obliteration, an encircling excision of subendocardial scar provides reasonably good control of arrhythmias. Electrophysiologic studies have found that trigger zones for arrhythmias usually are located in the scar within 1 to 3 cm of the border of the aneurysm. When severe preoperative ventricular arrhythmias are present, intraoperative electrophysiologic mapping to locate and ablate the irritable foci may be necessary. Otherwise, an implantable defibrillator is placed postoperatively to control the arrhythmias (see Implantable Defibrillators below).

Results

Prognosis is determined principally by the residual ventricular function, which in turn is influenced by the size of the aneurysm and the severity of the coronary disease. If angina was the prominent symptom before operation, 5-year survival rates are 60 to 70 percent or better, but if congestive heart failure was the principal indication for operation, 5-year survival rates are around 50 percent.

Most patients are significantly improved after operation, manifested by relief of angina and improvement in ventricular function. Significant improvement in ventricular function can be demonstrated with hemodynamic studies in some patients, while in others little change is noted. Improvement in ejection fraction may be better after intracavitary repair with reconstructive aneurysmorrhaphy.

Olearchyk and associates described experiences with 244 cases with a 5-year survival rate near 70 percent. Dobell and associates, in a series of 67 patients, found a 5-year survival rate of 84 percent when angina was the prominent symptom and 53 percent when congestive failure was significant.

Akins and associates, in a series of 100 ventricular aneurysms, reported a 2 percent operative mortality rate and a 6-year survival rate of 77 percent. Jatene reported repair in 508 patients from 1977 to 1983 using reconstructive principles for left ventricular aneurysm repair, with an operative mortality rate of 4 percent. Cooley reported repairs in 100 patients using the intracavitary reconstruction technique, demonstrating a significant improvement in left ventricular ejection fraction after this operation, with an operative mortality rate of 4 percent. Sosa and associates reported use of a similar repair in patients with associated sustained ventricular tachycardia, demonstrating an operative mortality rate of 5 percent and freedom from recurrent ventricular tachycardia of 95 percent. Other studies have suggested that aneurysms associated with sustained ventricular tachycardia should be repaired with intraoperative physiologic studies to guide ablation of the arrhythmogenic foci, particularly when sustained ventricular tachycardia is the presenting symptom. Combined aneurysm repair and map-guided ablation can be done with a 5 to 10 percent operative mortality rate.

VALVULAR HEART DISEASE

General Principles

“*When should an operation be performed?*” This basic question must be evaluated periodically during the medical therapy of any patient with valvular disease, because the disease process usually is progressive. Valvular heart disease results in a *pressure load* (valvular stenosis), a *volume load* (valvular insufficiency), or both (mixed stenosis and insufficiency). For example, aortic stenosis increases the afterload of the left ventricle, resulting in left ventricular hypertrophy, while both aortic and mitral insufficiency result in a significant volume overload of the left ventricle with cardiac dilatation. While the heart effectively compensates for these hemodynamic changes for some time, progressive deterioration in cardiac function eventually develops. The goal of the clinician is to decide when progressive deterioration in cardiac function is likely to occur and to intervene before significant irreversible cardiac injury has occurred. A decision for operation often is based on physiologic abnormalities found with diagnostic studies, such as echocardiography, cardiac catheterization, or radionuclide studies, rather than by waiting for severe symptoms to develop. Surgical therapy is recommended at a much earlier stage of the disease today than in earlier years in an attempt to maintain excellent long-term cardiac function after valve replacement.

The importance of this approach is emphasized by the fact that longevity after cardiac valve replacement is strongly influenced by the myocardial function at the time of operation. Patients with early functional disability (NYHA functional class II or early class III) have a 5-year survival rate near 90 percent, but only 60 percent of patients with class IV disability live 5 years. This striking difference is a result of irreversible changes in myocardial function that existed before operation, indicating that operation should not be postponed until symptoms are disabling.

For aortic stenosis, valve replacement is strongly indicated in all symptomatic patients and should be seriously considered in asymptomatic patients if the orifice cross-sectional area has decreased below 1.0 cm². This is especially true if the patient has developed significant left ventricular hypertrophy and is beginning to show signs of systolic dysfunction of the left ventricle. With mitral stenosis the left ventricle is protected from injury, and operation usually is indicated when early symptoms develop or when the left atrium becomes significantly enlarged and atrial fibrillation develops.

With aortic and mitral valve insufficiency, selecting the proper time for operation is more difficult, because it depends on the left ventricular function in response to chronic volume overload. The heart compensates for a significant period of time and the patient remains asymptomatic. The clinician must carefully monitor the patient, looking for signs of decreasing systolic function of the heart or for a rapid enlargement in the size of the heart. The demonstration of a decreased ejection fraction at rest (or a rise in the end-systolic volume by echocardiography) or a fall in the ejection fraction during exercise probably are the best signs that the systolic function of the heart is beginning to deteriorate and that operation should be performed promptly. Cardiac function generally will return to normal if operation is done at this time. Postponing operation until serious enlargement of the left ventricle develops and the systolic function has severely deteriorated is a mistake, because 5-year prognosis after successful operation is greatly decreased because of irreversible ventricular injury.

Even with advanced disease, NYHA class IV disability and pulmonary hypertension, patients with valvular heart disease are rarely inoperable. Valve replacement almost always can be performed with an operative mortality rate of 5 to 10 percent. After the hemodynamic burden on the ventricle has been removed with corrective surgery, the patient is treated with an aggressive medical heart failure regimen (digitalis, diuretics, and afterload reduction), usually with significant improvement. Except in the unusual case of intrinsic myocardial disease with late-stage “cardiomyopathy,” operation should not be denied to patients with advanced valvular heart disease.

Survival after valve repair or replacement is strongly influenced by the left ventricular function beforehand—i.e., the systolic function or ejection fraction—and the clinical functional classification. Other risk factors affecting survival are age, congestive heart failure, pulmonary hypertension, concomitant coronary disease, prior myocardial infarction, concomitant disease in another valve, and prior surgery.

Late complications from valvular surgery include heart failure, endocarditis, thromboembolic and anticoagulant-related complications, prosthetic valve failure, and reoperation.

Surgical Options: Prosthetic Valves, Homografts, Autografts, and Valve Repair

The ideal prosthetic valve does not exist. The two basic types of valves are *mechanical prostheses* and *bioprostheses* (heterografts). Both have advantages and disadvantages, which can be summarized as follows: mechanical prostheses are highly durable but require permanent anticoagulation therapy to minimize the risk of thromboembolic complications; bioprostheses have fewer thromboembolic complications but are more prone to structural failure and to require reoperation.

Mechanical valves (Fig. 18-8) commonly used in the United States include the St. Jude Medical bileaflet prosthesis, the Medtronic-Hall tilting disc valve, the CarboMedics bileaflet prosthesis, and the Omniscience disc valves. The older Starr-Edwards caged ball valve has limited use. The disk prostheses have excellent flow characteristics and durability with a low incidence of late valve-related complications.

With mechanical valves the basic long-term risks are thromboembolism and hemorrhage from anticoagulation therapy with warfarin. With proper anticoagulation, keeping the INR

(international normalized ratio) at 3.0 to 4.0 times normal, the incidence of thromboembolism is low, approximately 1 to 2 percent/patient/year. Antiplatelet therapy with aspirin may additionally lower the risk of thromboembolism, although this may not be necessary if close attention is paid to the INR. Anticoagulant-related hemorrhage resulting from warfarin therapy occurs with a incidence of 1 to 4 percent/patient/year, while combined thromboembolic and anticoagulant-related complications occur in 3 to 6 percent/patient/year. With careful supervision and close attention to INR monitoring of warfarin therapy, significant thromboembolic and anticoagulant complications are becoming progressively less common.

The decision in selecting a mechanical prosthesis depends on the patient's understanding and ability to take anticoagulant therapy safely and indefinitely. Mechanical prostheses are more durable than bioprostheses, with 15-year valve failure rates of approximately 5 percent. Mechanical prostheses usually are preferable in patients with a long life expectancy and in patients who want to minimize the risk of reoperation. The drawbacks of mechanical prostheses include a higher incidence of thromboembolic and anticoagulant-related complications and the need for permanent anticoagulation.

Several types of bioprostheses exist, with similar characteristics (Fig. 18-9 A, B). The glutaraldehyde-preserved stented porcine xenograft valve and the stented bovine pericardial xenograft valve are the two main types. The main advantage of the various bioprostheses is the low incidence thromboembolism, making anticoagulation therapy unnecessary in the majority of patients, but the principal limitation is poor durability because of late valve failure. The late failure rate of bioprostheses is approximately 20 percent at 10 years and over 50 percent at 15 years.

Surgical alternatives to prosthetic valve replacement have been developed in an attempt to utilize the body's natural tissue and lower the incidence of valve-related complications. In the 1960s Ross, in England, and Barrett-Boyes, in New Zealand, described a procedure for aortic valve replacement using antibiotic-preserved aortic homograft valves. Newer methods of homograft preservation that result in significant cellular viability are possible using cryopreservation techniques, widening the availability of homografts and potentially improving the long-term results. Homografts have been used more frequently for aortic and pulmonary valve replacements. Because the homograft valve is natural tissue, the thromboembolic rate is low and long-term anticoagulation therapy is not required. The late durability of homografts remains uncertain. Estimates suggest that the 10- to 15-year durability of aortic homografts may be similar to that of porcine valves, with a freedom from reoperation of approximately 50 percent at 15 years. Some reports suggest better durability when using cryopreserved tissue.

Ross described a potentially more durable, but more complicated, alternative for aortic valve replacement using the patient's own pulmonary valve as an autograft and replacing the pulmonary valve with an aortic or pulmonary homograft (Fig. 18-10). Initially the Ross procedure was not widely accepted, but it has been increasingly used in recent years as operative techniques and myocardial protection have improved and cryopreserved homografts have become more widely available. The Ross procedure has the advantage of placing a natural, autologous pulmonary valve into the aortic position, which should be extremely durable without the need for anticoagulation therapy. The results have been good. Ross reported late follow-up on 339 autograft patients with a rate of freedom from autograft replacement of 85 percent at 20 years and a rate of freedom from all valve-related events of 70 percent at 20 years. Elkins reported similar results with an actuarial freedom from any reoperation of 89 percent at 5 years, 92 percent for the pulmonary autograft, and 95 percent for the homograft in the pulmonary position. Given these excellent results, without the need for long-term anticoagulation, many surgeons recommend the Ross procedure for most younger patients who require aortic valve replacement.

Important alternatives to mitral valve replacement for mitral insufficiency also have also been developed. In the 1960s McGoon, Kay, and Reed each developed separate plication techniques for reconstruction of the mitral valve for mitral insufficiency, but these methods had limited applicability and were used primarily for treatment of isolated chordal rupture or annular dilatation. Renewed interest in surgical reconstruction of the mitral valve resulted from work by Carpentier in the 1970s. Carpentier's system for mitral valve reconstruction was applicable to a wider range of pathologic findings than prior methods, and valve reconstruction has subsequently proved to be highly reproducible. The long-term durability of mitral valve reconstruction has been approximately 90 percent at 10 years for patients with degenerative disease. Valve reconstruction may offer significant advantages over valve replacement in terms of lower thromboembolic and anticoagulant-related complication rates and overall valve-related morbidity and mortality (Fig. 18-11). Mitral valve reconstruction has assumed an important role in the treatment of mitral valve disease, offering an excellent alternative to valve replacement in selected patients.

Recommendations

Before operation the surgeon should give the patient a recommendation on the appropriate prosthesis or procedure. Some patients are capable of making this decision after the various risks and options have been explained, and others are not. The surgeon should attempt to learn from the patient whether there is any strong reason to avoid anticoagulation therapy and how the patient feels about the uncertainty of future reoperation. The surgical recommendation then is based on several factors, including the patient's age, lifestyle, intelligence, reliability, and access to follow-up health care. Socioeconomic factors, the desire to have children in female patients, and associated medical conditions also are considered.

For aortic valve replacement, a mechanical prosthesis or a Ross procedure is recommended for most younger patients, depending on their desire to avoid anticoagulation therapy. An aortic bioprosthesis often is recommended for patients older than 70 to 75 years of age because anticoagulation therapy may be hazardous in the elderly and valve failure is unlikely to occur before death from other causes.

For patients with mitral valve disease, commissurotomy or valve reconstruction may be performed if feasible, as long as a durable result is likely. Reconstructive procedures are not recommended, unless the likelihood of achieving a durable repair is high. Mitral valve replacement is required in 10 to 50 percent of patients, depending on the patient population and the experience of the surgical group. When valve replacement is necessary a mechanical prosthesis is recommended for most patients, and especially for patients in atrial fibrillation, since anticoagulation therapy already is required in this group. Bioprostheses usually are reserved for elderly patients in whom anticoagulation therapy is considered hazardous.

Mitral Valve Disease

Mitral Stenosis Etiology

Mitral stenosis is almost always caused by rheumatic heart disease, although a definite clinical history can be obtained in only 50 percent of the patients. Mixed stenosis and insufficiency also are usually rheumatic in origin, and other causes of mitral stenosis are exceedingly rare. Congenital mitral stenosis can result when a single papillary muscle is present (parachute mitral valve) or when the subvalvular apparatus is fused, producing restricted leaflet motion. Stenosis occasionally occurs from systemic lupus erythematosus, rheumatoid arthritis, mucopolysaccharide disease, or carcinoid.

Pathology

Although the rheumatic inflammatory process is associated with some degree of pancarditis, involving endocardium, myocardium, and pericardium, permanent injury results predominantly from the endocarditis with progressive fibrosis of the cardiac valves. The mitral valve apparently is almost always involved.

Rheumatic valvulitis produces three distinct degrees of pathologic change: fusion of the commissures alone, commissural fusion plus subvalvular shortening of the chordae tendineae, and extensive fixation of the entire valve and subvalvular apparatus, with calcification and scarring of both leaflets and chordae (see Fig. 18-12). The degree of pathology present should be determined preoperatively, because this predicts the suitability of surgical commissurotomy or valve replacement.

Mitral stenosis usually has a prolonged course after the initial rheumatic infection, and symptoms may not appear for 10 to 20 years. The progression to valvular fibrosis and calcification may be related to repeated episodes of rheumatic fever or may be from scarring produced by turbulent blood flow after the initial episode of inflammation. The flow alterations cause repeated trauma to the leaflets, resulting in the deposition of thrombin, platelets, and other blood components.

Mitral stenosis produces a functional narrowing of the valve orifice. This impairment to transvalvular diastolic blood flow results in an end-diastolic pressure gradient across the mitral valve. The most precise determination of the degree of mitral stenosis is the mitral valve area, which can be calculated from the diastolic pressure gradient across the mitral valve and the cardiac output using the formula developed by Gorlin. The normal cross-sectional area of the mitral valve is 4 to 6 cm², varying with body size. Significant hemodynamic changes from mitral stenosis usually do not appear until the cross-sectional area is less than 2 to 2.5 cm², and symptoms progressively increase when the cross-sectional area decreases to less than 1.5 cm². Severe mitral stenosis, producing NYHA functional class III or class IV symptoms, occurs when the mitral valve area is less than 0.8 to 1.0 cm², and an opening of approximately 0.5 cm² is said to be the smallest size compatible with life, producing symptoms at rest.

Pathophysiology

The pathophysiology associated with mitral stenosis results from a chronic elevation in left atrial pressure that produces pulmonary venous congestion and subsequent symptoms. The left ventricular function usually remains normal as the ventricle is protected by the stenotic valve. As the degree of stenosis progresses, the flow of blood into the left ventricle is limited and the cardiac output is reduced, resulting in poor exercise tolerance, fatigue, and muscular wasting. As the left atrium hypertrophies, atrial fibrillation eventually develops, which exacerbates the patient's symptoms and results in an increased likelihood of clot formation and embolization.

Clinical Manifestations

The main symptoms of mitral stenosis are exertional dyspnea and fatigue. Dyspnea occurs when the mean left atrial pressure and pulmonary venous pressure exceed the plasma oncotic pressure, producing interstitial edema and pulmonary congestion. Orthopnea, paroxysmal nocturnal dyspnea, and a cough that worsens when the patient lies down reflect the influence of upright position on pulmonary congestion. Severe pulmonary congestion can produce acute pulmonary edema or hemoptysis. Although hemoptysis often subsides, pulmonary edema may be fatal unless treated promptly. When pulmonary hypertension results in right-sided heart failure, findings such as jugular venous distention, ascites, or ankle edema may be seen.

Atrial fibrillation is present in 30 to 50 percent of the cases, and in some patients arterial embolization is the initial presenting symptom. Atrial thrombi result from stasis and dilatation of the left atrium, with the left atrial appendage being especially susceptible to clot formation. Atrial fibrillation does not need to be present for embolization to occur. Angina is a rare symptom that may result from coronary embolization or from decreased diastolic perfusion of the myocardium.

The characteristic auscultatory findings of mitral stenosis, called the "auscultatory triad," are an increased first heart sound, an opening snap, and an apical diastolic rumble. These findings are pathognomonic. The apical diastolic rumble, produced by blood flowing through the stenotic orifice, may be sharply localized to an area at the apex and usually is heard best when the patient is in the lateral decubitus position. A loud pansystolic murmur transmitted to the axilla usually indicates associated mitral insufficiency. A systolic murmur along the left lower sternal border, loudest near the xiphoid process, commonly occurs with tricuspid insufficiency.

Diagnostic Studies

The ECG may show atrial fibrillation, left atrial enlargement (P mitrale), and right-axis deviation, or it may be normal. The ECG often is an inadequate guide to the severity of the stenosis. In contrast, several abnormalities usually are seen on the chest x-ray. The earliest change is enlargement of the left atrium, typically seen on the posteroanterior film as a

double contour visible behind the right atrial shadow. The overall cardiac size may be normal, but enlargement of the left atrium and the pulmonary artery may obliterate the normal concavity between the aorta and the left ventricle, producing a “straight” left border of the heart. Calcifications of the mitral valve also may be seen. In the lung fields, the typical abnormalities consistent with pulmonary congestion may be present.

Echocardiography is the diagnostic method of choice. Transesophageal echocardiography is a particularly useful test, because it provides enhanced resolution and unobstructed visualization of the mitral valve and the posterior cardiac structures, including the left atrium and the atrial appendage. Echocardiography gives a very accurate measurement of the cross-sectional area of the mitral valve and allows assessment of the degree of leaflet mobility, calcification, and subvalvular fusion, each of which may be important in predicting the feasibility of commissurotomy.

Cardiac catheterization should be performed when the diagnosis is in doubt and in patients in whom concomitant coronary artery disease is likely. Catheterization also is valuable in patients with long-standing disease to assess other factors, such as pulmonary hypertension, the degree of mitral insufficiency, and the presence of associated pathology in another valve.

Operative Indications

Operation is indicated for all symptomatic patients with mitral stenosis and should be considered in patients with minor symptoms whose left atrium is significantly enlarged and in patients who have had recent onset of atrial fibrillation or an embolic event. Exercise studies may be useful in evaluating the need for early operation in minimally symptomatic patients.

In determining the timing of surgery in patients with mitral stenosis, it is important to remember that the disease often progresses insidiously. Patients may be treated medically for years with few symptoms other than a minor limitation of activities and be unaware that progressive destruction of the mitral valve is occurring. Close supervision and periodic echocardiographic evaluation are crucial in choosing the proper time for intervention.

Patients almost never have such advanced disease that operation cannot be done. Even in patients with NYHA functional class IV symptoms with pulmonary hypertension, cardiac cachexia, and ascites, the operative mortality rate is now less than 10 percent, and the 5-year survival rate is over 50 percent. Most patients with class IV symptoms obtain remarkable clinical improvement after operation. Successful operation on the majority of seriously ill patients with mitral stenosis is possible because the stenotic valve has restricted inflow into the ventricular cavity, protecting left ventricular function.

Mitral Insufficiency Etiology

Degenerative disease is the most common cause of mitral insufficiency in the United States, accounting for 50 percent of the patients operated on for it. Other causes include rheumatic fever (20 to 30 percent), ischemic disease (15 to 20 percent), endocarditis, congenital abnormalities, and cardiomyopathy.

Pathology

The four major structural components of the mitral valve are the annulus, leaflets, chordae, and papillary muscles. A defect in any of these components may create mitral insufficiency. A functional classification for mitral insufficiency was proposed by Carpentier, outlining three basic types of functionally diseased valves. Type I insufficiency occurs from annular dilatation or leaflet perforation with normal leaflet motion. Typically, this type is a result of ischemia or endocarditis, although isolated annular dilatation is occasionally seen with rheumatic and degenerative causes. Type II insufficiency occurs secondary to increased leaflet motion with leaflet prolapse. This may be from an elongated or ruptured chorda and is most commonly degenerative or ischemic in origin. Type III insufficiency occurs from restricted leaflet motion, with the involved leaflet not reaching the proper plane of closure during systole; this type is found most commonly in rheumatic patients. Annular dilatation often develops secondary to insufficiency from prolapse or restricted leaflet motion and usually is an additional part of type II or type III insufficiency.

The chordae tendineae are thinned and elongated in patients with degenerative disease. Excessive leaflet tissue usually is present and may be thin, or it may be thickened and whitish in patients with Barlow’s syndrome. The valve usually has a billowing appearance, and the annulus is invariably dilated.

With rheumatic disease the chordae tendineae are thickened and shortened, and leaflet motion is restricted. Varying degrees of commissural fusion, leaflet fibrosis, or calcification may occur, producing mixed stenosis and insufficiency. Asymmetric posterior annular dilatation usually is present.

Pathophysiology

The basic physiologic abnormality in patients with mitral insufficiency is regurgitation of part of the left ventricle stroke volume into the left atrium. This results in decreased forward blood flow and an elevated left atrial pressure, producing pulmonary congestion and volume overload of the left ventricle.

As mitral insufficiency progresses there is a corresponding increase in the size of the left atrium above the normal upper limit of 4 cm in diameter. As this occurs, atrial arrhythmias and eventually atrial fibrillation evolve, particularly after the atrium is dilated to more than 4.5 to 5 cm in diameter. Concurrently, the left ventricle dilates and the stroke volume increases. Eventually, however, this compensatory mechanism fails and the ejection fraction decreases. A decrease in the systolic function of the heart, manifested as a fall in the ejection fraction, is a relatively late finding, because the ventricle is “unloaded” as a result of the valvular insufficiency.

Clinical Manifestations

In patients with acute mitral regurgitation the left atrial pressure rises abruptly and congestive heart failure develops suddenly, while in patients with chronic mitral insufficiency symptoms do not develop until later in the course. As chronic mitral insufficiency progresses, the patient develops exertional dyspnea, orthopnea, easy fatigability, and palpitations. These symptoms arise from the combination of an increased left atrial pressure and a decreased cardiac output. As left ventricular dysfunction progresses, symptoms of pulmonary congestion become more prominent, ultimately producing pulmonary hypertension and right-sided heart failure.

On physical examination the characteristic findings of mitral insufficiency are an apical systolic murmur and a forceful apical impulse. The apical murmur usually is harsh and transmitted to the axilla, although if the left atrial pressure is extremely high, the murmur may not be audible. The severity of the insufficiency may not correlate with the intensity of the murmur.

Diagnostic Studies

The chest x-ray is excellent for assessing cardiac enlargement. Frequently the left atrium and left ventricle are enlarged and the cardiothoracic silhouette is increased.

Echocardiography is considered the most valuable diagnostic test. With echocardiography the severity of insufficiency can be determined accurately and the site of valvular prolapse or restriction identified. An important measurement is the size of the cardiac chambers. The size of the left atrium reflects the chronicity and severity of the insufficiency. The normal left atrium seldom has an internal diameter greater than 4 cm. With severe, chronic mitral insufficiency dilatation of the left atrium to 5 to 6 cm or greater is common. This is important because the propensity for atrial fibrillation is greatly increased when the left atrial size is more than 4.5 to 5 cm. The diastolic dimensions of the left ventricle become enlarged relatively early in patients with mitral insufficiency because of the volume overload inherent in the regurgitant lesion.

Once the echocardiogram demonstrates a decrease in ventricular systolic function, manifested as a rise in the end-systolic dimension of the heart, this is an indication that the left ventricle is no longer able to handle the volume load, and symptoms usually will develop shortly. If there is uncertainty about the physiologic significance of mitral insufficiency, an exercise radionuclide study or a stress-echocardiography study may be used. Normally the ejection fraction rises with exercise, but a fall in ejection fraction with exercise suggests systolic dysfunction in response to stress. If the ejection fraction is diminished at rest, significant systolic dysfunction is already present.

Cardiac catheterization is indicated before operation in patients with advanced mitral insufficiency and cardiac enlargement as well as in patients in whom the diagnosis is uncertain. Coronary angiography is performed in older patients and in patients with a history of angina or myocardial infarction. With catheterization it is important to note the degree of pulmonary hypertension, the left atrial and left ventricular end-diastolic pressures, the ejection fraction, and the presence of associated valvular pathology or coronary artery disease.

Operative Indications

Historically, operation was not recommended for patients with mitral insufficiency until functional disability was present. However, delaying operation until the patient is severely symptomatic and the heart is markedly dilated often results in a certain degree of irreversible ventricular injury. In order to better preserve ventricular function and obtain the best long-term results, a more aggressive approach is recommended today. Operation is recommended *at the first sign of symptoms* and for patients with minimal or no symptoms if there are *signs of diminishing systolic function of the heart*. Operation is not mandatory but should be considered in patients with significant insufficiency when the left atrium is enlarged to more than 4.5 to 5.0 cm and atrial fibrillation develops, particularly if progressive left ventricular enlargement is also occurring.

Recommending operation for minimally symptomatic patients on the basis of hemodynamic abnormalities has been discussed periodically in different reports. Studies of the natural clinical history of mitral insufficiency have found that totally asymptomatic patients with severe mitral insufficiency usually progress to significant disability within 5 years. Similarly, ventricular function appears to be correlated with late results after valve replacement, with the best results obtained in patients with normal ventricular function. Once the ejection fraction has deteriorated, results are significantly worse. These considerations were reviewed in detail by Levine and associates, and by Assey and associates, who proposed parameters of ventricular function to optimize the timing of operation.

Operative Techniques Standard Operative Approach

For mitral valve surgery the traditional operative approach has been through a median sternotomy incision (see Fig. 18-3). This approach has outstanding results providing excellent exposure to the heart and safe access to the great vessels for cannulation, cross-clamping, and administration of cardioplegic solutions. Cardiopulmonary bypass usually is established with cannulation of the ascending aorta and bicaval venous drainage. The aorta is cross-clamped and the heart is arrested with cardioplegic solution given in antegrade fashion into the aortic root or retrogradely into the coronary sinus.

The preferred technique for exposure of the mitral valve is through a left atrial incision, made posterior and parallel to the intraatrial groove. After the left atrial incision is made, self-retaining retractor blades are gently inserted, elevating and rotating the septum superiorly and leftward to expose the mitral valve. In certain patients exposure through the posterior left atrial approach may be not be optimal. This is particularly true in patients with a small left atrium, a deep chest, an aortic prosthesis in place, and in reoperations, in which the tissue is fixed and nonelastic. Alternative incisions for exposing the difficult mitral valve include a transseptal incision, a superior approach through the dome of the atrium, and the biatrial transseptal approach described by Smith.

Open Mitral Commissurotomy

The basic technique for open mitral commissurotomy was first described by Mullin in 1974. If commissural fusion alone is present, excellent results are predictable. In contrast, if the valve leaflets are fibrotic and calcified and the subvalvular apparatus is immobile, restoration of normal mobility usually is impossible, and prosthetic valve replacement usually is required. Shortening and fusion of the underlying chordae tendineae occur in approximately one-third of the patients by the time of surgery. This might explain why surgical commissurotomy may be superior to balloon dilation in many patients. With operation the fused chordae can be surgically divided and mobility restored, but with balloon valvuloplasty this is not feasible.

Once cardiopulmonary bypass has been established and the heart has been arrested, the left atrium is opened and the mitral valve is visualized. Initially the atrial cavity is examined for thrombi, especially within the atrial appendage. If the patient is in atrial fibrillation, the atrial appendage is closed with a continuous suture of 3-0 Prolene to minimize the subsequent risk of emboli.

The mitral valve is assessed by evaluating the leaflet mobility, commissural fusion, and the degree of fibrosis in the subvalvular apparatus. A right- angle clamp often is placed beneath the commissures, gently applying horizontal tension to evaluate commissural fusion. The fused commissure actually is a scar developing at the line of commissural closure of the mitral leaflets. In most patients this can be readily identified as a trench, although if fibrosis of the leaflets is severe, the exact location of the commissure may be difficult to identify. Examining the commissural chordae may provide additional landmarks.

Once the commissure has been accurately identified and the chordae noted, a right-angle clamp is introduced beneath the fused commissure, stretching the adjacent chordae and leaflets, after which the commissure is carefully incised with a No. 11 or No. 15 blade scalpel. The incision is made 2 to 3 mm at a time, serially confirming that the separated margins of the commissural leaflet remain attached to chordae tendineae. The usual commissurotomy curves slightly anteriorly and does not go directly laterally. The incision should stop 1 to 2 mm from the valve annulus where the leaflet tissue becomes thin, indicating the transition from the fused commissure to the normal commissural leaflet of the mitral valve. Once the commissurotomy is completed, the fused papillary muscle often is incised for 5 to 10 mm as necessary to minimize restriction and improve mobility of the attached leaflet.

After separation of the commissure and mobilization of the underlying chordae tendineae and papillary muscle, leaflet mobility is assessed visually. The anterior leaflet is grasped with a forceps and moved throughout the entire range of motion, looking for subvalvular restriction or leaflet rigidity. Movement of the leaflet is assessed by injection of cold saline into the ventricle. In some patients restricted motion may be further improved by dividing secondary chordae or by the selective debridement of calcium. Very thickened chordae can be mobilized by excision of a triangular portion of the fused cords. If extensive debridement and mobilization of the chordal structures is necessary, valve replacement usually is more appropriate. At least 30 percent of patients undergoing commissurotomy require more than simple incision of the commissure to produce an adequate mitral orifice and to restore mobility to the valve.

Mitral Valve Replacement

Mitral valve replacement is necessary when the extent of disease precludes commissurotomy or valve reconstruction (see Fig. 18-12). Valve replacement is most likely in patients with long-standing disease. The techniques of perfusion and the surgical approaches are identical to those described for open mitral commissurotomy.

After the left atrium is opened, the diseased mitral valve is exposed with a self-retaining retractor. If valve replacement is necessary, an incision in the anterior mitral leaflet is made, usually starting at the 12-o'clock position and resecting most of the anterior leaflet and associated chordae. The posterior leaflet and chordal attachments are preserved whenever feasible, although this is not always possible in patients with extensive rheumatic disease, because calcium or extensive valvular thickening is often present.

Although preservation of the subvalvular apparatus and chordae has been demonstrated to be valuable when performing valve replacement for mitral insufficiency, the role of chordal preservation to preserve postoperative left ventricular performance after valve replacement for mitral stenosis remains less certain. Okita compared ventricular functional recovery after mitral valve replacement with and without chordal sparing. Postoperative radionuclide angiography during exercise failed to demonstrate any difference in ejection fraction in the chordal preservation group versus those having chordal resection. Harpole and associates found no decrease in the ejection fraction or in the pressure volume relationship of the left ventricle after mitral valve replacement for mitral stenosis despite the resection of chordae. Posterior chordal preservation during mitral valve replacement does lower the risk of posterior left ventricular free wall rupture, which is a lethal complication of mitral valve replacement.

When the mitral valve is excised, a prosthesis of appropriate size is selected. The valve usually is inserted with 12 to 16 mattress sutures, simple or buttressed with a pledget. The mattress suture technique with pledgeted sutures is recommended because it virtually eliminates perivalvular leakage. The sutures often are inserted from the atrial to the ventricular side when a mechanical prosthesis is being used; it may be desirable to evert the annulus and seat the valve intraannularly (see Fig. 18-12). This technique avoids potential leaflet-annular entrapment.

In patients with friable tissue or a calcified annulus the sutures may be inserted from the ventricular side, through the annulus and sewing ring, seating the valve on top of the annulus. The supraannular technique may be stronger, because it sandwiches the annulus between the pledgets and the sewing ring without everting the annulus. This technique is used frequently for bioprostheses.

Care is taken to insert the sutures precisely into the annular tissue, because excessively deep insertion of sutures may injure critical structures, including the circumflex coronary artery posterolaterally (7 to 8 o'clock), the atrioventricular node anteromedially (1 to 2 o'clock), or the aortic valve anterolaterally (10 to 12 o'clock).

Mitral Valve Reconstruction

The basic techniques of valve reconstruction for mitral insufficiency, described by Carpentier and others, include the use of an annuloplasty ring, quadrangular resection of the posterior leaflet and chordal shortening, chordal transposition, and triangular resection for repair of the anterior leaflet. The remodeling ring annuloplasty concept was based on Carpentier's finding that the deformity resulting in mitral insufficiency was dilatation of the annulus. In association with annular dilatation, other anatomical abnormalities include elongation or rupture of chordae to produce increased leaflet motion, or localized restrictive disease, as described in Carpentier's functional classification of valvular pathology.

The intraoperative assessment of the valvular pathology is the most important step in valve reconstruction. A localized roughened area of atrial endocardium, a "jet lesion," may be present from regurgitant blood striking the endocardium, providing a guide to the location of the most serious insufficiency. The degree of annular enlargement is noted initially. The commissures are then examined, noting whether these are prolapsed, fused, or malformed. The closing plane of the leaflets in the area supported by commissural chordae is determined next, and the anterior and posterior leaflets are then serially examined, noting areas of prolapse or restriction. Abnormalities such as perforation, fibrosis, calcification, or clefts also must be recognized.

Proper evaluation of the degree of leaflet prolapse is critical. The "billowing" mitral valve originally described by Barlow has excessive leaflet tissue, but may remain competent if the chordae are not elongated. In such cases, the rough free edge of the leaflet closes at the proper level even though the midportion of the leaflet may contain excessive tissue. As emphasized by Carpentier, the anterolateral commissural chordae are seldom elongated, so elevating the commissural leaflet with a nerve hook provides a valuable reference point from which the degree of elongation of other chordae may be determined. A total lack of structural integrity from chordal rupture results in a completely flail leaflet.

In patients with rheumatic disease, the chordae may be contracted and foreshortened, producing decreased motion of the associated leaflet. When this finding is present, the leaflet does not reach the proper plane of coaptation during systole.

Posterior Leaflet Procedures

Quadrangular resection of the posterior leaflet has become the mainstay of mitral valve reconstruction (Figs. 18-13 and 18-14). Diseased leaflet tissue in the posterior leaflet is excised with a quadrangular excision, usually removing 2 to 4 cm of tissue and, occasionally, over 50 percent of the leaflet. Strong chordae of proper length are identified on each side of the excised leaflet and encircled with retraction sutures. A rectangular excision is performed, cutting directly down to the mitral annulus, but not excising annulus.

Once the quadrangular excision has been performed, the annulus of the excised segment of leaflet is plicated with several interrupted 2-0 Tevdek sutures placed about 5 mm apart. These are started centrally and extended to include a few millimeters of annulus adjacent to the remaining leaflets. When these annular sutures are tied, the leaflet margins are automatically brought into apposition without any tension. If tension were to exist on the leaflet tissues, there would be a serious hazard of dehiscence of the subsequent leaflet repair. Once the annular sutures have been tied, the leaflet margins are approximated with simple or figure-of-eight sutures, usually 4-0 or 5-0 polypropylene, depending on the thickness of the leaflets.

Often the valve appears quite competent after the leaflet repair has been completed. Gently injecting saline into the ventricle with a bulb syringe and noting both the mobility of the leaflets and their apposition is an excellent visual guide for assessing competency. If localized insufficiency remains in other areas, additional procedures can be done.

Anterior Leaflet Procedures

Different approaches must be used for reconstructing the anterior leaflet. Although the techniques must be well planned and exacting, the results of repair requiring anterior leaflet reconstruction have been excellent. Three primary techniques are used: (1) chordal shortening, (2) chordal transposition from the posterior leaflet, and (3) triangular resection of redundant, prolapsing anterior leaflet tissue.

Shortening of elongated anterior leaflet chordae has been used to correct anterior prolapse in patients with intact chordae, although chordal transposition has been used more frequently and probably is more reliable (Fig. 18-15). A segment of posterior leaflet directly opposite the prolapsed anterior leaflet may be identified for chordal transposition. A small quadrangular excision of the posterior leaflet with the attached chordae is performed, and the mobilized leaflet and chordae then are transposed onto the anterior leaflet to provide structural support. The quadrangular defect in the posterior leaflet then is repaired by the previously described techniques.

An alternative method for treating anterior leaflet prolapse involves triangular resection of the involved area of the anterior leaflet (Fig. 18-16). The adjacent intact chordae are identified, and triangular resection is performed to remove the redundant, flail, or prolapsing segments of the leaflet and any ruptured chordae. The technique seems particularly helpful when a large amount of redundant anterior leaflet tissue is present, and this method of repair may limit the incidence of postoperative systolic anterior motion of the mitral valve after valve repair.

Repair of Leaflet Perforation

Leaflet perforations can be repaired by primary suture closure or by closure with pericardium after debridement of irregular leaflet edges. Extensive leaflet destruction is best managed with mitral valve replacement.

Ring Annuloplasty

Mitral annular dilatation, which often is a significant component of mitral insufficiency, is reliably corrected by ring annuloplasty. The ring acts to remodel the annulus, correcting insufficiency and stabilizing the leaflet and annular repair sites. The size of the mitral orifice is measured with a ring sizer. Once the proper size has been selected, the ring is inserted with a series of 2-0 Tevdek sutures (usually 10 to 14 sutures) carefully placed tangentially in the mitral valve annulus (see Figs. 18-13 and 18-14). Particular care is required in inserting these sutures to be certain that they are in the mitral annulus and not in the leaflet tissue or in the atrial wall and to avoid placing sutures lateral to the annulus, because the aortic valve cusps are only 2 to 3 mm beyond the annulus anterolaterally and the circumflex coronary artery a similar distance posterolaterally.

After the annuloplasty ring has been tied in position, saline is again injected gently into the left ventricle with a bulb syringe to assess leaflet mobility. If focal insufficiency remains after completion of the annuloplasty, several ancillary procedures can be considered. These include leaflet plication with figure-of-eight sutures or triangular excision of small prolapsing segments of leaflet. Elongated chordae may be shortened, and the commissures may be plicated in selected cases after the annuloplasty ring is in place.

Results

After Commissurotomy

The operative mortality rate for open mitral commissurotomy is less than 1 percent. In 1981 Gross reported late results in 202 patients undergoing commissurotomy between 1967 and 1978 with a 98 percent complete follow-up. Five years after operation 87 percent of patients who had no residual valve dysfunction after commissurotomy were free of any complications. The frequency of thromboembolism was 0.3 percent per year in the first 10 years after operation.

Similar results were reported in 1980 by Halseth, who described experiences with 222 patients operated on over a period of 10 years. The operative mortality rate was 1.5 percent, with only 7 percent of the survivors later requiring mitral valve replacement. Open mitral commissurotomy is a well-established standard against which newer and less invasive procedures must be compared. Surgical commissurotomy is a safe procedure, and, by allowing effective correction of both leaflet and subvalvular disease, late results are excellent in 80 to 90 percent of patients at 10 years.

Commissurotomy Versus Balloon Valvuloplasty

Percutaneous mitral balloon valvuloplasty has recently become established as an alternative treatment for mitral stenosis. The choice between surgical commissurotomy and balloon valvuloplasty varies widely among centers. A study by Palacios suggests that in patients with favorable anatomy the immediate outcome after balloon valvuloplasty is similar to short-term results obtained after open commissurotomy. A randomized trial reported by Reyes and associates demonstrated equal immediate and 36-month results with open commissurotomy and balloon valvuloplasty in appropriately selected patients. In this study overall survival, functional capacity, restenosis, and reoperation rates were equivalent at 3 years.

The late results after balloon valvuloplasty are less well established than the results after open commissurotomy. It is well known that open commissurotomy gives a much better late functional result than closed surgical commissurotomy, and balloon valvuloplasty appears to be functionally similar to the older closed surgical commissurotomy technique. Cohen and associates, in a report on 164 consecutive patients undergoing balloon valvuloplasty, open commissurotomy, or mitral valve replacement, obtained significantly better results at 3 years in patients treated with open procedures. The actuarial freedom from subsequent mitral valve procedures at 3 years was 66 percent after balloon valvuloplasty, versus 87 percent after open surgical commissurotomy and 100 percent after valve replacement. The linearized yearly reoperation rates were 12 percent, 4 percent, and 1.2 percent for balloon valvuloplasty, commissurotomy, and valve replacement, respectively.

After Mitral Valve Replacement

The operative mortality rate for mitral valve replacement is 2 to 5 percent, depending on the number of risk factors present. The major predictors of increased operative risk are poor ventricular function, NYHA functional classification, advanced age, previous cardiac surgery, and the presence of associated coronary artery disease or concomitant disease in another valve. Similarly, the major factors influencing long-term survival after mitral valve replacement are age, NYHA functional status, urgency of operation, mitral insufficiency (versus stenosis), ischemic disease, the level of pulmonary hypertension (pulmonary vascular resistance), and the need for concomitant valvular or coronary bypass procedures. Five-year survival rates after mitral valve operation are 60 to 90 percent, and 10-year survival rates are 40 to 75 percent, with late survival varying widely, depending on the number and the severity of the risk factors present.

Improved engineering of mechanical prostheses has produced lower intrinsic thrombogenicity and better durability. These improvements have resulted in fewer late valve-related complications and a lower late incidence of valve-related death after mitral valve replacement in recent years. In 1994 Khan reported late results after 1,000 St. Jude valve implants, with 399 patients undergoing isolated mitral valve replacement. At 10 years 83 percent of the mitral group was free of thromboemboli, 87 percent free of valve-related hemorrhage, and 91 percent free of valve-related death. The study suggests that valve-related factors have significantly less impact on late survival than patient-related factors.

Valve Reconstruction Versus Valve Replacement

In an analysis of 685 patients undergoing Carpentier-type valve repair at NYU between 1979 and 1995, the hospital mortality rate was 2.5 percent for isolated valve reconstruction and 5.9 percent overall. The actuarial freedom from complications at 5 and 10 years, respectively, were as follows: thromboembolic, 92 and 88 percent; anticoagulant-related complications, 98 and 96 percent; endocarditis, 97 and 96 percent; and reoperation, 91 and 84 percent. Postoperative freedom from reoperation was significantly better in nonrheumatic patients (93 percent at 5 years, 88 percent at 10 years) than in rheumatic patients (86 percent at 5 years, 73 percent at 10 years; $p < .005$).

Similar long-term results after mitral reconstruction were reported by DeLoche and Carpentier in 1990. In 206 patients undergoing operative repair between 1972 and 1979 the 15-year actuarial survival was 72 percent, and the 15-year survival from valve-related death was 83 percent. At 15 years, 94 percent of patients were free from thromboembolism, and 97 percent were free from endocarditis. Freedom from reoperation was better in patients with degenerative disease than in those with rheumatic disease (93 percent versus 76 percent).

In 1991 Yun and Miller extensively reviewed published data comparing mitral valve repair with replacement, analyzing reports from several large series. They found that long-term freedom from late valve-related morbidity generally was better after valve repair than after valve replacement.

Theoretically, the long-term survival after valve repair also could be better than after valve replacement because of the lower risk of valve-related complications and because of better preservation of left ventricular function as a result of chordal preservation. A review of all mitral valve patients operated on between 1980 and 1995 at NYU demonstrated better late survival in nonrheumatic patients undergoing isolated mitral valve repair than in patients undergoing valve replacement. In contrast, in rheumatic patients and in patients requiring double- or triple-valve procedures, late results in terms of freedom from late cardiac death, reoperation, and all valve-related complications were better after valve replacement using the St. Jude mechanical prosthesis. These findings suggest that while isolated mitral valve repair may offer improved late results in patients with degenerative, single-valve involvement, patients with rheumatic pathology and those with multivalvular involvement have better long-term results if valve replacement is performed.

The data suggest that mitral valve reconstruction is a durable and reproducible procedure in patients with degenerative disease, producing excellent long-term results with a low risk of valve-related complications. Patients with degenerative disease undergoing valve repair may have a better survival rate and a lower risk of late complications than those undergoing valve replacement. Valve reconstruction may be the procedure of choice for many patients with degenerative disease. Valve replacement is preferable if a durable repair cannot be achieved because of advanced rheumatic disease, extensive valvular deformity, or surgical inexperience with valvular reconstruction. Mechanical valve replacement gives improved durability and improved freedom from late valve-related events in patients with rheumatic or multivalve disease.

Minimally Invasive Mitral Valve Surgery

Because of advances in technology and instrumentation, such as the development of an endovascular perfusion and cardioplegic arrest balloon catheter system by Stevens and coworkers, minimally invasive surgery can be applied in an expanding number of complex heart operations, including mitral valve surgery. Extensive experimental studies on the feasibility, safety, and reproducibility of minimally invasive mitral valve surgery formed the basis for subsequent clinical trials, which established the feasibility and reproducibility of minimally invasive mitral surgery, with experimental results equivalent to those obtained with standard open-chest surgery. The port-access method of minimally invasive mitral valve surgery was subsequently applied clinically with excellent initial results.

The operative approach is through a right anterior minithoracotomy with a 5- to 8-cm skin incision made in the inframammary skin fold. The technique uses femoral cannulation and a balloon endoaortic occlusion system for cardioplegia (Fig. [18-17](#)). This approach provides excellent exposure of the mitral valve, and valve repair or replacement are readily achievable.

The initial experience with minimally invasive mitral surgery has been promising. Over 100 minimally invasive mitral procedures were performed at NYU within 6 months of the technique's introduction with no operative deaths and with excellent results. The advantages of the minimally invasive approach are less pain, smaller scar, improved cosmetic result, and shorter recovery time.

Aortic Valve Disease

Effective surgical treatment of aortic valve disease became possible in 1960 with the development of satisfactory prosthetic valves by Starr and Edwards and by Harken and associates. Earlier attempts to correct aortic valvular disease by cusp replacement with prosthetic cusps of Teflon cloth or by extensive debridement of calcific material from calcified valve cusps initially gave satisfactory results in some patients, but a high failure rate within 1 to 2 years led to abandonment of these techniques as soon as satisfactory prosthetic valves became available.

Several modifications of valvular prostheses have been proposed over the years, and numerous effective mechanical and tissue valves have been developed. Thromboembolic events occur with a frequency of 1 to 2 percent per year after mechanical aortic valve replacement. Bioprostheses (stented glutaraldehyde-preserved xenografts), which were developed to minimize the risk of thromboembolism, do not require permanent anticoagulation therapy when used in the aortic position, but they undergo progressive deterioration, with an increasing incidence of valve failure between 5 and 15 years. The failure rate for bioprostheses is approximately 5 percent at 5 years, 20 percent at 10 years, and 50 percent at 15 years.

The Ross procedure (pulmonary autograft) has a low risk of thromboembolism and does not require anticoagulation therapy. The risk of reoperation after the Ross procedure lies between those achieved with mechanical and with tissue prostheses, approximately 20 percent at 15 years and 30 percent at 20 years.

Aortic Stenosis Etiology

The three main causes of aortic stenosis in the adult population are congenital malformation, rheumatic disease, and acquired calcific disease. Congenital aortic stenosis is discussed in detail in Chap [20](#) but, in general, about 50 percent of adults undergoing operation have a congenitally malformed valve, usually a bicuspid valve. These patients frequently have minimal disability for decades until the valve calcifies, usually in the fourth or fifth decade of life. Rheumatic fever is the primary cause in 30 to 40 percent of patients. With rheumatic disease the degree of stenosis progresses with time, and concomitant mitral valve disease is almost always present, although not always clinically significant. The third major cause of aortic stenosis is acquired calcific stenosis, which is the most frequent cause in older patients, occurring in the seventh or eighth decade of life. Calcific stenosis appears to be a part of the aging process, with valve damage and calcification developing from years of flow and turbulence.

Pathophysiology

A normal aortic valve has a cross-sectional area of 2.5 to 3.5 cm², depending on body size. Moderate stenosis is present when the valve orifice has narrowed to less than 1 to 1.2 cm², and the patient often becomes symptomatic once the valve area is less than 0.8 cm². Cross-sectional areas of less than 0.5 cm² may be found in advanced disease, often with a systolic gradient of 100 mmHg or greater across the valve. In most patients with significant aortic stenosis the gradient is at least 50 mmHg, although the gradient may be misleading when the cardiac output is low. The valve area should always be calculated using the Gorlin formula. Nevertheless, gradients greater than 60 to 70 mmHg are almost always significant.

The increased workload on the myocardium imposed by aortic stenosis results in progressive concentric ventricular hypertrophy, but little ventricular dilatation. Left ventricular hypertrophy results in a thick, noncompliant ventricle that develops early diastolic dysfunction. The systolic function of the ventricle remains well-preserved early, decreasing later from “afterload mismatch” once the thickened ventricle cannot produce the tension required to effectively eject blood through the stenotic valve. Once systolic dysfunction occurs, the patient usually develops symptoms within 12 to 24 months, with congestive heart failure progressing thereafter.

Myocardial ischemia develops in some patients with severe aortic stenosis, usually in response to exercise. The left ventricular mass and left ventricular systolic wall tension are increased, resulting in increased oxygen demand. Simultaneously, the cardiac output often is low and does not increase in response to the demands of physical activity, while the end-diastolic pressure of the ventricle is high, resulting in poor perfusion of the subendocardium. This results in an imbalance between oxygen supply and demand. Symptoms include angina and arrhythmias, and sudden death can occur.

Clinical Manifestations

Characteristically, patients with aortic stenosis have a long, asymptomatic, latent period, sometimes for 10 to 20 years. A patient with moderate aortic stenosis with classic physical findings may exhibit only slight dyspnea or no symptoms at all for many years. Once the ventricle develops systolic dysfunction and begins to deteriorate, three classic symptoms may occur: exertional dyspnea (or heart failure), angina, and syncope. Once these symptoms develop, the risk of death exceeds 30 to 50 percent over the next 5 years, and operation is urgently indicated. Sudden death, which accounts for 15 to 20 percent of fatalities from aortic stenosis, becomes much more of a threat once these symptoms are present.

The most common symptom is exertional dyspnea with decreased exercise capacity and easy fatigability. True left ventricular failure, manifested as severe congestive heart failure with pulmonary edema, is a more ominous finding. The risk of death is high once heart failure is present, approximately 40 percent over the next 2 to 3 years. Atrial fibrillation, a consequence of prolonged elevation of left atrial pressure, is also a significant finding, because it can indicate an advanced left ventricular failure unless mitral valve disease is present.

Angina pectoris develops in 30 to 40 percent of the patients, a manifestation of myocardial ischemia. Many ischemic episodes are associated with “silent” muscle necrosis, because some patients with surprisingly few symptoms are found to have large amounts of myocardium replaced by scar tissue. Associated coronary artery disease should always be excluded when angina is present. The average life expectancy once angina or syncope has appeared is about 3 years.

Syncope develops in about 10 percent of the patients, apparently from decreased cerebral blood flow. In some patients it may occur after minimal effort and with little warning. More often syncope develops in response to exercise or increased physical activity because the cardiac output cannot respond appropriately. In a small proportion of patients it results from a conduction abnormality, apparently an intermittent heart block from involvement of the atrioventricular node by calcium arising from the stenotic valve.

The principal physical finding is a harsh, systolic murmur at the base of the heart (right second interspace) with radiation to the carotid arteries. The pulse pressure is narrow and sustained (pulsus parvus et tardus). The apical impulse has been described as a “prolonged heave,” not a “forceful thrust” that is found with ventricular dilatation from aortic or mitral insufficiency.

Diagnostic Studies

The heart size usually is normal or only slightly enlarged on chest x-ray, and calcification of the valve often is visible in older patients. The ECG is not reliable, because of the wide variation of findings. In some patients left ventricular hypertrophy is evident, but in some seriously ill patients with severe aortic stenosis the ECG is virtually normal. Conduction abnormalities are common, apparently from spicules of calcium projecting into the conduction bundle located just beneath the base of the noncoronary sinus. Some patients develop complete heart block. Atrial fibrillation generally indicates the presence of mitral stenosis or advanced disease with a prolonged elevation of intracardiac pressures.

The diagnosis of aortic stenosis can be made by echocardiography, which: (1) demonstrates the presence of calcium in the valve and demonstrates leaflet mobility; (2) provides an estimate the peak and mean transvalvular systolic gradients; (3) documents the ventricular wall thickness and the degree of left ventricular hypertrophy; and (4) measures the end-systolic and end-diastolic dimensions of the left ventricular cavity. The Doppler echocardiogram measures flow velocity to determine the peak instantaneous pressure gradient, but it may overestimate the peak transvalvular gradient. By echocardiography the mean gradient usually is a more accurate measure of the true transvalvular gradient and a better predictor of the degree of stenosis.

Cardiac catheterization readily confirms the diagnosis by measuring the transvalvular gradient and permitting calculation of the cross-sectional area of the valve. Operation usually is indicated when the cross-sectional area is less than 0.8 cm² and the gradient exceeds 50 mmHg. At catheterization coronary arteriography should be done in patients over 55 to 60 years of age and in patients with angina or a strong family history of coronary disease. Concomitant coronary artery disease is found in 30 to 50 percent of adult patients undergoing aortic valve surgery, with the frequency increasing with age. The mitral and tricuspid valves should be assessed for stenosis or insufficiency, particularly when rheumatic disease is present or when the patient has significant heart failure and pulmonary hypertension. The left ventricular function, left ventricular end-diastolic pressure, and the pulmonary artery pressure should be measured in patients with heart failure. Catheterization is not always necessary in younger patients.

Treatment

Operation is recommended for all symptomatic patients with aortic stenosis. For asymptomatic patients, periodic echocardiographic studies are done to assess the transvalvular gradient, the cardiac size, and the ventricular function. If a patient with a significant aortic gradient develops left ventricular hypertrophy and a drop in the ejection fraction (or a rise in the end-systolic volume by echocardiography), operation usually is indicated. Similarly, operation is recommended once the cross-sectional area is less than 0.8 cm², especially if significant left ventricular hypertrophy is present. Sudden death remains a small but definite hazard in such patients, and progressive clinical deterioration is likely to occur soon once this degree of stenosis is reached.

Aortic Insufficiency

Etiology and Pathology A variety of diseases can produce aortic insufficiency, including endocarditis, rheumatic fever, degenerative valvular disease, congenital disease, aortoannular ectasia, aortic aneurysm, and aortic dissection. Mixed disease, with stenosis and insufficiency, can develop in any patient with aortic stenosis.

Inflammatory diseases are the most common causes of aortic insufficiency. Perhaps the most frequent cause is bacterial endocarditis, occurring from infections with streptococci, staphylococci, or enterococci, in decreasing order of frequency. Rheumatic fever is another common cause of aortic insufficiency, while other inflammatory diseases such as syphilis are increasingly rare causes.

One variant of degenerative valvular disease is the so-called floppy valve, a type of myxomatous degeneration of the valve that produces thin and elongated tissue. The aortic leaflets sag into the ventricular lumen, often with no other histologic abnormality. The gross and histologic appearances suggest that this is a variant of the more common mitral valve prolapse.

Congenital aortic insufficiency rarely is present at birth, but it may develop in older patients if stiffening and calcification of the malformed bicuspid valve produces an insufficient rather than a stenotic valve. This accounts for 10 to 25 percent of the patients operated on for aortic insufficiency. Aortic insufficiency also occurs in association with subaortic ventricular septal defects from a venturi effect that results in prolapse of the aortic leaflet into the septal defect.

Aortoannular ectasia is an unusual type of connective tissue disease seen with increasing frequency as the average age of the population increases. It is often of unknown cause. Aortoannular ectasia occurs in its most extreme form with connective tissue disorders, such as Ehlers-Danlos syndrome and Marfan's syndrome. These diseases involve defective collagen synthesis or cross-linking that results in extensive cystic medial necrosis throughout the aorta, with particular involvement of the aortic root. These patients have a greatly increased risk of aortic dissection. The aortic root gradually enlarges, starting in the sinuses of Valsalva and progressing to a discrete aneurysm in the ascending aorta. The pathologic process is unusual because dilatation decreases and almost stops at the level of the innominate artery. The size and shape of the aneurysm is quite characteristic, resembling a truncated cone with the narrow apex near the level of the innominate artery. Aortic insufficiency results from dilatation of the aortic annulus.

In less severe forms of aortic degenerative disease there may be a localized aneurysm in the ascending aorta, with or without aortic insufficiency. Often no other signs of connective tissue disease are present, although histologic examination of the excised aneurysm usually demonstrates characteristic cystic medial necrosis. Atherosclerotic aneurysms also may produce aortic insufficiency by dilatation of the annulus, although aortic insufficiency is less common in patients with pure atherosclerotic disease. Aortic dissection produces insufficiency by dissection of the aortic wall with detachment and prolapse of the valve cusps, usually the noncoronary cusp. Aortic dissections and aneurysms are discussed in more detail in Chap. [19](#).

Pathophysiology

Surprisingly large volumes of blood regurgitate into the ventricle with severe aortic insufficiency. This results in an increased ventricular end-diastolic volume and a compensatory increase in the left ventricular stroke volume that may be two or three times greater than the normal stroke volume of 60 to 75 mL. As the ventricular diastolic volume increases

(increased preload), dilatation and eccentric hypertrophy develop to maintain the ratio of wall thickness cavity radius at normal levels. The diastolic pressure may not increase until later in the clinical course, so symptoms of pulmonary congestion appear only with advanced disease. The process of volume overload and ventricular dilatation contrasts with aortic stenosis, in which systolic pressure overload and concentric ventricular hypertrophy develop early, without dilatation. With a continued volume overload the left ventricular end-diastolic pressure rises to 20 to 30 mmHg. Eventually the ratio of wall thickness to cavity radius decreases as cardiac dilatation continues, the increased systolic wall tension exceeds the ability of the ventricular muscle to eject, resulting in “afterload mismatch,” and the systolic function of the ventricle begins to deteriorate.

With marked dilatation of the left ventricle and increased intracardiac pressures mitral insufficiency may begin to develop. When a rheumatic history is present, it may be impossible to determine preoperatively whether the mitral insufficiency represents dilatation with secondary insufficiency or rheumatic disease. Secondary mitral insufficiency usually regresses satisfactorily after replacement of the aortic valve.

Clinical Manifestations

Symptoms develop at a variable rate in patients with aortic insufficiency, depending on the acuity and severity of the insufficiency and the characteristics of the heart muscle. Frequently the patient who gradually develops moderate to severe insufficiency remains asymptomatic for many years, often ten or more. Once symptoms appear, however, the ventricular function usually is significantly depressed and rapid clinical deterioration occurs over the next 4 to 5 years. In the presurgical era, about 40 percent of patients died within 10 years and 90 percent within 20 years after the onset of aortic insufficiency. The terminal illness usually is progressive cardiac failure; sudden death is much less common than with aortic stenosis.

The most common symptom is dyspnea on exertion and a decreased exercise capacity, gradually increasing in severity as the ventricle deteriorates. Palpitations also are common, apparently arising from forceful contraction of the dilated left ventricle. Biventricular heart failure and angina pectoris occur with advanced disease, usually with severe aortic incompetence in which the regurgitant flow exceeds 50 percent of forward flow.

Palpation readily discloses an enlarged heart and a prominent cardiac impulse, located inferiorly and to the left of the normal location. The hallmark of aortic insufficiency is a high-pitched decrescendo diastolic murmur along the left sternal border, starting immediately after the second sound. The length of the murmur may correspond with the severity of the insufficiency, but the intensity does not. If the murmur is loudest to the right of the sternum, dilatation of the aortic ring, as in Marfan's syndrome, is likely. An ejection systolic murmur of moderate intensity is frequent even in the absence of aortic stenosis because of increased flow. An S₃ gallop indicative of heart failure may be present. In some patients a middiastolic rumble is noted in the mitral valve area—the Austin Flint murmur that simulates mitral stenosis but is produced by the diastolic flow of blood from the aortic insufficiency and impedes opening of the mitral valve during diastole.

Examination of the peripheral arterial circulation usually reveals several abnormalities. The pulse pressure is widened, partly from an increase in systolic pressure, but principally from a decrease in diastolic pressure, which may be in the range of 30 to 40 mmHg. The true diastolic pressure, measured by direct arterial puncture, is never less than 30 to 35 mmHg, even though on auscultation a diastolic pressure of zero may be obtained in some patients. The exact level of diastolic pressure is not closely correlated with the severity of the aortic insufficiency, because of the influence of peripheral resistance. The peripheral pulses usually are visible, forceful, and bounding. The pulse is described as “water-hammer,” or quickly collapsing, referred to as a Corrigan's pulse. “Pistol shot” sounds are readily heard with the stethoscope over peripheral arteries. A wide variety of other auscultatory phenomena have been described, some over a century ago, all indicating vasodilatation and a hyperactive peripheral circulation.

Diagnostic Studies

The chest x-ray usually shows impressive cardiac enlargement, with the apex displaced downward and to the left. If the cardiothoracic ratio is 0.5 or less, asymptomatic patients may be monitored periodically with biannual x-rays as long as the heart size is normal. The ECG is normal early in the disease, but with cardiac enlargement signs of left ventricular hypertrophy become prominent. The cardiac rhythm remains sinus early in the disease, although atrial fibrillation is common after the onset of advanced disease and has an ominous prognosis, usually representing chronic severe volume overload with severely elevated intracardiac pressures.

The degree of the aortic insufficiency and the size and function of the left ventricle can be evaluated more precisely with echocardiography, which measures the degree of valvular insufficiency, the end-systolic and end- diastolic ventricular dimensions, the left atrial size, and the ejection fraction. The echocardiogram also can assess the degree of mitral or tricuspid insufficiency, rule out associated valvular stenosis, and estimate the pulmonary artery pressure.

The classic finding on cardiac catheterization is the visible reflux of dye from the aortic root into the ventricle with angiography, graded from 1+ to 4+. In patients with chronic disease or cardiac failure the left ventricular end- diastolic pressure usually is elevated to 15 to 20 mmHg and sometimes higher, often with an associated rise in the pulmonary artery pressure. The aortic root should be assessed for aneurysmal disease, and coronary angiography should be performed to rule out coronary artery disease.

Operative Indications

While surgery is indicated in all symptomatic patients with aortic insufficiency, it has long been recognized that postponing operation until symptoms are disabling is not satisfactory. Many patients already will have developed substantial ventricular enlargement and cardiac dysfunction by that time. Clinical investigation has sought a noninvasive measurement that would identify the proper time for operative intervention. This is best done by periodic echocardiography.

Operation is recommended for asymptomatic patients with aortic insufficiency at the first sign of deteriorating systolic cardiac function, manifested as a rise in the end-systolic dimension of the ventricle echocardiographically or by a drop in the ejection fraction measured by radionuclide studies. All indications suggest that cardiac function will return to normal and long-term survival will be improved significantly if an operation is done at this stage. Intervention should be performed before the ventricular function severely deteriorates, usually when the end-systolic dimension of the left ventricle exceeds 45 to 50 mmHg or when the resting ejection fraction drops below 50 percent. Similarly, a rapid increase in cardiac size or a progressive increase in the diastolic volume of the heart over a 6- to 12-month interval is an indication for operative intervention before systolic dysfunction develops.

The development of dyspnea or other symptoms of congestive heart failure are clear indications for operation. In some NYHA functional class IV patients with advanced left ventricular dysfunction, how much improvement can be expected from valve replacement is uncertain; the principal symptoms may be from progressive cardiomyopathy. Available studies do not permit a precise scientific decision in this regard. In the authors' experience, valve replacement can be performed even in the most advanced patients with an operative mortality rate of less than 10 percent. Many of these patients will improve significantly after surgery, although the degree of improvement may be uncertain for 6 to 12 months. Postoperative medical therapy with afterload reduction, digoxin, and diuretics reduction is beneficial in this group of patients after valve replacement to control heart failure and to help remodel the ventricle.

Operative Techniques

Aortic Valve Replacement At NYU The St. Jude Medical bileaflet mechanical valve is the most widely used prosthesis in younger patients, although the Medtronic-Hall, the CarboMedics, and the Omniscience valves are equally efficacious. While the flow characteristics are excellent with all of the mechanical valves, new designs that allow supraannular rather than intraannular placement (St. Jude HP valve and CarboMedics “Top Hat”) have allowed better flow when the aortic annulus is small. Bioprostheses (the Carpentier-Edwards porcine valve, the Carpentier-Edwards pericardial valve, or the Hancock porcine valve) are used primarily in older patients with acquired calcific stenosis.

Aortic valve replacement is performed through a median sternotomy, using cardiopulmonary bypass for perfusion and cardioplegia for myocardial protection. This technique has been firmly established for over 20 years with a high degree of safety and reproducibility. After cardiopulmonary bypass and moderate systemic hypothermia have been initiated, the aorta is cross-clamped and cardioplegic solution is delivered directly into the coronary ostia or retrogradely through the coronary sinus venous system. Topical hypothermia is used routinely to prevent cardiac rewarming, wrapping the heart with a cold gauze pad or cooling device. This technique results in total recovery of cardiac function postoperatively in most patients. The heart usually is vented with a catheter placed through the right superior pulmonary vein or through the apex of the left ventricle.

After the heart has been arrested, an oblique aortotomy incision is made, beginning approximately 1 cm above the right coronary artery and extending medially toward the pulmonary artery and inferiorly into the noncoronary sinus of Valsalva. The aortic valve is excised totally (Fig. 18-18), with care being taken to avoid losing any calcific fragments detached during removal of the valve. After the valve is excised with a scissors or a knife, calcium in the annulus is debrided as necessary to achieve a smooth annular surface for suture placement. A gauze pack is placed routinely in the ventricle before removal of the valve is begun to minimize the risk of embolism.

A pledgeted horizontal mattress suture technique is used almost exclusively at NYU (see Fig. 18-18); other techniques include figure-of-eight sutures, simple interrupted sutures, and continuous sutures. The pledgeted mattress suture technique probably is unnecessary in many patients, but it virtually eliminates the hazard of periprosthetic leaks. Care is taken to avoid damage to the coronary ostia, to the conduction bundle (which is adjacent to the membranous septum between the right and the noncoronary valve cusps), and to the mitral valve posteriorly.

Most mechanical valves are placed in the intraannular position, and the sutures are placed from above downward, everting the annulus. When the sutures are tied the pledgets remain visible in the aorta and the valve sits completely intraannularly. This technique avoids tissue overhang, which could interfere with leaflet movement, but it requires placing a slightly smaller valve size.

The supraannular technique is used for most bioprostheses and for newer supraannular mechanical valves, designed for the small annulus. The sutures are placed from below the annulus upward, passing the needles through the annulus and then through the valve sewing ring. The valve is seated on the top of the annulus, allowing placement of a slightly larger prosthesis, which is important if the annulus is small (Fig. 18-19).

Coronary bypass with the internal mammary artery or saphenous vein graft is performed in standard fashion whenever associated coronary disease is present.

Aortic Valve Repair

Though aortic valve repair was one of the first attempted cardiac operations, it had only limited success, and these procedures were infrequently used once prosthetic valves became available. Surgeons have occasionally proposed newer techniques of aortic valve repair in an effort to avoid the long-term complications inherent with valvular prostheses, but most of these techniques have been abandoned because of the high incidence of late failure. The various treatments of aortic stenosis involving debridement of the aortic valve resulted in a high incidence of progressive valvular restenosis or insufficiency, and these techniques are now seldom used. Two newer techniques of valve repair for pure aortic insufficiency have achieved some success and deserve mention.

Fraser and associates reported results with repair of insufficient bicuspid aortic valves in 72 patients. The operative technique involves a triangular resection of the redundant

segment of the involved valve cusp in an attempt to achieve cusp symmetry, followed by annular plication of one or both commissures in an attempt to achieve greater leaflet coaptation. Short-term results have been good, with approximately 90 percent of patients remaining free from reoperation at 2 years.

David and associates proposed a different but similarly innovative technique for repair of insufficient aortic valves in patients with aortic root aneurysms. This technique is based on the principle that aortic insufficiency in patients with aortoannular ectasia is from annular dilatation or distortion of the sinotubular ridge or both. Patients with annular dilatation and sinotubular distortion were treated with aortic annuloplasty and tubular graft placement, with reimplantation of the native aortic valve into the graft. Patients without annular dilatation were treated by sinotubular remodeling. Early results have been good in the initial 45 patients reported by David, with only two patients requiring reoperation.

Ross Procedure

The Ross procedure involves removal of the aortic valve and replacement of the valve (free-hand technique) or removal of the entire aortic root (cylinder root replacement technique) (see Fig. 18-10) and replacement with a pulmonary valve autograft transplanted from the native pulmonary position; the pulmonary valve and main pulmonary artery are replaced with a pulmonary homograft.

The cylinder root replacement technique, which is the most reproducible, will be described briefly. The native aorta is transected 5 mm above the sinotubular ridge. The aortic valve and sinuses of Valsalva are removed totally, preserving the left main and right coronary arteries on buttons of aortic tissue. The main pulmonary artery is transected at the bifurcation, and a separate incision is made below the pulmonary valve in the right ventricular outflow tract. The pulmonary valve and artery are enucleated en bloc from the outflow tract bed, with care being taken to avoid injury to the first septal perforator as the valve is removed from the septum medially. The pulmonary autograft annulus is sutured to the aortic annulus with continuous or interrupted sutures, and the coronary arteries are reimplanted into the autograft. The pulmonary valve and right ventricular outflow are reconstructed in standard fashion with a homograft.

The operative risk and perioperative complications with the Ross procedure have been similar to those with standard valve replacement. These patients do not require long-term anticoagulation therapy, and the risk of thromboembolism is negligible. The late durability is better than that achieved with bioprostheses. The Ross procedure has become an attractive option for many young patients requiring aortic valve replacement.

Minimally Invasive Aortic Valve Surgery

Recent attempts have been successful at using a “minimal incision” for aortic valve replacement. While these methods are still under development, most have involved femoral cannulation for perfusion and the use of a small transverse or vertical incision in the right third costal interspace. The small incision limits pain and potentially shortens the recovery time.

Postoperative Care

Postoperative care usually is uneventful after aortic valve replacement. Arrhythmias are relatively common, so ECG monitoring is routine for 24 to 48 h. Anticoagulation therapy is started 2 to 3 days after mechanical valve replacement, targeting an INR (international normalized ratio) of approximately 3 to 3.5 times normal. Antiplatelet therapy with aspirin is used with bioprostheses and for 3 to 6 months after homografts or the Ross procedure. The hospital stay usually is 5 to 7 days, depending on the patient's age and overall physical condition.

Except for patients with severe ventricular dysfunction, patients usually become asymptomatic and regain a normal range of physical activity within 2 to 3 months after operation. Periodic medical supervision should be instituted for all patients because of the problems inherent with any prosthetic valve. The patient should have long-term monitoring for proper anticoagulation therapy (with the INR followed in patients on warfarin), and the valve function should be periodically assessed by echocardiography. In patients with homografts or the Ross procedure, yearly echocardiographic evaluation is recommended to assure continued proper valve function. Thromboembolism, anticoagulant-related hemorrhage, endocarditis, and prosthetic valve failure are the principal late complications that require periodic monitoring. With good anticoagulation therapy, thromboembolism occurs with an incidence of 1 to 2 percent per year after mechanical valve replacement and 0.5 to 1 percent per year after bioprosthetic valve replacement. The risk of thromboembolism approaches zero after the Ross procedure. Endocarditis remains an infrequent but serious late hazard after valve replacement, and routine antibiotic prophylaxis is recommended lifelong for any invasive procedure that might produce a transient bacteremia.

Results

The operative mortality rate for aortic valve replacement is 1 to 2 percent in uncomplicated patients and seldom exceeds 10 percent even with advanced disease and pulmonary hypertension. In patients over 70 years of age the operative risk is approximately 5 percent for isolated aortic valve replacement and 10 to 15 percent when associated coronary artery disease is present. Christakis and associates, Kirklin and Barrett-Boyes, and the Society of Thoracic Surgeons National Database have each reported the clinical variables affecting the operative risk for aortic valve replacement. The major risk factors are age, NYHA functional status, congestive heart failure, ventricular function (ventricular ejection fraction), cardiomegaly, pulmonary artery pressure, unstable hemodynamics, angina or associated coronary disease, concomitant mitral or tricuspid valve disease, aortic insufficiency (as opposed to stenosis), preoperative renal insufficiency, and reoperation. The mode of perioperative death usually is stroke, operative hemorrhage, or arrhythmia. Heart block has become an uncommon complication.

The 10-year survival rate after aortic valve replacement exceeds 80 percent for most patient groups, but it is significantly worse in patients with severely impaired ventricular function, NYHA class IV symptoms, and pulmonary hypertension, emphasizing the need for early operation when signs of deteriorating ventricular function are found in diagnostic studies.

Tricuspid Stenosis and Insufficiency

Etiology

Organic disease of the tricuspid valve is almost always a result of rheumatic fever. With the exception of septic endocarditis, usually in intravenous drug abusers, it virtually never occurs as an isolated lesion, but only in association with extensive disease of the mitral valve. With mitral disease the frequency of associated tricuspid disease is 10 to 15 percent, although an incidence as high as 30 percent has been reported. Rarely, blunt trauma produces rupture of a papillary muscle or chordae tendineae with resulting tricuspid insufficiency.

Tricuspid insufficiency is the more common lesion encountered; pure stenosis is infrequent, because stenotic lesions usually have concomitant insufficiency. Functional tricuspid insufficiency is much more common than insufficiency from organic disease. It develops from dilatation of the tricuspid annulus and right ventricle as a result of pulmonary hypertension and right ventricular failure. These abnormalities result from left ventricular failure and chronic elevation of left atrial pressure.

Pathology

With tricuspid stenosis, the pathologic changes are similar to those found with the more familiar mitral stenosis. There is fusion of the commissures to form a small central opening 1 to 1.5 cm in diameter. As right atrial pressure is normally only 4 to 5 mmHg, significant tricuspid stenosis may be present with a valve orifice considerably larger than that seen with mitral stenosis. With rheumatic disease, combined stenosis and insufficiency or pure insufficiency results from fibrosis and contraction of the valve leaflets, often in association with shortening and fusion of chordae tendineae. Calcification is rare. More commonly functional dilatation of the tricuspid annulus results in tricuspid insufficiency. The valve leaflets appear stretched but otherwise are pliable and seemingly normal even though serious regurgitation is present. The dilatation and deformity of the annulus are irreversible. Valves with severe functional insufficiency usually do not regain competency, even though the mitral valve disease is corrected and pulmonary artery systolic pressure returns to normal.

Pathophysiology

With tricuspid stenosis or severe insufficiency the mean right atrial pressure becomes elevated to 10 to 20 mmHg and sometimes higher. The higher pressures are found with a tricuspid valve orifice smaller than 1.5 cm² and a mean diastolic gradient between the atrium and ventricle of 5 to 15 mmHg or in patients with pulmonary hypertension and severe insufficiency. When the mean right atrial pressure remains above 15 mmHg, hepatomegaly, ascites, and leg edema usually appear.

A moderate degree of tricuspid insufficiency may be tolerated with little adverse influence on the circulation except for a decrease in cardiac output. This is in striking contrast to mitral insufficiency, in which the regurgitating blood and elevation of left atrial pressure produce pulmonary congestion. The unusual patient with isolated tricuspid insufficiency produced by a traumatic injury may do well for years, as the only physiologic disturbance is elevation of venous pressure and a decrease in cardiac output. The purest example of the surprising tolerance for tricuspid insufficiency is seen in the intravenous drug abuser with septic endocarditis who has been treated by total excision of the tricuspid valve. Some but not all patients tolerate absence of the tricuspid valve with total tricuspid insufficiency for months or years. In most patients chronic, severe tricuspid insufficiency eventually produces signs of right heart congestion.

Clinical Manifestations

The symptoms and signs of tricuspid valve disease are similar to those of right heart failure resulting from mitral valve disease. These result from chronic elevation of right atrial pressure above the range of 15 to 20 mmHg. The most familiar ones are edema, ascites, jugulovenous distention, and hepatomegaly. Characteristic murmurs are present and may be associated with hepatic pulsations. With long-standing severe tricuspid insufficiency hepatic dysfunction and clotting abnormalities may develop. As similar findings result from right heart failure without tricuspid disease, the concomitant presence of tricuspid disease in the patient in heart failure with mitral valve disease may be easily overlooked.

The characteristic murmur of tricuspid stenosis is best heard as a diastolic murmur at the lower end of the sternum. It is a low-pitched murmur of medium intensity and can easily be overlooked, because it is well localized at the lower end of the sternum. During inspiration the intensity of the murmur increases as the volume of blood returning to the heart is temporarily raised by an increase in intrathoracic negative pressure. Tricuspid insufficiency produces a prominent systolic murmur at the lower end of the sternum and also at the cardiac apex, where it may be confused with the systolic murmur of mitral insufficiency. The murmur often is found in association with an enlarged, pulsating liver and prominent, engorged peripheral veins. A prominent jugular pulse, especially when the cardiac rhythm is sinus, may be the best clue to unsuspected tricuspid disease. The liver often is enlarged and pulsatile, and a hepatojugular reflex may be noted.

Diagnostic Studies

The x-ray shows enlargement of the right atrium and right ventricle. Prominent P waves may be visible on the ECG if a sinus rhythm is present. Echocardiography confirms enlargement of the right atrium and ventricle and readily detects tricuspid stenosis or insufficiency, although it may overestimate the degree of insufficiency. Cardiac catheterization frequently is required to confirm the diagnosis.

If the blood volume and cardiac output are adequate, intraoperative palpation has been useful in helping to determine the significance of tricuspid insufficiency, correlating this

information with clinical, physical, and echocardiographic findings.

Treatment

Usually the surgical decision in tricuspid insufficiency is a tentative one until the valve is examined at operation. Mild degrees of tricuspid insufficiency usually are left alone, especially in the absence of pulmonary hypertension. The degree of hypertrophy of the right atrial wall is a helpful guide, because the absence of significant right atrial hypertrophy indicates that chronic severe elevation of right atrial pressure has not been present.

With significant tricuspid insufficiency, annuloplasty or tricuspid replacement usually should be done. In some cases this must be performed as an isolated procedure, often as a reoperation for persistent right heart failure after previous surgery on the other valves. Accordingly, the authors prefer to perform concomitant tricuspid valve repair when severe tricuspid insufficiency accompanies mitral disease or multiple valve disease. Effective tricuspid valve repair in these circumstances minimizes the risk of late right-sided heart failure and reoperation.

In the majority of patients tricuspid disease is from dilatation of the annulus, which is evidenced not only by the large annulus, but also by the absence of fibrotic changes in the leaflets. Typically the leaflets appear entirely normal. Virtually all such patients can be treated by annuloplasty. Data on more than 300 patients from our institution show that the posterior leaflet annuloplasty is simple, safe, and reproducible in the absence of significant intrinsic leaflet disease. In our experience, tricuspid valve repair was 98 percent durable at 7-year follow-up using this technique. The Carpentier ring annuloplasty may be used if the annulus is extremely distorted and the right heart pressures are expected to remain chronically elevated. Excellent results with tricuspid repair using the Carpentier technique have been reported.

In the minority of patients whose tricuspid stenosis is from commissural fusion, a commissurotomy may be performed. This often is combined with the Carpentier annuloplasty. More commonly valve replacement is performed when stenosis is present.

Valve replacement is seldom necessary for pure tricuspid insufficiency, except in patients with significant pulmonary hypertension and leaflet disease precluding annuloplasty. In a 1986 report of experiences with 151 valve replacements and 63 valve repairs, the prosthetic valve subsequently had to be replaced in 20 patients, principally because of progressive thrombosis. The 10-year durability of tricuspid bioprostheses is higher than the 80 percent 10-year durability with mitral or aortic bioprostheses, in which higher pressures are present. The choice of optimal prosthesis in the tricuspid position is uncertain.

When the prosthetic valve is inserted, care is required in suture placement along the septal leaflet, where the conduction bundle is located. In this area sutures should be placed more superficially, through the base of the septal leaflet, to avoid injury to the conduction bundle. Because of the proximity of the conduction tissue to the tricuspid annulus in this area, heart block is more frequent than after replacement of other valves.

In the patient with tricuspid endocarditis, Arbulu demonstrated that total excision of the tricuspid valve without replacement could be tolerated. This approach permitted removal of all infected tissue without insertion of a foreign body, increasing the likelihood of cure of the endocarditis with antibiotics. Of the 50 long-term survivors, 11 subsequently required prosthetic replacement.

Some have questioned this approach, because there are few data to indicate that insertion of a prosthetic valve is associated with higher frequency of recurrent endocarditis if the proper antibiotics are given, yet the overall hemodynamic response is better when valve replacement is performed. Our experience with tricuspid endocarditis supports the use of valve replacement at the initial operation.

The operative mortality rate for isolated tricuspid valve replacement is 1 to 2 percent. In earlier years the reported mortality rate for patients undergoing tricuspid surgery in conjunction with aortic or mitral surgery was high, running from 25 to 40 percent, primarily because the presence of tricuspid disease represented far advanced cardiac failure. An additional cause of high mortality was probably inadequate myocardial preservation of the hypertrophied right ventricle. With present techniques, however, tricuspid surgery seems to add only a small additional risk to aortic or mitral surgery. A study from our institution with data from 1976 to 1985 noted an operative risk of 6.3 percent for combined mitral and tricuspid surgery.

Multivalvular Disease

Disease involving multiple valves is relatively common, particularly in patients with rheumatic disease. Prominent signs in one valve can readily mask disease in others. With aortic valve disease, functional mitral insufficiency can result from a progressive rise in the left ventricular end-diastolic pressure and volume. Similarly, mitral valve disease may result in pulmonary hypertension, right heart failure, and functional tricuspid insufficiency. Often these secondary functional changes resolve without treatment if the primary pathology is corrected. Multiple-valve surgery has a higher operative risk than single-valve procedures, because the condition often represents more advanced disease or is associated with significant cardiac dysfunction.

In a 1992 report by Galloway and colleagues, 513 patients with multiple-valve disease treated surgically between 1976 and 1985 were followed to assess factors influencing operative risk and long-term survival. Three groups accounted for the majority of the cases: 58 percent had aortic and mitral valve disease (AV + MV), 29 percent had mitral and tricuspid valve disease (MV + TV), and 12 percent had triple-valve disease (AV + MV + TV). Preoperative congestive heart failure was present in 91 percent, 41 percent were NYHA class III, and 54 percent were NYHA class IV. The average pulmonary artery systolic pressure was 60 mmHg. Despite chronic symptoms and severe disease the overall operative mortality rate was 12.5 percent and the 5-year survival rate was 67 percent. The variables predicting decreased survival time were systolic pulmonary artery pressure, age, triple-valve procedure, concomitant coronary disease requiring bypass operation, previous heart surgery, and diabetes. After operation 80 percent of the patients improved to NYHA class I or class II, demonstrating that most patients had significant clinical improvement despite advanced disease. The 5-year freedom from late cardiac-related complications or death was 82 percent.

AV + MV

Nine combinations of valvular pathology can produce AV + MV, because each valve can be stenotic, insufficient, or both. Stenosis in both valves may lead to underestimation of the degree of aortic stenosis, because return of blood to the left ventricle is limited as a result of mitral stenosis. Aortic insufficiency, which produces the Austin Flint murmur, might overshadow and mask true mitral stenosis. With functional mitral insufficiency resulting from severe aortic disease, aortic valve replacement can lead to resolution of insufficiency in some patients, but patients with more severe mitral insufficiency may require mitral annuloplasty or valve replacement. Each patient should be examined closely for these and other considerations, and cardiac catheterization usually should be performed before operation. The operative mortality rate for isolated AV + MV replacement is 4 to 5 percent.

MV + TV

Nine combinations of valvular pathology are possible with mitral and tricuspid disease, but mitral disease with functional tricuspid insufficiency is the common scenario, resulting from chronic pulmonary hypertension and right heart failure. The operative mortality rate for isolated MV + TV replacement is approximately 6 percent.

Triple-Valve Disease

Triple-valve surgery can be challenging, because the clinical condition usually is a result of chronic aortic and mitral disease with severe pulmonary hypertension, biventricular failure, and functional tricuspid insufficiency. Occasionally rheumatic disease affects all three valves. The degree of pulmonary hypertension is the most significant predictor of survival with triple-valve disease. The overall operative mortality for 61 triple-valve procedures in the above-cited NYU report (Galloway et al.) was 23 percent, but only 5.6 percent when the pulmonary artery systolic pressure was less than 60 mmHg. This emphasizes the value of early operation, before the development of irreversible ventricular dysfunction and chronic pulmonary hypertension.

CARDIAC TUMORS

General Considerations

Primary cardiac neoplasms are rare, reported to occur with incidences ranging from 0.001 to 0.3 percent in autopsy series. Benign tumors account for 75 percent of primary neoplasms, and malignant tumors account for 25 percent. The most frequent primary cardiac neoplasm is myxoma, comprising 30 to 50 percent. Other benign neoplasms, in decreasing order of occurrence, include lipoma, papillary fibroelastoma, rhabdomyoma, fibroma, hemangioma, teratoma, lymphangioma, and others. Most primary malignant neoplasms are sarcomas (angiosarcoma, rhabdomyosarcoma, fibrosarcoma, leiomyosarcoma, liposarcoma), with malignant lymphomas accounting for 1 to 2 percent. Metastatic cardiac neoplasms are more common than primary neoplasms, occurring in 4 to 12 percent of patients dying of cancer.

Symptoms include dyspnea, fever, malaise, weight loss, arthralgias, and dizziness. Clinical findings may include murmurs of mitral stenosis or insufficiency, heart failure, pulmonary hypertension, and systemic embolization.

The diagnosis usually is readily established by 2-D echocardiography. Transesophageal echocardiography may be useful when transthoracic findings are equivocal or confusing. MRI has been of value in diagnosis, providing excellent cardiac definition. Cardiac catheterization is not necessary in the majority of cases but may be necessary when other cardiac disease is suspected or if other diagnostic studies are equivocal.

Excision is the treatment of choice and is possible for most benign tumors. Care is taken to avoid deformity or destruction of adjacent cardiac structures, and reconstruction of the involved cardiac chamber occasionally is necessary. Total excision of metastatic or primary malignant neoplasms is less frequently possible but should be attempted. Otherwise incisional diagnostic biopsy is performed. Multimodality therapy with excision, chemotherapy, and radiotherapy is indicated for most malignant cardiac neoplasms.

Myxoma

Sixty to 75 percent of cardiac myxomas develop in the left atrium, almost always from the atrial septum near the fossa ovalis. Most other myxomas develop in the right atrium. Fewer than 20 have been reported in the right or left ventricle. The curious predilection for a myxoma to develop from the rim of the fossa ovalis in the left atrium has been studied, but no satisfactory explanation has been found.

Myxomas are true neoplasms, although their similarity to an organized atrial thrombus has led to considerable debate. The occurrence of myxomas in the absence of other organic heart disease, histochemical studies demonstrating mucopolysaccharide and glycoprotein, and a distinct histologic appearance indicate that myxomas are true neoplasms. The tumors recur locally, but they do not invade or metastasize and are considered benign.

Pathology

The tumors usually are polypoid, projecting into the atrial cavity from a 1- to 2-cm stalk attached to the atrial septum. The size ranges from 0.5 cm to larger than 10.0 cm. Only the superficial layer of the septum is involved; invasion of the septum does not occur. Some myxomas grow slowly; a few patients have symptoms for many years. There is no tendency to invade other areas of the heart; distant metastases are rarely reported. The friable consistency of a myxoma is of particular significance, because fatal emboli have occurred after digital manipulation of the tumor at operation.

Histologically, a myxoma is covered with endothelium and composed of a myxomatous stroma with large stellate cells mixed with fusiform or multinucleated cells. Mitoses are infrequent. Lymphocytes and plasmacytes are regularly found. Hemosiderin, a result of hemorrhage into the tumor, also is common.

Sporadic myxomas usually present in the fifth or sixth decade of life but have been described in younger and older patients. Autosomal dominant genetically transmitted familial myxomas can occur, usually presenting before 30 years of age. Familial “myxoma syndrome” includes myxomas, freckles, pigmented nevi, nodular adrenal cortical disease, and mammary myomatous fibroadenomas. Testicular tumors and pituitary adenomas with two or more components are required for diagnosis.

Pathophysiology

A myxoma might cause no difficulty until it grows large enough to obstruct the flow of blood through the mitral or tricuspid valve or fragments to produce peripheral emboli. The frequency of embolization, previously estimated to occur in 40 to 50 percent of patients, is not surprising, for an astonishing degree of to-and-fro motion of a myxoma, swinging on a small pedicle with each cardiac contraction, may be seen with echocardiography or angiography. Intermittent acute obstruction of the mitral orifice has been reported to produce syncope or even sudden death. Some myxomas produce generalized symptoms resembling an autoimmune disorder, including fever, weight loss, digital clubbing, myalgia, and arthralgia. These patients may have an immune reaction to the neoplasm, as elevated levels of interleukin-6 and elevated levels of antimyocardial antibodies have been described.

Clinical Manifestations

Symptoms may be those of mitral valve obstruction, resembling mitral stenosis, except for acute exacerbations, presumably because of transient lodging of the myxoma in the mitral orifice, peripheral embolization, or generalized autoimmune symptoms. The diagnosis often is made after an embolic episode, from histologic examination of the surgically removed embolus or as a result of subsequent diagnostic studies to determine the reason for embolism. The precision and reliability of 2-D echocardiography has greatly simplified diagnosis. Angiography is optional unless additional disease is suspected. CT has been reported to be helpful with small tumors, but MRI is more definitive. Abnormalities usually are found on examination of the heart and also on the electrocardiogram, but these are not diagnostic.

Treatment

Operation should be performed as soon as possible after the diagnosis has been established, because a disabling or fatal cerebral embolus is an ever-present hazard. A sternotomy incision is used. Once extracorporeal circulation has been established, the aorta is clamped to avoid embolism. Palpation is avoided. The right atrium is opened and the fossa ovalis incised to expose the stalk of the myxoma. The left atrium is then opened in the interatrial groove. With the tumor visualized, the segment of atrial septum from which the tumor arises is excised, after which the tumor is removed through the incision in the left atrium (Fig. 18-20). The defect in the atrial septum is closed primarily or with a small patch. The technique is simple and permits exploration of atria and ventricles.

A few cases of recurrent myxoma have been reported, some of which have been successfully operated on. These were thought to have resulted from inadequate excision of the site of origin, but some have recurred at more remote sites in the atrium, indicating the multipotential source of these unusual neoplasms. It seems prudent to perform periodic echocardiography routinely for several years after operation.

Larrieu and associates described experiences with 18 myxomas in a series of 25 cardiac tumors over a period of 24 years. Fyke and associates treated 21 patients with mitral myxoma in the first 7 years after the introduction of 2-D echocardiography. The operative mortality rate today is 1 to 2 percent.

Metastatic Neoplasms

Cardiac metastases have been found in 4 to 12 percent of autopsies performed for neoplastic disease. Although they have occurred from primary neoplasms developing in almost every known site of the body, the most frequent have been carcinoma of the lung or breast, melanoma, and lymphoma. Cardiac metastases involving only the heart are very unusual. Similarly, a solitary cardiac metastasis is infrequent; usually there are multiple areas of involvement. Cardiac involvement is common with leukemia or lymphoma, developing in 25 to 40 percent of patients. All areas of the heart are involved with equal frequency except the cardiac valves, perhaps because lymphatics are absent in valves.

The diagnosis of a primary cardiac malignant tumor may be suspected in a patient in whom an unexplained hemorrhagic pericardial effusion develops, especially in association with a bizarre cardiac shadow on the radiograph. Echocardiography should confirm the presence of an abnormal cardiac mass. Thoracotomy or sternotomy usually is required to establish the diagnosis. Combined chemotherapy and radiation is indicated, but only rarely is effective therapy possible.

Rhabdomyoma

A cardiac rhabdomyoma probably is not a true tumor but a hamartoma, representing a focal arrest and maturation of cardiac muscle. The nodules have also been termed *nodular glycogenic degeneration*, thought to be a manifestation of glycogen storage disease. About one-half of the patients have tuberous sclerosis of the brain. On histologic examination cells with large vacuoles are found in which the nuclei appear suspended by threads of cytoplasm, giving origin to the term “spider cell.” Associated adenoma sebaceum and benign kidney tumors occasionally are noted.

Although rhabdomyoma is said to be the most common cardiac tumor in children, it is a rare lesion. Reece and associates found that only about 110 cases were reported in the literature before 1984. The cardiac lesions may be solitary or multiple nodules or may present a diffuse infiltration of the cardiac muscle. The lesions do not grow.

Most rhabdomyomas on record were recognized in infancy, at an average age of 5 months. The disease is fatal, but whether the death is from the tumor or from associated disease is uncertain. Symptoms may result from obstruction of a ventricular chamber or from arrhythmias, such as recurrent ventricular tachycardia. Complete excision has been accomplished in a few patients. If tuberous sclerosis of the brain is not present, some infants may be operated on successfully and cured of potentially fatal arrhythmias.

Miscellaneous Tumors

Unusual benign lesions of the heart include fibromas, lipomas, angiomas, teratomas, and cysts. Fewer than 50 of each of these types of lesion have been reported. Fibromas have been found most frequently in the left ventricle, often as 2- to 5-cm nodules within the muscle. Sudden death, probably from a cardiac arrhythmia, has been reported with these tumors and may be the reason that only 18 percent of the reported tumors have been found in adults.

Lipomas are rare asymptomatic tumors found projecting from the epicardial or endocardial surface of the heart in older patients. Angiomas are small, focal, vascular malformations of no clinical significance, although in the literature four were reported to have been associated with a heart block. Pericardial teratomas and bronchogenic cysts are rare lesions that can cause symptoms from compression of the right atrium and obstruction of venous return. Most of these occur in children. Some of the larger cysts, up to 10 cm in diameter, may produce grotesque deformities from extensive invagination of the right atrial wall. Myxomas are by far the most common benign tumor in adults; they are seldom found in children except as part of the familial syndrome described in the previous section.

PERICARDITIS

Acute Pericarditis

Pericarditis results from acute inflammation of the pericardial space, resulting in substernal chest pain, ECG changes, and a pericardial friction rub. The pain often is inspiratory, worse in the supine position, and relieved by leaning forward. Associated ECG changes frequently occur, most commonly sinus tachycardia with concave upward ST segment elevation throughout the precordium. The ECG typically progresses to T-wave inversion, followed by the total resolution of all changes. The cause of acute pericarditis is variable, including infection, myocardial infarction, trauma, neoplasm, radiation, autoimmune diseases, drugs, nonspecific causes, and others (Table 18-3). Untreated pericarditis may result in progressive development of a pericardial effusion with subsequent cardiac tamponade. Infectious causes may result in septic complications. Chronic constrictive pericarditis may develop after resolution of the acute process.

Diagnosis

The diagnostic work-up should attempt to determine the underlying cause of the pericarditis. Blood tests should include erythrocyte sedimentation rate, hematocrit level, white blood cell count, bacterial cultures, viral titers, heterophile antibody, blood urea nitrogen, T₃, T₄, thyroid-stimulating hormone, antinuclear antibody, rheumatoid factor, and myocardial enzyme levels. The ECG may be typical or nonspecific. The chest x-ray may be normal or may demonstrate an enlarged cardiac silhouette or a pleural effusion. An echocardiogram to evaluate the degree of pericardial effusion is essential. A pericardiocentesis or pericardial biopsy may be necessary when the diagnosis is uncertain.

Treatment

The preferred treatment depends on the underlying cause. Purulent pyogenic pericarditis requires drainage and prolonged intravenous antibiotic therapy. Postpericardiotomy syndrome, post-myocardial infarction syndrome, viral pericarditis, and idiopathic pericarditis often are self-limiting but can require a short course of treatment with nonsteroidal anti-inflammatory agents. If a significant pericardial effusion is present, surgical drainage is indicated if signs of tamponade are present or if resolution is not prompt with anti-inflammatory agents. A 5- to 7-day course of steroids occasionally is necessary. Follow-up studies should be done to document resolution of pericardial effusion or to assess for late constrictive pericarditis.

Chronic Constrictive Pericarditis

Etiology

In the majority of patients, the cause of chronic constrictive pericarditis is unknown and probably is the end stage of an undiagnosed viral pericarditis. Tuberculosis is a rarity. Intensive radiation is a significant cause in some series. Constrictive pericarditis may develop after an open-heart operation. Previous cardiac surgery was reported in one study to be the cause in 39 percent of the patients treated surgically for constrictive pericarditis at NYU.

Pathology and Pathophysiology

The pericardial cavity is obliterated by fusion of the parietal pericardium to the epicardium, forming dense scar tissue that encases and constricts the heart. In chronic cases, areas of calcification develop, adding an additional element of constriction.

The physiologic handicap is limitation of diastolic filling of the ventricles. This results in a decrease in cardiac output from a decrease in stroke volume. The right ventricular diastolic pressure is increased, with a corresponding increase in right atrial and central venous pressure, ranging from 10 to 30 mmHg. The venous hypertension produces hepatomegaly, ascites, peripheral edema, and a generalized increase in blood volume.

Clinical Manifestations

The disease is slowly progressive with increasing ascites and edema. Fatigability and dyspnea on exertion are common, but dyspnea at rest is unusual. The ascites often is severe, and the diagnosis is easily confused with cirrhosis.

Hepatomegaly and ascites often are the most prominent physical abnormalities. Peripheral edema is moderate in some patients, but severe in others. These findings are manifestations of advanced congestive failure from any form of heart disease. With constrictive pericarditis, however, the usual cardiac findings are a heart of normal size without murmurs or abnormal sounds. Atrial fibrillation is present in about one-third of the patients, and a pleural effusion is common in more severe cases. A paradoxical pulse is found in a small proportion of patients.

Laboratory Findings

Venous pressure is elevated, often to 15 to 20 mmHg or higher. The ECG, though not diagnostic, usually is abnormal, with a low voltage and inverted T waves. The chest x-ray usually shows a heart of normal size, but pericardial calcification may be seen in a significant proportion of cases and often is the first clue to the diagnosis. Echocardiogram, MRI, or CT scan may demonstrate a thickened pericardium.

Findings on cardiac catheterization are highly characteristic. There is elevation of the right ventricular diastolic pressure with a change in contour, showing an early filling with a subsequent plateau, the “square root” sign. There also is “equalization” of pressures in the different cardiac chambers, because right atrial pressure, right ventricular diastolic pressure, pulmonary artery diastolic pressure, pulmonary wedge pressure, and left atrial pressure are similar. The one condition that cannot be excluded without myocardial biopsy is a restrictive cardiomyopathy.

Treatment

When the diagnosis has been made, pericardiectomy should be done promptly, because the disease relentlessly progresses. Operation can be done through a sternotomy incision or a long left anterolateral thoracotomy. The constricting pericardium should be removed from all surfaces of the ventricle, mobilizing the heart to where it can be held freely upward in the hand. Removal of the pericardium over the atria and the venae cavae is considered optional, although this usually is done as well. The heart-lung machine is kept on a standby basis in the event of significant hemorrhage. If this occurs, the patient can be heparinized and the blood aspirated and returned to the patient until the laceration is repaired. The pericardium is removed from the pulmonary veins on the right to the pulmonary veins on the left. Both phrenic nerves are mobilized and protected. Particular care is taken to remove pericardium over the pulmonary artery, where residual constriction can seriously impair the operative result.

As the constricting scar develops from organization of an exudate between the pericardium and the epicardium, the plane of dissection may be external to the epicardium, which will greatly decrease operative hemorrhage. If the epicardium is thickened, it must be removed from the underlying myocardium, though this is tedious and results in diffuse bleeding.

Intracardiac pressures should be measured by direct needle puncture before and after pericardiectomy. Often with a complete pericardiectomy the characteristic pressure abnormalities are eliminated or greatly improved. If significant abnormalities remain, the operative field should be carefully checked for any residual sites of constriction. In the past, slow recovery over many months probably was a result of inadequate pericardiectomy, not underlying “ventricular atrophy.”

Results

After a radical pericardiectomy that corrects the hemodynamic abnormalities, patients improve promptly with a massive diuresis. The risk of operation varies with the age of the patient and the severity of the disease; the mortality rate usually is less than 5 percent. A good result can be anticipated for more than 95 percent of the patients.

Culliford reported 62 patients treated surgically with total pericardiectomy. The majority of patients were NYHA class III or class IV preoperatively. Sixteen percent required cardiopulmonary bypass for treatment of associated pathology, usually tricuspid insufficiency. The operative mortality rate was 3 percent. After operation hemodynamic abnormalities promptly corrected, ascites and peripheral edema resolved, and functional status improved dramatically.

ARRHYTHMIA SURGERY

Surgical treatment of cardiac arrhythmias began in 1968 when Sealy first successfully interrupted the bundle of Kent in a patient with Wolff- Parkinson-White (WPW) syndrome. This was possible because of the newly developed electrophysiologic studies and mapping techniques described by Durrer, of Amsterdam, who mapped a patient with WPW syndrome in 1967, demonstrating electrical conduction through an accessory pathway in the area of ventricular preexcitation. Daniel and associates and Kaiser and associates independently reported intraoperative mapping for ischemic ventricular arrhythmias in 1969, but advances in surgical treatment of ventricular tachycardia did not occur until the late 1970s, when Guiraudon described the encircling endocardial ventriculotomy and Josephson and associates proposed map-guided subendocardial resection for ablation of recurrent ventricular tachycardia.

Subsequently, surgical methods for treatment of arrhythmias were advanced by Cox and others. The development of multipoint (160 to 256 channel) computerized cardiac mapping systems greatly improved the speed and efficiency of arrhythmia mapping, and better operative skills and adjuvant cryoablation techniques improved the efficacy of surgical methods. The development of radiofrequency catheter ablation allows treatment of cardiac arrhythmias in the cardiac catheterization laboratory.

Most ablative therapy for arrhythmias, surgical or nonsurgical, is done for one of four entities: (1) WPW syndrome with preexcitation; (2) paroxysmal supraventricular tachycardia (PSVT) from (a) atrioventricular (AV) node reentry or (b) concealed AV pathways; (3) sustained ventricular tachycardia; and (4) chronic atrial fibrillation. Patients with cardiac arrhythmias, with the exception of atrial fibrillation, require electrophysiologic studies before treatment, and additional intraoperative mapping is done at the time of surgery. Catheter ablation methods, when appropriate, can be applied in the catheterization laboratory after electrophysiologic studies are completed.

An epicardial and an endocardial approach have been used for surgical treatment of WPW syndrome and for paroxysmal supraventricular tachycardia, frequently using operative cryoablation as an adjunct. Both operative approaches have been highly successful. The majority of patients with WPW syndrome or PSVT are now treated nonoperatively with less morbidity using radiofrequency catheter ablation techniques. Mapping of the slow posterior pathway has allowed successful catheter treatment of PSVT associated with AV node reentry tachycardia without heart block, and radiofrequency ablation has been successful for WPW syndrome.

Implantable cardioverter defibrillators (ICD) are used for ventricular tachycardia unless it is sustained and unresponsive to medical therapy. A small but significant number ventricular tachycardias can be treated with catheter ablation. These arrhythmias often are life threatening. The ventricular arrhythmia is induced by electrophysiologic studies and attempts are made to suppress the arrhythmia with pharmaceutical agents. If the arrhythmia remains nonsuppressible, surgery is indicated.

Operative map-guided methods have been highly effective for the treatment of sustained ventricular tachycardia associated with left ventricular aneurysms. While results have been encouraging using “blind” (non–map-guided) endocardial resection and remodeling endoaneurysmorrhaphy for treatment of this combination of diseases (see Left Ventricular Aneurysm, above), most reports note lower arrhythmia recurrence rates when mapping methods are used.

Nonsuppressible ischemia-related sustained ventricular tachycardia without aneurysmal disease also may be treated by definitive map-guided operative ablation or by placement of the implantable defibrillator. Coronary bypass is performed concomitantly if reversible myocardial ischemia is present. Although the operative mortality rate is lower with placement of the implantable defibrillator than with open ablative surgery (3 percent versus 8 to 10 percent), the incidence of late sudden death is lower after operative ablation, and the choice of appropriate treatment is controversial. When ventricular fibrillation and not ventricular tachycardia is the primary event, the implantable defibrillator becomes the treatment of choice. Ventricular fibrillation cannot be mapped, and electrophysiologic studies cannot be used to follow the response to medications. Nonoperative radiofrequency ablation has not been effective in the treatment of sustained ventricular tachycardia, and neither operative nor radiofrequency ablation methods have been effective for ventricular fibrillation.

Chronic atrial fibrillation is the most common cardiac arrhythmia. The incidence in the general population is 0.4 to 2 percent, and 10 percent for those older than 60 years of age. Sequelae of chronic atrial fibrillation include: (1) stasis of flow in the left atrium, which may result in thromboembolism; (2) loss of atrioventricular synchrony, which causes a decrease in cardiac output that may be significant; and (3) an irregularly irregular heartbeat that causes discomfort for many patients. Atrial fibrillation is considered *chronic* when the fibrillating atria cannot be converted to and maintained in sinus rhythm or maintained in sinus rhythm by electrical or pharmacologic means.

Medical treatment of chronic atrial fibrillation is focused on ventricular rate control and anticoagulation therapy. Digitalis, beta blockers, and calcium- channel antagonists are first-line drugs for controlling ventricular rate by slowing atrioventricular conduction. Warfarin is used for long-term anticoagulation therapy to prevent thromboembolism. Despite these measures, patients may suffer congestive heart failure from loss of atrioventricular synchrony, thromboembolism from inadequate anticoagulation, bleeding complications from over-anticoagulation, and a sense of anxiety and discomfort from an irregular heartbeat.

Because of the suboptimal results of medical therapy, surgical investigators sought to develop a surgical technique to restore sinus rhythm to chronically fibrillating atria. Using electrophysiologic mapping of the atria, Cox and colleagues determined that atrial fibrillation was caused by macroreentrant circuits of electrical activity that required a large area of atrium in order to form. They also determined that atrial automaticity functioned independently from the macroreentrant circuits. On the basis of these studies they concluded that (1) surgical interruption of the reentrant circuits would abolish the ability of the atria to fibrillate, and (2) automaticity should be preserved to restore sinus rhythm.

Cox then developed a surgical procedure based on strategically placed full- thickness incisions in both atria. Dividing atrial wall and sewing it back together interrupts the macroreentrant circuits and eliminates the atrial fibrillation. With strategically placing the incisions, “corridors” are created for the propagation of depolarization and contraction. The configuration of the corridors results in a single pathway from the SA node to the AV node, which restores AV synchrony. From this “true pathway,” other pathways direct depolarization into “blind alleys” configured to cover nearly all remaining areas of the atria, resulting in widespread atrial contraction. The configuration of a true pathway giving rise to multiple blind pathways resembles a maze. By restoring sinus rhythm and atrial contractility, the maze procedure results in active transport of blood through the atrium and into the ventricles. This improves cardiac output and eliminates the stasis of blood flow, thereby reducing the risk of thromboembolism and eliminating the need for permanent anticoagulation therapy.

The maze procedure has been used for nearly 10 years and has been modified twice. A report on long-term follow-up has been published. The report includes 164 patients followed for at least 3 months (100 percent follow-up) and includes maze-I, maze-II, and maze-III procedures. Sinus rhythm was restored in 100 percent of patients. Recurrence of

atrial fibrillation or flutter occurred in 12 patients (7 percent) but was converted back to sinus rhythm and maintained medically in all 12 patients. At 6 months, atrial function was evaluated in 125 patients using transesophageal echocardiography or dynamic MRI. Right atrial function was preserved in 98 percent of patients, and left atrial function in 86 percent. Follow-up in patients undergoing the maze-III modification revealed right and left atrial function preservation of 98 and 94 percent, respectively. There have been no strokes, and two patients reported transient ischemic attacks. No patients in the series are receiving permanent anticoagulation therapy.

PACEMAKERS/IMPLANTABLE DEFIBRILLATORS/INTRAAORTIC BALLOON PUMP/ASSIST DEVICES

Pacemakers

Indications

Permanent pacemakers initially were used for patients with complete (third- degree) AV heart block that developed spontaneously in older patients or that occurred secondary to operative trauma. Pacemakers have been used for a variety of other arrhythmias, including the sick sinus syndrome and for any second-degree, bifascicular, or trifascicular heart block associated with intermittent bradyarrhythmias or symptoms. Asymptomatic patients with Mobitz II second-degree block or with Mobitz I second-degree block at the level of the bundle of His also are considered by most authorities to be candidates for a pacemaker. The most common indication for permanent pacemaker implantation is sick sinus syndrome, accounting for 50 percent. Temporary pacemakers are used for transient heart block associated with myocardial infarction or cardiac surgery or in symptomatic patients awaiting a permanent pacemaker.

Pathophysiology of Heart Block

In normal cardiac conduction tissue the cardiac impulse originates in the sinoatrial node near the junction of the superior vena cava and the right atrium. The impulse is propagated through the right atrium to the AV node, which lies medial to the ostium of the coronary sinus beneath the septal portion of the tricuspid valve at the apex of the triangle of Koch. From the AV node, conduction travels near the annulus of the tricuspid valve and along the bundle of His to pass through the central fibrous body of the ventricular septum, coursing slightly leftward on the crest of the muscular component of the ventricular septum. Here the bundle divides into the left and right bundles that travel to the left and right ventricles, respectively.

The most common surgical trauma producing complete heart block is valve surgery, but the incidence is the highest during the repair of a ventricular septal defect or an ostium primum atrial septal defect. Complete heart block may rarely follow prosthetic replacement of the aortic, mitral, or tricuspid valves, either from direct injury to conduction tissue or from traction with subsequent fibrosis. Endocarditis can result in complete heart block from an annular abscess. Intrinsic disease may develop in the SA node and the AV node in the aging population, and it can develop as a result of ischemic disease.

Symptoms attributed to first-degree heart block (PR interval > 0.2 s) are rare. With second-degree heart block (intermittent lack of AV conduction), symptoms become apparent if the bradycardia is severe or persistent. Two types of second-degree heart block are described: Mobitz I (Wenckebach) typically has progressive prolongation of the PR interval followed by a nonconducted P wave. This usually is benign unless it occurs at the His- Purkinje conduction level. Mobitz II second-degree block is more ominous, presenting with normal PR intervals, but with sporadic nonconduction or third-degree block. Mobitz II often progresses to complete third-degree block with associated symptoms.

Third-degree (complete) heart block seriously impairs cardiac output in several ways. The resulting bradycardia varies from 25 to 60 beats/min, which may compromise the coronary and the cerebral circulation. Refractory congestive failure may exist with heart rates of 45 beats/min or less. Patients with complete heart block frequently have extreme exercise intolerance, and dangerous symptoms of cerebrovascular insufficiency can occur, such as syncope and convulsions (Adams-Stokes syndrome).

Once complete AV dissociation and third-degree heart block occurs, patients often develop a slow ventricular escape rhythm, typically with heart rates of 25 to 60 beats/min. Cardiac arrest and death can occur. Others may develop ventricular tachycardia or ventricular fibrillation instead of a stable ventricular escape mechanism. Although a few patients remain asymptomatic with a rate as low as 35 beats/min, most have symptoms once the heart rate is less than 45 beats/min. With Mobitz II second- degree block the heart rate may be normal most of the time, but can abruptly change into complete AV disassociation.

Diagnosis

Milder symptoms of light-headedness or near-syncope may be caused intermittent bradycardia or heart block. The differential diagnosis for syncope or near-syncope includes second- or third-degree heart block, sick sinus syndrome, paroxysmal atrial tachycardia, sustained ventricular tachycardia, spontaneous ventricular fibrillation, vasovagal reaction, aortic stenosis, carotid sinus syndrome, occlusive arterial disease of the carotid or cerebral circulation, epilepsy, and idiopathic syncope. The diagnosis of complete heart block can be established with an ECG, but the diagnosis of intermittent heart block, sick sinus syndrome (tachy-brady arrhythmias), or paroxysmal atrial tachycardia requires serial ECGs or a 24-h Holter monitor.

Pacemaker Physiology

The efficacy of cardiac pacing depends on the ability of the heart to respond to short bursts (2 ms) of electrical stimulation, ideally with less than 1.5 to 2 mA of current. There is a direct linear relationship between stimulation threshold and the electrode surface area, with more concentrated charge present in smaller electrodes. Characteristics of modern pacemaker generators include a voltage ranging from 2½ to 7½ V, delivering a current of 3.5 to 15 mA. The pulse width of delivered charge can be adjusted from 0.15 to 2.0 ms. External programming capabilities usually are present.

The pacing threshold determination is particularly important with intravenous implantations to ensure that the pacing lead is in the proper position. With satisfactory location of the intravenous intracardiac electrode determined by fluoroscopy, an optimal stimulation threshold for pacing is 0.4 to 1.0 mA at 0.2 to 0.5 V with a pulse width of 0.5 ms. The resistance of the pacing electrode is calculated from the voltage and current threshold required to achieve cardiac pacing, with 500 to 1,000 ohms considered acceptable.

Cardiac pacing may be done with unipolar or bipolar electrodes. Today most pacing is done with bipolar leads. In a bipolar system the positive and negative electrodes are in contact in the heart. When unipolar pacing is used, the tip of the electrode is the stimulating pole, and the ground is the metal of the generator. This is the most common method, although sensing occasionally is improved by switching to bipolar leads.

Types of Pacemakers

Demand pacemakers are now used almost exclusively. A widely used type of pacemaker is the atrial-ventricular pacemaker, or dual-chamber pacemaker, which requires electrodes in both the atrium and the ventricle. Dual-chamber units allow AV sequential pacing (Fig. 18-21) by sensing both chambers and pacing as necessary. Newer pacemakers have rate-responsive capabilities, increasing their rate in response to the patient's bodily activity. Almost all pacemakers are now externally programmable.

A standardized code to describe pacemakers has been developed. It usually includes three to five letters that describe the functions of the pacemaker. The first letter indicates the chamber(s) paced (V = ventricle, A = atrium, D = dual-chamber pacing). The second letter indicates the chamber(s) sensed (V = ventricle, A = atrium, D = dual-chamber sensing, 0 = none). The third letter describes the response of the pacemaker to a sensed beat (T = triggered, I = inhibited, D = dual-chamber inhibition, 0 = not applicable). For example, a DDD pacemaker has dual-chamber pacing, sensing, and inhibition. Occasionally a fourth letter is used to indicate programmability or rate modulation (P = simple programmable, M = multiprogrammable, C = communicating, R = rate modulation), and a fifth letter may be used if antitachycardia or defibrillation functions are available (P = antitachycardia pacing, S = shock, D = dual function). These later designs often are combined in the automatic implantable defibrillator.

Operative Technique

For transient heart block after myocardial infarction a temporary pacing electrode is percutaneously introduced through the subclavian or femoral vein and advanced into the right ventricle. Fluoroscopy may be helpful but usually is not necessary. Temporary pacemaker wires are also commonly placed on the epicardial surface of the heart after cardiac surgical procedures. Temporary transvenous or epicardial pacemaker wires are easily removed before hospital discharge.

Most permanent pacemakers are inserted by the transvenous route. The cephalic vein, the external jugular vein, or the internal jugular vein can be used, but more commonly a direct sterile subclavian puncture is used for venous access. When the needle is in the subclavian vein, a dilator and sheath are passed over a guide-wire, and the pacing electrode is advanced through the sheath. The transvenous electrode can be advanced into the appropriate chamber of the heart, guided by fluoroscopic control. Pacing threshold and P-wave or R-wave sensitivity are determined. Once the electrode or electrodes have been positioned, the pacemaker is implanted in a subcutaneous pocket on the anterior chest wall.

Open surgical approaches use one of three routes: left intercostal, left subcostal, or subxiphoid incision. Because the morbidity is greater with the placement of epicardial leads, this method is reserved for special circumstances, such as in patients with prosthetic tricuspid valves or in neonates.

Postoperative Considerations

The morbidity with transvenous pacemaker implantation is small. Migration of the pacing electrode is the most common complication. Perforation of the ventricle can occur, resulting in diaphragmatic pacing or even cardiac tamponade. Infection is surprisingly rare. Recovery after pacemaker placement usually is uneventful. Antibiotics are given for 24 h, but there are no data to support their use. Difficulties with the subcutaneous implantation of the generator are rare with the small generators now available.

After discharge from the hospital, the patient should be seen in 2 to 3 weeks to assess for a rise in the pacing threshold. Telephonic electrocardiographic surveillance has been a significant advance. Data indicate that this type of periodic monitoring is mandatory for good care. As the battery is wearing down, the pacemaker's output changes, switching to a fixed mode; this can be determined by telephonic monitoring. When generator failure begins to occur, the pacemaker is electively replaced. Continuous follow-up is essential.

Special problems may occur in pacemaker patients requiring other surgical procedures in which the electrocautery is needed. Electromagnetic fields as high as 60 V/m are commonly induced with the electrocautery. This far exceeds the energy level required to activate the sensing circuit in most pacemakers. As a result this electrical signal may inhibit the pulse generator, damage the circuitry, or reset the pacing programs. A variety of arrhythmias have been observed as a result of subsequent pacemaker malfunction, such as ventricular asystole, multiple premature ventricular contractions, ventricular tachycardia, and ventricular fibrillation. Total destruction of some pacemakers has occurred. In patients for whom use of the electrocautery is necessary, brief bursts of cautery should be used, keeping as far from the generator as possible. A bipolar cautery probably is safest. A pacemaker programming system should be available, and the pacemaker should be interrogated and reprogrammed after use of electrocautery. For patients who are totally pacemaker dependent, a backup transvenous wire and external pacemaker should be available before the electrocautery is used.

Implantable Defibrillators

More than 500,000 deaths occur in the United States each year from sudden cardiac arrest; the majority of these are a consequence of ventricular fibrillation and ventricular

tachycardia. Less than 10 percent of these deaths occur in the setting of acute myocardial infarction. The survival rate for an out-of-hospital cardiac arrest not associated with myocardial infarction is less than 25 percent.

Spontaneous ventricular tachycardia and ventricular fibrillation commonly are a result of underlying structural heart disease, such as coronary artery disease, ventricular aneurysm, or cardiomyopathy. Long-term survival in these patients is poor and their risk for sudden death can be stratified by ejection fraction and electrophysiologic studies. For patients with inducible ventricular arrhythmias that are not suppressed with medication, the 2-year survival rate is less than 40 percent. These patients are candidates for arrhythmia surgery or for placement of the internal cardioverter defibrillator.

Internal Cardioverter Defibrillator (ICD) Devices

The early development of the implantable defibrillator was achieved by Mirowski, with the first human implantation being performed in 1980. In early 1982 a second-generation device, the automatic internal cardioverter defibrillator, became available. Third- and fourth-generation ICD devices are in clinical trials (Fig. 18-22). These units allow programmability, backup bradycardia pacing, antitachycardia pacing, and standard cardioversion. Until recently ICD placement required open implantation, but totally transvenous systems are available now.

Indications

The indications for ICD therapy are similar to those for arrhythmia surgery:

- (1) One or more episodes of spontaneous sustained ventricular tachycardia (VT) or ventricular fibrillation (VF) in a patient in whom electrophysiologic testing or spontaneous ventricular arrhythmias cannot be used accurately to predict the efficacy of other therapies.
- (2) Recurrent episodes of spontaneous sustained VT or VF in a patient despite antiarrhythmia drug therapy (guided by electrophysiologic testing or noninvasive methods).
- (3) Spontaneous sustained VT or VF in a patient in whom antiarrhythmia drug therapy is limited by intolerance or noncompliance.
- (4) Persistent inducibility of clinically relevant sustained VT or VF on electrophysiologic studies on the best available drug therapy or despite surgical/catheter ablation in a patient with spontaneous sustained VT or VF.

Results

Several reports have documented the efficacy of ICD therapy. In a large series by Winkle the rate of freedom from sudden death after 5 years with use of the ICD was 96 percent. Although the risk of sudden cardiac death is substantially reduced by ICD therapy, overall survival rates are eroded by congestive heart failure, with a 5-year survival rate of 70 to 75 percent. Similar results are obtainable with arrhythmia surgery, and the appropriate treatment of these patients is controversial.

Intraaortic Balloon Pump

The intraaortic balloon pump (IABP) is the most common and effective technique for assisted circulation. Approximately 70,000 are inserted annually in the United States. The technique was developed by Alstin and associates. A balloon catheter is inserted through the femoral artery and advanced into the thoracic aorta. Insertion of the balloon may be done by direct arteriotomy or with a percutaneous insertion kit. With electronic synchronization, the balloon is alternately inflated during diastole and deflated during systole. Coronary blood flow is increased by improved diastolic perfusion, and afterload is reduced. The cardiac index typically improves after insertion, and the preload decreases. Total myocardial oxygen consumption is diminished by approximately 15 percent.

The IABP can be used for several days or weeks with minimal morbidity. Ischemia of the extremity through which the IABP is inserted is the most serious complication, and the extremity must be examined frequently for viability. Platelet consumption also may occur.

Assist Devices

Temporary assisted circulation is a valuable clinical modality in the treatment of transient cardiac injury. The most common indication for temporary assisted circulation is cardiac failure after cardiac surgery, manifested by a cardiac index of less than 1.5 L/min/m² despite inotropic support. Often an IABP has been used without success. In such instances, mechanical cardiac assistance for 24 to 48 h or longer may permit recovery of cardiac function, probably by minimizing reperfusion injury and from resolution of myocardial edema. Some patients who are dying while awaiting heart transplantation may benefit from circulatory support with a cardiac assist device.

Several types of ventricular assist devices (VAD) are available, including temporary extracorporeal centrifugal pumps, implantable left ventricular assist systems (LVAS), external heterotopic pulsatile ventricular assist devices, and total artificial hearts (TAH). The most commonly used devices are simple pumps (DeBakey roller pump or BioMedicus centrifugal pump) connected to the left atrium and the aorta. These devices can provide effective support for 5 to 7 days, and, in rare cases, up to 2 weeks. External pulsatile assist devices deliver blood flow in synchrony with the native heart and are used for short-term support after cardiac surgery or primarily as a bridge to cardiac transplantation. Attempts at total cardiac replacement with an artificial heart, such as the Symbion Jarvik-7 TAH, have been largely unsuccessful because of long-term thromboembolic complications, infection, and trauma to blood elements.

Results

At NYU over 100 temporary ventricular assist devices have been used for postcardiotomy shock in the past 15 years, with an overall survival rate of approximately 30 percent. Prompt insertion is necessary before irreversible myocardial damage has occurred. The authors currently use the centrifugal pump with heparin-bonded tubing for short-term cardiac assistance. Similar results were reported by Pae and associates from a combined registry experience of 965 cases, with a hospital survival rate of 25 percent. Eighty-six percent of the surviving patients were subsequently NYHA functional class I or class II. The survival rate was equivalent to those obtained with the use of pulsatile pumps or nonpulsatile centrifugal devices. All ventricular assist devices are associated with a significant risk of thromboembolism and infection.

Data are available from the registry of the International Society of Heart Transplantation on the use of ventricular assist devices in 291 patients as a bridge to cardiac transplantation, 213 patients (73 percent) of whom survived transplantation. Pretransplant cardiac support ranged from 19 to 159 days. Data from St. Louis University indicate a survival rate of 48 percent when assist devices were used as a bridge to transplantation. All patients undergoing cardiac support, either after cardiac surgery or as a bridge to transplantation, would have died had ventricular assist devices not been available. With steady improvements in technology, long-term cardiac assistance or totally artificial hearts may have an expanded role in the future.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 19 - Thoracic Aneurysms and Aortic Dissection

Aubrey C. Galloway
Jeffrey S. Miller
Frank C. Spencer
Stephen B. Colvin

THORACIC ANEURYSMS

General Considerations

Modern surgical treatment of arterial aneurysms was introduced nearly a century ago by Rudolf Matas, who described a method of internal repair of aneurysms termed *reconstructive endoaneurysmorrhaphy*. Nearly fifty years later, in 1952, notable work by Cooley and DeBakey demonstrated the feasibility of repairing aneurysms involving the thoracic aorta. Subsequently, *excisional therapy* became the mainstay of the modern surgical approach to aneurysm disease, with resection of the aneurysmal aortic segment followed by restoration of blood flow through the placement of an interposition Dacron graft.

More recently excisional therapy and endoaneurysmorrhaphy with internal graft placement have been used for repair of thoracic aortic aneurysms. When the aneurysm is well localized, total excision of the diseased aorta usually is feasible, and the involved area is replaced with an end-to-end Dacron graft (*graft interposition technique*). Many surgeons prefer this technique for aneurysms involving the aortic root or the ascending aorta, because the excisional method is associated with a lower risk of subsequent reoperation for pseudoaneurysm. By contrast, in patients with more extensive aneurysmal disease, i.e., aneurysms that involve a long segment of aorta or invade adjacent structures, excisional therapy is unnecessary and may be hazardous. In these patients a simpler and equally effective method is replacement of the diseased segment of aorta from within by placing a Dacron graft, without excising the aorta itself. An end-to-end anastomosis between the graft and the aorta is performed, working from within the open aneurysm. The wall of the aneurysm subsequently is wrapped around the graft for tissue coverage. This technique, termed the *graft inclusion method*, has the advantage of limited operative dissection and is associated with less risk of bleeding or injury to adjacent structures.

Etiology and Pathogenesis

An aortic aneurysm can be defined as a localized or diffuse aortic dilatation, usually exceeding 5 to 6 cm in diameter. Aneurysms develop from a weakness or defect in the aortic wall, which has a tendency to dilate progressively. This is to some extent a self-perpetuating process, because the lateral wall tension of a vascular tube is related to the radius by Laplace's law (tension = pressure × radius). The larger the aneurysm, the greater the wall tension. Past concepts of the pathogenesis of aneurysm formation were oversimplified and inaccurate. Aneurysms do not form simply from tension related to passive dilation but are a result of a complex remodeling process involving the vessel wall, as was recently outlined in a review by Halloran and Baxter. This remodeling process involves a change in the structure of collagen and elastin within the vessel wall with an increased proteolytic enzyme activity that changes the balance between proteases and protease inhibitors (Table 19-1). Serine protease, metalloprotease, and neutrophil elastase activity are increased in the aneurysmal vessel, resulting in a net degradation of elastin and collagen. Protease inhibitor activity is simultaneously decreased, leading to an additional imbalance in the ratio of protease to protease inhibitor. These biochemical changes produce progressive dilation of the vessel wall. Inflammatory mediators also are activated, especially in patients with atherosclerotic, infectious, or autoimmune diseases, additionally enhancing the proteolytic process. The net effect is vascular remodeling with dilation of the vessel wall. Initially there is a loss of elastin in the media with compensatory thickening of the adventitia, which maintains vessel wall integrity despite increased wall tension. Later in the process, increased collagenase activity leads to degradation of the adventitial layer with additional dilation or rupture of the aneurysm.

The main causative factors associated with aortic aneurysms are age, hypertension, smoking, atherosclerosis, aortic dissection, and connective tissue disorders. Although atherosclerosis has long been thought to be a primary cause of aneurysmal formation, this may be an oversimplification. While atherosclerosis is associated with aneurysmal disease, the presence of atherosclerosis alone probably is not sufficient for an aneurysm to develop. Chronic atherosclerosis results more often in a proliferative, obstructive vascular disease, with aneurysmal dilatation occurring in a minority of individuals. The available data suggest that other factors, such as inflammation or a genetic tendency for increased proteolytic activity, also must be present for an aneurysm to develop in atherosclerotic patients.

Aortic dissection may result in acute or chronic aneurysmal dilatation. This is discussed in detail later in this chapter. The most common connective tissue disorder associated with aneurysm formation is Erdheim's cystic medial necrosis, in which the underlying defect is idiopathic degeneration of the media with progressive aneurysmal dilation. Marfan syndrome and Ehlers-Danlos syndrome, both with identifiable genetic defects, are well-defined connective tissue diseases associated with aortic aneurysms. In Marfan syndrome, the defect is in the gene coding for fibrillin, which is a major structural component of the microfibrils in connective tissue. Marfan syndrome patients have a well-described phenotype that includes a tall stature, high palate, joint hypermobility, lens disorders, mitral valve prolapse, and aneurysms involving the aortic root or other parts of the aorta. Patients with Marfan syndrome have a significantly increased risk of aortic valvular insufficiency, aortic rupture, and aortic dissection. A similar but less common defect is Ehlers-Danlos syndrome, which is a consequence of a group of genetic mutations leading to defective collagen synthesis, resulting in aneurysm formation in multiple locations throughout the body. Similar to Marfan patients, these patients develop aneurysms of the aortic root with an increased risk of spontaneous aortic rupture. The surgeon should be familiar with the various genetic defects associated with aneurysmal disease, because identification of a genetic disorder can lead to earlier and more effective operative treatment of these high-risk patients.

Less common causes of thoracic aneurysms include trauma, infection (syphilitic or other mycotic), inflammatory diseases (granulomatous and Takayasu's arteritis), and autoimmune diseases.

Classification

Aneurysms of the thoracic aorta are best classified in terms of anatomic location; the clinical significance and the surgical approach vary widely with the location of the aneurysm. The four major locations are (1) the aortic root and ascending aorta, (2) the transverse aortic arch, (3) the descending thoracic aorta, and (4) the thoracoabdominal aorta. Aneurysms of the aortic root or ascending aorta are the most common, representing over 40 percent of thoracic aneurysms in a large series. The descending aorta is involved in about 35 percent of thoracic aneurysms, and the transverse arch and the thoracoabdominal aorta are involved in 10 to 15 percent each. Localizing aneurysms to a specific location may be misleading, because the entire aorta may be diseased and the most obvious location might be only a portion of a more diffuse disease process.

Other descriptive terms have been helpful in identifying specific areas of aortic involvement. *Aortoannular ectasia* is a descriptive term for a degenerative dilatation of the aortic annulus and the sinuses of Valsalva, producing aortic insufficiency and a localized aneurysm involving the aortic root. This lesion often is associated with connective tissue diseases such as Marfan syndrome or Ehlers-Danlos syndrome. *Traumatic aneurysms* may occur after blunt trauma to the chest. They are located just distal to the subclavian artery at the site of insertion of the ligamentum arteriosum. While the management of acute traumatic aortic transection is discussed in Chap 6 (Trauma), the surgical management of elective repair of a traumatic aneurysm is similar to that of all other descending thoracic aneurysms. The term *dissecting aneurysm* is a misnomer, usually meaning *aortic dissection*, which results from an intimal tear in the aortic wall with subsequent disruption of the media, producing a true and a false lumen throughout the dissected aorta. When a dissected aorta dilates to over 5 to 6 cm in diameter, it is appropriately termed an *aortic aneurysm secondary to aortic dissection*. These may be acute or chronic.

Clinical Manifestations

The classic symptom associated with a large or an expanding aneurysm is pain, which may be excruciating and severe. With chronic aneurysms the symptoms may be more subtle, such as chronic pressure or a low-grade aching pain. Because symptomatic patients have an increased risk of rupture, pain possibly related to aneurysmal expansion must be thoroughly investigated. New onset of pain in a patient with a known aneurysm is highly significant and may indicate rapid expansion, leakage, or impending rupture.

Large aneurysms may produce symptoms from compression of adjacent structures, such as the trachea, mainstem bronchus, superior vena cava, pulmonary artery, recurrent laryngeal nerve, or vertebral bodies. Syphilitic aneurysms, seen less now than in the past, were known for their tendency to invade bone, producing back pain from erosion of the thoracic spine, which occurs less frequently with other aneurysms. The most common symptoms in patients with thoracic aneurysms are, in decreasing order of frequency, pain, pulmonary symptoms, and hoarseness.

The majority of patients with moderate-sized thoracic aneurysms are asymptomatic unless significant enlargement has occurred. These aneurysms usually are discovered incidentally in a routine chest radiograph or during catheterization or imaging studies performed for other reasons. Usually there are no physical abnormalities or hemodynamic disturbances, except in aneurysms of the aortic root, which may be associated with aortic valvular insufficiency.

Diagnostic Studies

Once an abnormal shadow has been identified on chest x-ray, an attempt to establish the diagnosis of an aneurysm is made with magnetic resonance imaging (MRI), computed tomography (CT), or echocardiography. Aortography, previously the primary diagnostic tool, is seldom necessary because of newer imaging techniques. The full extent of aneurysmal involvement can be well demonstrated with MRI, CT, or MR angiography. Echocardiography is used to determine the size of the ascending aorta and aortic root and is especially helpful in detecting any associated valvular insufficiency. When surgery involving the aortic root is necessary, cardiac catheterization usually is indicated to additionally define the degree of aortoannular ectasia and the degree of coronary artery displacement. An aortic root angiogram with runoff images of the aortic arch is recommended, evaluating the involvement of the sinuses of Valsalva, the degree of aortic insufficiency, the amount of coronary artery displacement, and the degree of aneurysmal involvement of the aortic arch distally.

Patients with thoracic aneurysms have a high incidence of associated coronary artery disease. These patients should have screening studies or coronary angiography before aneurysm repair.

Natural History and Operative Indications

The natural history of aortic aneurysms is one of progressive enlargement with eventual rupture. Factors related to the risk of rupture include aneurysmal size, change in size, age of the patient, pain, symptoms of aneurysmal expansion, smoking, and chronic obstructive pulmonary disease (COPD). Juvonen and associates analyzed multiple risk factors to develop a multivariate equation to calculate the probability of rupture for thoracic aneurysms. Patients with large aneurysms (more than 6 to 7 cm in diameter) with rapid aneurysm expansion or with pain generally underwent elective repair, while others with aneurysms of moderate size were monitored with serial CT scans over approximately 5 years. The risk of rupture for aneurysms of various sizes was as follows: less than 3 cm, 0 percent; 3–4 cm, 6 percent; 4–5 cm, 12 percent; 5–6 cm, 36 percent; 6–7 cm, 50 percent; and 7–8 cm, 100 percent. Age and COPD were factors other than aneurysm size that independently increased the risk of rupture. The approximate risk of rupture for an individual patient with a thoracic aneurysm can be estimated on the basis of clinical variables and compared with the estimated operative risk. Although the majority of aortic aneurysms steadily enlarge and elective surgery can be planned, sudden rupture occurs in a small proportion of patients with smaller aneurysms that were not progressively enlarging. Patients treated nonoperatively should be warned of this possibility.

Other clinical trials have demonstrated a similar risk of rupture for thoracic aneurysms. In a large series reported by Bickerstaff and associates in which patients were treated nonoperatively, the 1-year survival rate was 60 percent, and the 5-year survival rate was 13 percent. Over 70 percent of the deaths were caused by rupture of the aneurysm. Pressler and McNamara reported observations in 260 patients with thoracic aneurysms, 126 of whom were treated surgically. The 5-year survival rate in the patients treated nonoperatively was only 21 percent, with high rupture rates noted for atherosclerotic and dissecting aneurysms. Similarly, Finkelmeier demonstrated that 41 percent of patients with chronic traumatic thoracic aneurysms died or developed symptoms from the aneurysm within 5 years.

Data support elective operative treatment for most thoracic aneurysms larger than 5 to 6 cm in diameter, because of the significant risk of rupture and death within 5 years. Operation should be strongly considered in patients with saccular aneurysms or connective tissue diseases once the aneurysm exceeds 4.5 to 5 cm, depending on the estimated operative risk, because the rupture rate is higher in these patients.

Aortic Root and Ascending Aortic Aneurysms

General Considerations

Aneurysms localized to the aortic root are often a result of cystic medial necrosis or connective tissue disease involving the aortic wall. Atherosclerotic aneurysms occur in the ascending aorta but are seldom limited to this area. Aortic dissection, inflammatory diseases (syphilis, autoimmune diseases, and aortitis), and congenital defects (bicuspid aortic valve and aortic coarctation) are other causes of ascending aortic aneurysms. In a series of 165 ascending aortic and aortic arch aneurysms reported from New York University (NYU), 29 percent were from aortic dissection, 22 percent from atherosclerosis, 22 percent from cystic medial necrosis or connective tissue diseases, and 27 percent from other causes.

Aneurysms involving the ascending aorta (Fig. 19-1) may be isolated to the supracoronary position, may extend distally to involve the aortic arch, or may extend proximally with involvement of the aortic root and sinuses of Valsalva. With aneurysms originating in the supracoronary part of the ascending aorta (i.e., not involving the sinuses of Valsalva), the aortic valve usually is normal and uninvolved, unless separate pathology such as congenital aortic stenosis is present. In contrast, aneurysms originating in the aortic root by definition involve the sinuses of Valsalva, causing distortion of the sinotubular ridge (the point where the commissures of the aortic valve attach to the ascending aorta). The process of aortoannular ectasia stretches and distorts the valvular commissures and cusps, producing central aortic insufficiency. In these patients the annulus, the sinuses of Valsalva, and the sinotubular junction are all dilated, and the coronary ostia usually are displaced superiorly by 2 to 3 cm. In some patients aortic insufficiency is severe, and cardiac failure may be the initial clinical finding. Chronic ascending aneurysms from aortic dissection often produce a similar finding, because the dissection process may distort the commissural attachments of the aortic valve at the sinotubular ridge, producing aortic insufficiency.

Before surgical therapy was available, most patients with Marfan syndrome died in the third decade of life from aortic dissections or rupture. Early surgical intervention for aortic root aneurysms is therefore recommended in patients with Marfan syndrome or Ehlers-Danlos syndrome, usually when the diameter of the aortic root exceeds 4.5 to 5 cm.

Operative Treatment

The standard operative approach for aneurysms involving the aortic root or ascending aorta is through a median sternotomy incision, using cardiopulmonary bypass and cardioplegia. When the aneurysm is confined to the ascending aorta without involvement of the aortic root, the ascending aorta is replaced with a woven Dacron graft, beginning just distal to the sinotubular ridge and ending proximal to the innominate artery (Fig. 19-2). Concomitant valve replacement is performed if aortic valve disease is present.

Replacement of the aortic root is performed when the aortic root is aneurysmal (aortoannular ectasia). The operation most commonly used for root replacement is the composite valve graft procedure, which involves replacement of the aortic valve and aortic root with an aortic valve–Dacron graft conduit, placed from the aortic annulus to the distal aorta beyond the aneurysm. This technique necessitates reimplantation of the coronary arteries into the composite graft. The need for aortic root replacement often can be determined preoperatively by MRI, CT, or echocardiographic studies, or by an angiogram of the aortic root during cardiac catheterization. The final decision often is made at operation; if the sinuses of Valsalva and the aortic annulus are dilated and the ostia of the coronary arteries are displaced more than 2 cm from the annulus, then the aortic root should be replaced.

The technique for aortic root replacement initially was described by Bentall and Debono. With the Bentall method, the composite valve graft is placed from within the aorta without cutting out the aneurysm, reimplanting the coronary arteries into the graft. The aneurysmal sac is closed around the composite graft as part of a graft inclusion technique (Fig. 19-3). This classic Bentall method is used less frequently today because of a significant incidence of late false aneurysm formation within the aneurysmal wrap. The most widely used technique for root replacement is a modification of the Bentall procedure in which the aortic root is totally excised, leaving “buttons” of aortic tissue around each coronary ostium for subsequent direct reimplantation into the composite graft (Fig. 19-4). Cabrol described an alternative method for coronary artery reimplantation during aortic root replacement using a Dacron tube 10 mm in diameter to connect the left and right coronary arteries, subsequently performing a side-to-side anastomosis between the small tube graft and the composite valve graft to reestablish coronary flow (Fig. 19-5). Aortic root replacement also can be performed using a cryopreserved homograft or a pulmonary autograft.

Results

Aortic root replacement with a composite graft can be safely performed, with excellent long-term results. The operative mortality rate is 2 to 5 percent, depending on the number of associated risk factors, such as coronary artery disease, reoperation, New York Heart Association (NYHA) class IV functional status, Marfan or Ehlers-Danlos syndrome, aneurysm rupture, or superimposed acute aortic dissection. The operative risk is less for isolated ascending aneurysm repair without root involvement. Because the operative risk is markedly increased once rupture or dissection has occurred, elective operation should be recommended whenever an aneurysm larger than 4.5 to 5.5 cm in diameter is identified.

Aortic Arch Aneurysms

Aortic arch aneurysms may be isolated or may be part of a continuous aneurysmal process involving the ascending and descending aorta. The most common causes are atherosclerosis, aortic dissection, and connective tissue disorders. The diagnosis usually is suspected after identification of an abnormality on a chest radiograph and confirmed by MRI, CT, or aortography. The innominate, carotid, or subclavian arteries also may be aneurysmal. The proximal and distal extent of disease and the degree of great-vessel involvement should be clearly defined preoperatively, because these characteristics strongly influence the operative approach.

Treatment

The operative procedure for arch replacement is complex, involving issues of surgical technique and cerebral protection. Depending on the extent of arch involvement, it may be necessary to perform *total arch replacement* (Fig. 19-6) with reimplantation of the arch vessels into the graft, or *partial hemi-arch replacement* (Fig. 19-7), with placement of a beveled graft leaving the arch vessels to arise from the native aortic arch superiorly. The hemi-arch repair has been used with increasing frequency at NYU over the past several years. This technique, initially described by Cooley, is simpler to perform and allows the surgeon the freedom to tailor the repair as necessary, replacing anywhere from 25 to 85 percent of the aortic arch. A third technique of arch replacement, the *elephant trunk technique* described by Borst and associates, is used in a small number of patients who will require a subsequent operation for a descending aneurysm. The elephant trunk technique is similar to standard total arch replacement, but the distal graft is invaginated into itself while the anastomosis is constructed, and the invaginated portion is subsequently unfolded so that it lies free in the descending aorta. This allows the surgeon easier access to the distal graft, or elephant trunk, during the subsequent descending aneurysm repair, which usually is performed several weeks later via a left thoracotomy.

In various large clinical trials reported by Coselli, Svensson, Ergin, and Galloway the overall operative risk for arch aneurysm surgery has ranged from 6 to 15 percent, depending on the number of risk factors present. The risk factors for poor outcome have been identified as advanced age, emergency surgery, acute dissection, aneurysm rupture, cardiac tamponade, Marfan syndrome or other connective tissue diseases, previous aortic surgery, and presence of a second aneurysm in the descending thoracic aorta. Other operative variables influencing survival include cardiopulmonary bypass time, operative bleeding, renal failure, and stroke.

Before the introduction of deep hypothermia and circulatory arrest, arch aneurysm repair had one of the highest operative risks of any surgical procedure, with mortality rates exceeding 75 percent. The demonstration that the brain could safely tolerate circulatory arrest for periods of up to 45 min if the temperature was carefully lowered to 15 to 17°C formed the basis for an improved surgical approach. This technique, termed deep hypothermia and circulatory arrest, was first applied by Griep and associates in 1975. The hypothermic circulatory arrest technique involves core cooling of the blood to 10 to 15°C while the head is packed in ice and the body is cooled externally with a cooling blanket. The patient is cooled until the tympanic membrane temperature is 15 to 17°C or until electroencephalogram (EEG) silence is achieved, after which cardiopulmonary bypass is stopped. Use of hypothermic circulatory arrest for cerebral protection allows the surgeon to stop all blood flow during arch repair so that a precise, sound technical arch anastomosis can be performed. The technique has the added advantage of avoiding clamping and manipulation of the diseased aorta, which lessens the risk of aortic injury or embolization. The circulatory arrest technique has resulted in a dramatically lower mortality rate.

The incidence of permanent neurologic injury (stroke) after circulatory arrest is low if the cerebral ischemic time does not exceed 45 min. A report by Galloway in 1989 described repair of 165 arch aneurysms over 10 years using hypothermia and circulatory arrest in the majority of cases; 85 percent of the arch repairs had circulatory arrest times of less than 45 min. The operative mortality rate was approximately 10 percent, and the frequency of stroke was less than 2 percent.

Ergin and associates analyzed neurologic outcome in patients undergoing hypothermic circulatory arrest for arch repair and demonstrated a correlation between the risk of temporary neurologic dysfunction (transient confusion, agitation, or obtundation) and the circulatory arrest interval. Temporary neurologic dysfunction occurred in 20 to 30 percent of the patients in whom the circulatory arrest interval had exceeded 50 min. The most common cause of permanent neurologic injury in the patients was embolic events, not ischemia from the circulatory arrest.

The incidence of neurologic injury after aortic arch repair varies from 0 to 15 percent in different series. The risk factors for stroke are age, complexity of the arch repair, circulatory arrest time, and the amount of clot or atheromatous debris in the arch.

A new technique of cerebral protection involving retrograde cerebral perfusion (through the superior vena cava) has been used to minimize the risk of cerebral ischemia during arch aneurysm surgery. Retrograde perfusion uses venous perfusion of the brain with blood that has been cooled to 7 to 12°C. Flow rates usually are 200 to 500 mL/min during the period of systemic circulatory arrest, maintaining a cerebral venous perfusion pressure at 30 to 40 mmHg. Reports by Bavaria, Ergin, Lytle, and Coselli have shown a diminished risk of stroke with the use of retrograde cerebral perfusion. The technique is of most importance in cases in which the circulatory arrest time is expected to exceed 45 min.

Descending Thoracic Aortic Aneurysms

General Considerations

Descending aneurysms may result from atherosclerosis, cystic medial necrosis, connective tissue diseases, aortic dissection, infection, inflammation, or prior trauma.

Atherosclerosis and degenerative diseases probably are the most common causes of descending aneurysms. The majority of aneurysms begin in the proximal descending aorta and extend for varying distances, often involving the entire descending thoracic aorta. Atherosclerotic aneurysms generally are fusiform, but some are localized and saccular. Concomitant abdominal aortic aneurysms occur in 25 to 30 percent of the patients, and the entire thoracoabdominal aorta may be involved in 10 percent. The aneurysmal disease may extend proximally to involve the aortic arch. Traumatic aneurysms arise from prior aortic transection after blunt trauma to the chest, developing in those few patients who survived the initial injury. While acute traumatic aortic transection usually follows trauma, a chronic aneurysm can develop after acute traumatic aortic injury, usually occurring immediately distal to the left subclavian artery, where the aorta is fixed by the ligamentum arteriosum. A chronic traumatic aneurysm of the descending aorta generally is repaired with a prosthetic graft and reconstruction techniques identical to those used for other forms of chronic descending aortic aneurysms.

Clinical Manifestations

Most patients with descending aortic aneurysms are asymptomatic, and the diagnosis is made after a chest radiograph, usually performed for other reasons. Symptoms, when present, usually result from aneurysmal enlargement or compression of adjacent structures. Large aneurysms may compress the left main bronchus, resulting in cough and dyspnea, or erode into the lung parenchyma, producing hemoptysis. Enlarging aneurysms near the left recurrent laryngeal nerve may produce vocal cord paralysis and hoarseness. Pain is the most common and concerning symptom, because it usually suggests significant compression of an adjacent vital structure or aneurysmal expansion with impending rupture.

The physical examination usually is normal. Rarely a bruit is audible in the left paravertebral area. Peripheral pulses usually are normal, although a large aneurysm compressing the left subclavian artery may result in diminished pulses in the left arm. A chronic dissection may shear off the origin of the subclavian or iliac vessels, resulting in diminished or unequal pulses in the arms or legs.

Diagnostic Studies

The diagnosis of a thoracic aneurysm usually is suspected after identification of a mass in the region of the descending aorta on a chest radiograph. The differential diagnosis includes bronchogenic carcinoma, metastatic carcinoma, or mediastinal tumors. Laminar calcifications may be visible in the wall of the aorta.

The diagnosis is confirmed by standard CT or MRI (Fig. 19-8), which readily demonstrate the size and extent of aortic aneurysmal involvement. These studies also are especially useful to evaluate aneurysms periodically for progressive enlargement so that the proper timing of surgery can be determined. Aortography may be used to confirm the diagnosis and further delineate the precise extent of the aneurysmal involvement before surgery. Today standard aortography is seldom necessary, however, because newer forms of enhanced MR or CT angiography are available. These imaging studies allow three-dimensional reconstruction of the aorta with scan times of less than 3 min and have largely replaced conventional aortography in the evaluation of thoracic aortic aneurysms.

Because concomitant atherosclerosis often is present in the coronary, renal, or carotid arteries, a thorough preoperative evaluation should be undertaken. Provocative stress testing, coronary arteriography, and carotid studies should be performed as indicated to determine whether concomitant cardiovascular disease requires treatment before thoracic aneurysm repair. Patients with symptomatic coronary artery disease or with a positive stress test should undergo cardiac catheterization and revascularization before aneurysm repair.

Operative Indications

In patients with descending aneurysms larger than 5 to 6 cm in diameter elective operative repair is recommended. This is especially important because of the high risk of subsequent rupture in these patients and because emergency surgery has mortality rates four to five times higher than elective surgery. In patients with aneurysms less than 5 cm in diameter, a policy of observation may be used, but these patients require frequent follow-up imaging studies, and operation is indicated if the aneurysm expands or if the patient becomes symptomatic.

Operative Technique

A variety of techniques have been used for repair of descending aortic aneurysms, with technical advances resulting in a marked reduction in the operative mortality over the past 10 to 15 years. Two major operative techniques are in use. *Unprotected cross-clamping* is performed with cross-clamps placed proximally and distally without distal perfusion, or with a single proximal cross-clamp with controlled distal exsanguination. In *perfusion or shunting techniques* bypass or passive shunts are used to maintain distal aortic perfusion during the cross-clamp time. Perfusion may be done with left atriofemoral bypass, with femoral–femoral bypass with an oxygenator, or with a Gott shunt from the proximal to the distal aorta. Perfusion and shunting techniques were introduced to limit spinal cord ischemia during the cross-clamp time.

Operative exposure is achieved with a left posterolateral thoracotomy through an appropriate interspace, usually the fourth, fifth, or sixth. For most mid–descending aortic aneurysms a fifth interspace incision is used, although the sixth rib may be resected if more extensive exposure is required. Initially the aorta is mobilized and encircled proximal and distal to the aneurysm. In some cases the aorta is controlled proximally between the left carotid and left subclavian arteries; approximately 25 percent of descending aneurysms require placement of the cross-clamp proximal to the left subclavian artery. Operative dissection in this area is facilitated by opening the pericardium to expose the intrapericardial portion of the transverse aortic arch. The vagus nerve and recurrent laryngeal nerve should be mobilized and protected.

The technique for graft placement is standard. An initial dissection is performed to isolate the aorta proximally and distally. The aorta must be sufficiently mobilized to allow precise placement of the cross-clamps, with proximal control usually obtained first. The surgeon must then decide whether a simple cross-clamp technique or a perfusion technique is to be used.

When the aorta is clamped and opened widely, thrombus is removed from the lumen. Ostia of intercostal vessels are oversewn from within the aneurysm unless they are to be reimplanted into the side of the graft. In most patients with descending aortic aneurysms the aneurysm is not resected, but rather the graft is placed internally after the aneurysmal contents are evacuated. A woven Dacron graft is inserted, and end-to-end anastomoses are performed proximally and distally with continuous vascular suture. Rarely, an interrupted or felt-reinforced suture technique is used if the aorta is especially friable. After the graft is placed the clamps are temporarily opened to remove air or thrombus, after which the suture lines are securely tied. The cross-clamps are removed slowly, and the distal body is perfused through the graft. Reperfusion often is accompanied by transient hypotension, which is corrected by volume infusion and sodium bicarbonate to correct reperfusion-related acidosis. Once hemostasis is obtained the aneurysmal sac usually is wrapped around the graft to provide additional hemostasis and tissue coverage.

The most dreaded complication after repair of descending thoracic aneurysms is paraplegia, which continues to occur in a small proportion of patients despite numerous technical efforts to minimize the risk. Simple cross-clamping without distal perfusion can be performed relatively safely if occlusion lasts less than 20 min, but longer unprotected cross-clamp periods are associated with an increased risk of paraplegia. If the cross-clamp period exceeds 40 to 45 min, the risk of paraplegia is over 20 percent, and it approaches 100 percent once the cross-clamp time exceeds 60 min. Research suggests that when the cross-clamp time exceeds 30 to 40 min, perfusion of the distal aorta reduces the risk of paraplegia, renal insufficiency, intestinal ischemia, and reperfusion-related white blood cell activation with multiorgan dysfunction.

Perfusion Technique with Spinal Cord Monitoring

For over 15 years at NYU, a distal perfusion bypass technique with femoral–femoral or atriofemoral bypass (Fig. 19-9) has been used for most patients undergoing elective repair of descending aneurysms. The technique initially was proposed by Cunningham and associates in 1982. Perfusion options include using femoral vein–femoral artery perfusion with an oxygenator, or left atriofemoral artery bypass using heparin-bonded circuits, no oxygenator, minimal heparin, and a cell-saver–rapid infuser. The left atriofemoral bypass technique minimizes the use of heparin and allows better control of the proximal blood pressure, because the left heart is unloaded by the left heart bypass, minimizing the need for nitroprusside. Left atriofemoral bypass is preferred for most patients.

In addition to distal perfusion, a key component of this approach has been the use of *somatosensory evoked potential (SEP)* monitoring to evaluate spinal cord ischemia or the adequacy of perfusion while the aorta is occluded. Research indicates that paraplegia usually results from spinal cord ischemia because of low aortic perfusion pressure or from the direct interruption of critical segmental blood supply to the spinal cord. Diminution of spinal cord SEPs during the cross-clamp time suggests that spinal cord perfusion pressure is inadequate or critical intercostal arteries have been interrupted. If the distal perfusion pressure is low and the SEPs become abnormal, the flow rate is increased until the distal pressure exceeds 55 mmHg. This often results in a return of the SEP amplitude to normal. In patients with refractory shock it is often impossible to achieve an adequate distal perfusion pressure. In a small number of patients the SEPs are lost despite an adequate distal perfusion pressure. In this situation intercostal arteries arising from the aneurysmal segment of aorta should be reimplanted into the Dacron graft. In these patients the loss of SEPs suggests that the blood supply to the distal spinal cord is dependent on critical intercostal arteries arising from the aneurysmal segment of aorta. In approximately one-third of the population the anterior spinal artery is inadequate to supply the distal spinal cord, which receives its major blood supply from segmental arteries arising between T8 and L2. Whether this is because of a true *arteria radicularis magna* (artery of Adamkiewicz), or whether the segmental blood supply is from a cluster of intercostal arteries is debatable. The majority of available data support the idea that several segmental arteries usually are present.

The NYU experience with 78 patients undergoing descending aneurysm repair was reported by Galloway and associates in 1996. A selective operative approach was used in these patients, using bypass with distal perfusion and SEP monitoring for most elective cases and simple cross-clamp with controlled distal exsanguination for others. With this technique, the incidence of paraplegia was low, and it was zero in patients undergoing elective repair. In perfused patients, paraplegia did not occur as long as the perfusion pressure was above 55 mmHg and the SEPs remained intact. The flow rate required to maintain an adequate distal pressure and spinal cord perfusion varied from 3 L/min to as high as 6 L/min.

Many surgical groups recommend the use of a distal perfusion method routinely, especially when the cross-clamp time is expected to exceed 30 to 40 min. Borst recommended the routine use of left atriofemoral bypass for descending aneurysm repair to minimize the risk of paraplegia, renal insufficiency, and death. Lawrie and colleagues reviewed 659 patients who underwent descending thoracic aortic aneurysm resection from 1953 to 1993 at Baylor College of Medicine. The incidence of neurologic injury was reduced by the use of a bypass technique, which they recommend whenever the cross-clamp time is likely to exceed 30 min. The routine use of distal perfusion and SEP monitoring is controversial. Crawford and associates reported no difference in the incidence of paraplegia when using distal perfusion and SEP monitoring compared to the simple cross-clamp method.

Single Cross-clamping with Controlled Distal Exsanguination

The technique of single proximal cross-clamping with controlled distal exsanguination (Fig. 19-10) was popularized by Cooley and Crawford. Theoretically, controlled distal exsanguination reduces venous distention and lowers cerebrospinal fluid pressure, improving cord perfusion during the cross-clamp period and limiting spinal cord ischemia and edema. Experimentally this has been shown to provide some measure of spinal cord protection and to lengthen the amount of time that cross-clamping can be maintained safely, without neurologic injury. Published clinical data are encouraging, showing an acceptably low risk of paraplegia with this approach as long as the cross-clamp period does not exceed 30 to 40 min. In reports from Galloway, Coselli, Crawford, and Cooley the risk of paraplegia was less than 2 percent in selected patients undergoing isolated descending

aneurysm repair with use of the simple cross-clamp with controlled distal exsanguination method. Since no bypass circuit for distal perfusion is required and no heparin is used, the risk of bleeding and coagulopathy may be lower.

Results

Mortality rates are 3 to 6 percent for elective descending aneurysm repair, and 20 percent for emergency surgery in patients with rupture, hypotension, or shock. Variables associated with an increased risk of death include emergent operation, acute rupture, shock, age, and untreated concomitant coronary artery disease.

The most common perioperative complications in descending aneurysm repair are bleeding, renal failure, stroke, myocardial infarction, respiratory failure, and paraplegia. In the NYU analysis of descending thoracic aortic aneurysm repair, the factors most highly associated with an increased risk of paraplegia were emergency presentation and preoperative shock. In patients in whom a repair technique using distal perfusion and SEP monitoring was performed, paraplegia was related to the inability to maintain an adequate distal perfusion pressure above 55 mmHg and to the loss of SEPs. In patients undergoing repair with a nonperfused single- cross-clamp technique, paraplegia was primarily related to the length of the cross-clamp time.

Both perfusion techniques and single-cross-clamping methods are viable options for descending aortic aneurysm repair. The authors recommend using a selective operative strategy that includes both options but prefer a bypass technique with distal perfusion for most patients undergoing elective repair. The risk of paraplegia is exceedingly low with this approach as long as distal perfusion is adequate (above 55 to 60 mmHg) and the SEPs remain intact. The single-cross-clamp technique is particularly useful in selected patients with acute rupture, bleeding, aneurysmal erosion into the lung, or anatomically complicated aneurysms.

The long-term prognosis after aneurysm repair depends on the presence of concomitant coronary artery or cerebrovascular disease and on the presence of aneurysmal disease in other parts of the aorta. Approximately 30 percent of patients with descending thoracic aneurysms develop aneurysmal disease elsewhere, so follow-up imaging studies on a yearly basis are strongly recommended.

New Developments

A technique using endovascular stented grafts for the treatment of descending thoracic aortic aneurysms has been described. A report by Mitchell and associates from Stanford University demonstrated successful results in 44 patients with three deaths (7 percent). Expandable stented grafts are placed percutaneously to exclude the aneurysmal segment of aorta. Further long-term follow-up is necessary to evaluate the effectiveness of this procedure.

Thoracoabdominal Aneurysms

Etiology and Classification

Thoracoabdominal aneurysms usually are associated with atherosclerosis, connective tissue disease, or aortic dissection. Crawford classified thoracoabdominal aneurysms on the basis of the anatomic extent of aortic involvement as follows:

Type I—proximal descending to upper abdominal aorta

Type II—proximal descending aorta to below the renal arteries

Type III—distal descending and abdominal aorta

Type IV—suprarenal and infrarenal abdominal aorta

Physiology of Repair

Thoracoabdominal aneurysms typically involve the segment of aorta where the celiac axis, superior mesenteric artery, and renal arteries arise. Surgical repair of these aneurysms is challenging and may result in transient spinal cord, renal, and visceral organ ischemia, with subsequent ischemic reperfusion-related white blood cell activation, leading to multiorgan injury. The risks of postoperative renal insufficiency and multiple organ failure are higher in repair of thoracoabdominal aortic aneurysms than of other aneurysms. Coagulopathy and bleeding may be significant in these patients. If massive transfusion is required, the risk of multiple organ failure is increased.

The blood supply to the lower spinal cord is segmental and arises from the aorta between T8 and L2 in approximately one-third of the population. Understanding spinal cord ischemia during aortic cross-clamping is particularly important in patients with thoracoabdominal aneurysms, because these aneurysms involve the aortic segment in which the blood supply to the lower spinal cord most commonly arises. Replacement of the thoracoabdominal aorta is associated with the highest risk of paraplegia of any aneurysmal surgery, with a risk ranging from 5 to 40 percent, depending on the extent of aortic involvement and the cross-clamp time.

Operative Treatment

Etheredge and DeBakey described the first surgical approach for thoracoabdominal aneurysms in 1955, which involved replacement of the aorta with a homograft. The DeBakey technique subsequently evolved into reconstruction of the aorta with a long tube graft, from which smaller grafts were individually attached to the celiac, superior mesenteric, and renal arteries. The operative mortality was high, exceeding 50 percent for many years.

In 1965 Crawford described the basic repair technique for thoracoabdominal aortic aneurysms that is still used today. This technique uses the graft inclusion method, with reimplantation of the ostia of the visceral vessels as an island into the side of the aortic Dacron graft (Fig. [19-11](#)). Crawford summarized a 20-year experience with 605 such operations in 1986, with an operative mortality of approximately 9 percent.

The operative exposure requires a large thoracoabdominal incision. The sixth or seventh intercostal space of the left chest is entered, the costal cartilage is divided, and the incision is extended below the umbilicus. The diaphragm is divided circumferentially along the periphery, preserving the central innervation. In the left chest the descending thoracic aorta is isolated proximal to the aneurysm. The spleen, left colon, and kidney are reflected medially, and the retroperitoneum is entered to expose the aorta distally to the bifurcation. The retroperitoneal lymphatics often are tied in continuity.

Once the aneurysm has been exposed the aorta is clamped proximally. Distally the aorta is left open or clamped and perfused. The aneurysm is incised, and an anastomosis is performed between the graft and proximal aorta. Segmental intercostal arteries or clusters of arteries above and below the diaphragm are reimplanted into the graft. The cross-clamp is sequentially moved distally beyond the reimplanted intercostal arteries. The visceral vessels are reimplanted and perfused in a similar fashion. After reimplantation of the visceral vessels an anastomosis between the graft and the distal aorta is performed. This may be done immediately adjacent to the superior mesenteric and renal arteries, immediately below the renal arteries, or at the aortic bifurcation.

Crawford found that the risk of paraplegia could be significantly reduced by reimplanting large intercostal vessels, or clusters of vessels, into the aortic graft. With Crawford's technique intercostal arteries are reimplanted and perfused within 20 to 30 min.

Additional techniques for spinal cord protection include sequential reimplantation of intercostal arteries, bypass methods to increase spinal cord blood flow, the use of intrathecal vasodilators such as papaverine, systemic steroids, intravenous lidocaine for membrane stabilization, systemic hypothermia, barbiturate administration to lower neurometabolism, calcium-channel blockers and oxygen radical scavengers to minimize reperfusion injury, and cerebrospinal fluid (CSF) drainage to decrease spinal cord pressure.

The bypass techniques and the single-cross-clamp technique with controlled distal exsanguination have been used to minimize the risk of paraplegia. Distal perfusion methods, however, are used less frequently with thoracoabdominal aneurysms. The technique is cumbersome because the aorta is clamped above and below the area from which the cord segmental blood supply arises. Separate perfusion must be provided to the visceral vessels, or the cross-clamp must be sequentially moved so that the intercostal vessels are intermittently perfused before reimplantation. Nevertheless, some surgeons advocate perfusion of the aorta whenever feasible.

The single cross-clamp with controlled distal exsanguination is used more commonly. By emptying blood from the lower body (distal exsanguination), the CSF pressure is lowered significantly, which maximizes cord perfusion and increases the safe ischemic interval. For thoracoabdominal aneurysm repair the controlled distal exsanguination method is often combined with the techniques of CSF drainage to lower CSF pressure and administration of intrathecal papaverine to increase blood supply to the spinal cord by a factor of five. With single cross-clamping and controlled distal exsanguination, CSF drainage, and intrathecal papaverine, the incidence of paraplegia has been as low as 0 to 5 percent in some reports. Despite these advances and the variety of technical options available, the optimal method of spinal cord protection during thoracoabdominal aneurysm repair is controversial.

Results

In 1993 Svensson and associates reported the cumulative Baylor College of Medicine experience with 1,509 patients undergoing thoracoabdominal aneurysm repair over 30 years. The operative mortality rate was 8 percent, and the incidence of paraplegia was 16 percent. The other major complications included renal insufficiency (18 percent), gastrointestinal complications (7 percent), and myocardial infarction. Variables associated with increased operative risk were age, preoperative creatinine level, concurrent proximal aortic aneurysm, concomitant coronary artery or pulmonary disease, and aortic cross-clamp time. Variables associated with an increased risk of paraplegia were cross-clamp time, extent of the aneurysmal involvement (Crawford types I and II), aortic rupture, age, proximal aortic aneurysm, and a history of preoperative renal insufficiency.

The extent of aneurysmal involvement is strongly associated with the risk of paraplegia. In Svensson's report the incidence of paraplegia was 31 percent in patients with Crawford type II thoracoabdominal aortic aneurysms. Similarly, in a report by Griep and colleagues the incidence of paraplegia was 30 percent in patients with Crawford type I or type II aneurysms. In large aneurysms with involvement of more than 10 intersegmental arteries, the risk of paraplegia increased by a factor of 29.

AORTIC DISSECTION

Etiology

Aortic dissection begins as a tear in the intima, with entry of blood and separation of the media for a variable distance, resulting in blood flow down a “false lumen.” A localized aneurysm may develop immediately, or months or years later where the aortic wall has become weakened and enlarged from the original dissection. The disease is 3 to 4 times more common in males than in females and occurs predominantly in older patients. Dissection may occur in any age group, but with certain cases occurring in childhood, usually secondary to coarctation of the aorta.

Aortic dissection usually results from a combination of hypertension and degenerative connective tissue disease. Roberts has emphasized that a history of hypertension is obtainable in 60 to 75 percent of patients, with hypertrophy of the left ventricle present in approximately 90 percent. Roberts determined that hypertension frequently is the precipitating factor in patients with Marfan syndrome who develop aortic dissection and predicted that proper control of hypertension would significantly lower the risk of dissection in all patients.

The strongest predisposing factor to aortic dissection is cystic medial necrosis, which may be idiopathic (Erdheim's cystic medial necrosis) or secondary to a known connective tissue disease, such as Marfan syndrome or Ehlers-Danlos syndrome (see above, Thoracic Aneurysms, subsection Etiology and Pathogenesis). Other factors associated with aortic dissection include aortic coarctation and congenital bicuspid aortic valve. Rarely, an atherosclerotic plaque or traumatic injury serves as the initiating tear site for aortic dissection.

Aortic dissection usually is not caused by atherosclerosis. Atherosclerosis is a proliferative disease of the intima and media, occurring most frequently in the abdominal aorta. Aortic dissection is a disease of the media, almost always originating in the thoracic aorta, although the dissection may continue distally to the aortic bifurcation. Although aortic dissection and atherosclerosis occur in older patients, they are distinctly different and unique disease processes. Rarely, an atherosclerotic plaque serves as a lead site for aortic dissection.

Pathology

The major initiating pathologic event is a tear in the intima and media, usually involving half the circumference of the aorta. The intimal tear permits blood to enter the media and dissect distally. The aortic wall progressively separates ("dissects") with an inner lumen composed of intima and an outer false lumen composed of the media and adventitia. In the classic pathologic analysis published by Roberts, the intimal tear was located in the ascending aorta in about 70 percent of patients, in the aortic arch in 10 percent, in the upper descending thoracic aorta near the ligamentum arteriosum in 20 percent, and in the abdominal aorta in about 2 percent. In a study by Miller and associates the intimal tear was found to be in the ascending aorta in 60 percent of patients, in the aortic arch in 10 percent, and in the descending aorta in 30 percent.

Once the dissection begins, it usually extends rapidly through the thoracic and abdominal aorta into the peripheral arteries. The dissection process extends into a peripheral artery in more than 50 percent of patients. Roberts has estimated that the entire aorta will dissect within minutes unless some structural abnormality that has disrupted continuity of the aortic wall, such as atherosclerosis or coarctation, halts the dissection. If this theory is correct, younger patients with less atherosclerosis would more frequently have dissection involving the entire aorta. A "reentry" tear can be identified in most patients, located in the aorta in about one-half of the patients, and in a peripheral artery in the others.

As the dissection progresses, branch vessels are sheared off and obliterated unless a communication with the false lumen is established. Proximally, the coronary arteries may be involved. Often one or more aortic valve commissures are detached, creating aortic insufficiency. The commissure between the right sinus and the noncoronary sinus is most commonly involved. Distally, any vessel may be involved. Innominate or carotid artery involvement may produce neurologic injury. Obstruction of a subclavian artery may produce arm ischemia and a differential pressure between the two arms. Occlusion of intercostal arteries may cause spinal cord injury with paraparesis or paraplegia. Dissection of renal arteries may produce renal insufficiency, hematuria, oliguria, or anuria. Distally, acute obstruction of the iliac or femoral arteries may cause leg ischemia, manifested with pain, sensory loss, or even gangrene. Overall, approximately 5 percent of patients with aortic dissection have some degree of visceral ischemia, peripheral neurologic injury, or paraplegia on admission.

The dissection may result in a fatal complication at any time. Rupture into the pericardial cavity with cardiac tamponade is the most common fatal complication, probably because the velocity of blood flow and aortic diameter are greatest in the ascending aorta. Rupture into the left pleural cavity or the retroperitoneum occurs less commonly.

Clinical Manifestations

The abrupt onset of excruciating pain, almost immediately reaching its peak intensity, is characteristic of aortic dissection. A patient with a myocardial infarction, by contrast, may gradually develop pain of increasing severity. Sutton and associates reported that chest pain, usually in the anterior chest, occurred in nearly 80 percent of 113 patients with aortic dissection. Back pain occurred in about one-third of the patients, suggesting that absence of back pain does not rule out a dissection of the thoracic aorta. Another characteristic is the tendency for the pain to migrate into different areas as the dissection extends distally. Many different pain syndromes can occur. The pain may radiate to the neck, the arms, the epigastrium, or the legs. It might mimic myocardial infarction or pulmonary embolus. The diagnosis of aortic dissection must be considered in patients with suspected myocardial infarction or pulmonary embolus in order to avoid a fatal treatment error. Pain is seldom completely absent in a patient with acute aortic dissection, although Spittell and coworkers reported that 15 percent of 236 patients presented with painless dissection.

Other presenting symptoms include congestive heart failure, tamponade, syncope, stroke, peripheral neurologic injury, leg or arm ischemia, paraplegia, gastrointestinal hemorrhage, hematuria or anuria, hoarseness, dysphagia, superior vena cava syndrome, and aortic insufficiency. In the Spittell study the initial clinical impression was a diagnosis other than aortic dissection in 38 percent of the patients. An awareness of the variety of symptoms associated with aortic dissection is essential if the diagnosis is to be made promptly.

Classification

DeBakey classified aortic dissections into types I, II, and III (Fig. 19-12). In the DeBakey type I dissection the tear site originates in the ascending aorta, usually just above the left main coronary artery, and the dissection continues distally into the descending or abdominal aorta. In the type II dissection the tear site is in a similar location in the ascending aorta, but the dissection stops distally at the innominate artery. In the type III dissection the tear site originates in the upper descending thoracic aorta, just distal to the subclavian artery; in type IIIA the dissection is localized in the thoracic aorta, and in type IIIB the dissection proceeds into the abdominal aorta.

The Stanford classification proposed by Miller is based on the clinical course and the surgical significance of the dissection (see Fig. 19-12). Stanford type A dissection includes any dissection involving the ascending aorta (DeBakey types I and II), and Stanford type B dissection involves only the descending aorta (DeBakey type III). Overall, approximately two-thirds of patients with acute dissection have Stanford type A, and one-third have Stanford type B. This is significant, because the prognosis without treatment is much worse for Stanford type A dissections.

Diagnostic Studies

On the chest radiograph a widened mediastinum or a left pleural effusion from extravasation of blood frequently is seen. In some patients the chest radiograph is completely normal. The electrocardiogram (ECG) is of value in distinguishing a dissection from a myocardial infarction, but there are no characteristic features of aortic dissection. Occasionally the ECG is misinterpreted because of the presence of ST-segment elevation secondary to hypertension and ventricular strain. If patients are treated for myocardial infarction with thrombolytic therapy on the basis of this misinterpretation, catastrophic exsanguination may occur. The most common abnormalities on the ECG are sinus tachycardia or left ventricular hypertrophy from the antecedent hypertension.

Transesophageal echocardiography (TEE) is the initial diagnostic procedure of choice for most patients with suspected aortic dissection (Fig. 19-13). The examination can be performed in the emergency department immediately whenever a diagnosis of aortic dissection is considered. TEE is sensitive and specific, establishing the diagnosis with a 99 percent accuracy. The tear site usually can be located, and the ascending aorta and aortic arch can be assessed and potential complications such as aortic valvular insufficiency and cardiac tamponade identified.

If the cardiology team is not equipped to make the diagnosis promptly by TEE, other valuable tests include a rapid-sequence CT scan with contrast, MRI, or enhanced MR angiography (Fig. 19-14). Aortography is highly accurate, but it is seldom necessary.

Medical Treatment

Once the diagnosis of aortic dissection is established, immediate drug therapy to control the blood pressure and decrease the forceful contractility of the left ventricle (dp/dt) is initiated. Antihypertensive therapy should be started when the diagnosis is suspected because lowering the blood pressure may stop the dissection process and prevent exsanguination. Various pharmacologic regimens are available, using combination drug therapy to achieve beta-adrenergic blockade and afterload reduction. The primary goal of medical therapy is to reduce shear stress by reducing left ventricular dp/dt. The systolic blood pressure should be kept below 110 to 120 mmHg.

Natural History and Operative Indications

Patients with acute Stanford type A dissection are at risk for early death because of aortic rupture or cardiac and neurologic complications. These patients have an extremely poor prognosis with medical therapy, and they should be operated on emergently. In contrast, patients with acute Stanford type B dissection have a relatively good prognosis, and early operation in this group generally is recommended only for patients with complications.

The high mortality rate for unoperated patients with Stanford type A dissection is documented in virtually every report. Thirty to 50 percent of these patients die within 24 h, 50 to 75 percent within 1 to 2 weeks, and 90 percent within 3 months. In a classic review of 425 cases by Hirst, 74 percent of patients with type A dissection died within 2 weeks and 91 percent within 6 months. Of the 62 patients reported by Lindsay and Hurst, almost all of the untreated type A patients died within 1 month.

The survival rate for patients receiving nonoperative therapy for Stanford type B dissection is much better, with a 1-month survival rate of 85 to 90 percent. Consequently, most patients with type B dissections are treated medically unless a complication develops. Complications requiring immediate operative treatment are rupture, hemodynamic compromise, prolonged pain, aneurysm expansion, visceral or limb ischemia, and new neurologic signs. Close observation is mandatory in patients receiving medical therapy. This includes serial hematocrit determinations, chest x-rays, and follow-up imaging studies during the initial hospital stay. Approximately 10 percent of patients with acute type B dissection treated medically develop a serious complication within 2 weeks, and prompt surgery is indicated for these patients.

In patients who survive acute aortic dissection a chronic aneurysm may develop, usually with a double lumen in the distal aorta. Blood flow often is present in both lumens, although the false lumen clots and heals in approximately 20 percent of the patients. When flow is present in both aortic lumens, the visceral blood supply may arise from the true lumen or the false lumen. For example, one renal artery may arise from the false lumen and the other from the true lumen, or both renal arteries may arise from the false lumen while all other vessels arise from the true lumen. Frequently, the aorta gradually becomes aneurysmal, especially if hypertension is poorly controlled. For patients with chronic aneurysms secondary to dissection, surgical intervention is recommended when the aorta grows larger than 5 to 6 cm in diameter, similar to the recommendations for other patients with chronic aneurysms. Patients with chronic aortic dissection must be monitored closely with yearly imaging studies; approximately 30 to 40 percent of these patients require surgery within 5 years.

Operative Treatment of Stanford Type A Dissection

Modern surgical treatment of ascending aortic dissection evolved from the work of DeBakey and Cooley, who reported successful excision and grafting of a chronic ascending dissection in 1955. In the early years the operative mortality exceeded 50 percent. Today the mortality rate for repair of acute type A dissection is less than 10 percent, which

compares favorably to the poor prognosis with medical therapy. Prompt operation is recommended for almost every patient with acute Stanford type A aortic dissection.

The main objectives of operation are to remove the intimal tear site, to replace diseased or dilated aorta as necessary, to obliterate the false lumen and redirect blood flow into the true lumen, and to correct associated valvular insufficiency or coronary ischemia. These goals usually are best accomplished with use of deep hypothermia and circulatory arrest (described above for arch aneurysm repair). Use of circulatory arrest allows the surgeon to avoid clamping and potentially injuring the diseased aorta and allows performance of the distal anastomosis “open,” under direct vision. At NYU a variation of the hemi-arch replacement technique (Fig. 19-15) is used almost exclusively for repair of acute type A dissections. The technique includes internal replacement of the dissected aortic segment with a Dacron graft.

The most frequent cause of operative death is hemorrhage. Performing the distal arch anastomosis with an “open” technique during circulatory arrest permits precise inclusion of all layers of the dissected aorta. The anastomoses usually are performed with a continuous suture of 4-0 polypropylene; excessive tension on the suture line, which may lacerate the friable intima, is avoided. Once the distal anastomosis is completed, the prosthetic graft is occluded and flow to the brain is restored while the proximal anastomosis is performed. The aortic wall is closed around the graft to complete the graft inclusion technique. This technique limits the risk of operative exsanguination.

Alternatively, the aorta may be excised totally, with removal of the tear site and any aneurysmal segments. The intima and adventitia are reapproximated proximally and distally with Teflon felt reinforcement, obliterating the false lumen. An end-to-end interposition Dacron graft is then inserted proximally and distally. This method may be associated with fewer late false aneurysms, but the risk of early bleeding may be increased.

Miller reported that excision of the intimal tear site does not influence late survival. In contrast, the NYU experience suggests that removal of the tear site from the circulation is a basic principle in the treatment of aortic dissection and should be done whenever possible.

When the aortic valve is involved with aortic dissection, resulting in aortic insufficiency, resuspension of the valve has been highly effective. Fann and associates reported satisfactory durability after aortic valve resuspension. In the NYU experience reported by Galloway, patients with severe valvular insufficiency underwent successful valve repair using valve reconstruction techniques adopted from the homograft experience. If a competent valve cannot be assured, however, aortic valve replacement should be performed; the long-term results are excellent after valve replacement under these circumstances. Overall, the risk in acute type A aortic dissection repair is relatively low, with an operative mortality rate of 5 to 10 percent in most major medical centers. Factors that increase operative risk are rupture, shock, and visceral organ ischemia.

Operative Treatment of Stanford Type B Dissection

With Stanford type B dissections originating in the descending thoracic aorta, most groups advocate initial medical therapy. In 10 to 15 percent of these patients urgent operation is indicated for complications such as recurrent pain, progressive mediastinal hematoma, leakage, acute expansion, rupture, visceral organ ischemia, extremity ischemia, progressive neurologic dysfunction, and retrograde dissection with aortic valve involvement.

The operation for repair of type B dissection is performed through a left thoracotomy. The goals of therapy are to exclude the tear site from the circulation, to obliterate the false lumen, and to redirect blood flow through the graft into the true lumen of the aorta distally. The operative techniques used to replace the descending aorta for a Stanford type B dissection are identical to those described previously for aneurysms of the descending thoracic aorta.

The operative mortality rate for repair of acute type B dissection is 10 to 15 percent, primarily because most type B repairs are done emergently. If initial medical therapy is successful and urgent surgery is not necessary, the patient is reevaluated in 1 to 2 months and yearly thereafter. Elective repair is recommended if the patient develops symptoms or progressive aneurysmal disease.

Prognosis

Aortic dissection usually occurs in patients with chronic hypertension and in those with chronic degenerative disease of the aorta. Aggressive treatment of hypertension and careful follow-up for the remainder of the patient's life is mandatory. The residual false lumen beyond the site of repair may gradually enlarge and become aneurysmal within several years. In the series reported by Miller and associates the 5-year and 10-year survival rates were 76 and 37 percent, respectively. Late rupture in another segment of the aorta accounted for 30 percent of the late deaths.

Tight control of hypertension is essential. More than one fatal aortic rupture has resulted from inadvertent cessation of antihypertensive therapy years after recovery from surgical treatment of aortic dissection. Late aneurysm formation in another part of the aorta occurs in at least 30 percent of patients with aortic dissection within 5 years. Therefore, patients with aortic dissection must be monitored carefully for any sign of aortic enlargement; a yearly MRI or CT evaluation for the remainder of the patient's life is recommended.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 20 - Arterial Disease

Kenneth Ouriel
Richard M. Green

The years following World War II have seen a logarithmic increase in the number of patients seeking treatment for diseases of the peripheral arteries. Today, approximately 1 million arterial reconstructions are performed in the United States annually. As the geriatric population grows at a disproportionate rate compared to the general population, the incidence of peripheral vascular disease will increase accordingly.

Of necessity, the evolution of vascular surgery awaited the development of arteriography, anesthesia, anticoagulation, blood transfusion, and synthetic graft materials. Nevertheless, Hallowell of England reported the first successful arterial operation in 1759 when he performed a lateral arterial repair of a traumatic wound. Eck is credited with the first formal blood vessel anastomosis when he sutured the portal vein to the inferior vena cava of a dog in 1877. Murphy performed the first end-to-end arterial anastomosis in a human being two decades later when he successfully rejoined the femoral artery by invagination of the proximal into the distal end after the excision of an arteriovenous fistula of the thigh. Alexis Carrel and Charles Guthrie made remarkable achievements as a result of collaborative efforts beginning in 1904. They pioneered the use of Dorfler's technique of through-and-through sutures of all layers of the vascular wall, and in 1912 Carrel received the Nobel Prize in physiology and medicine for his work on blood transfusion, vascular suture technique, and organ transplantation in experimental animals.

Rapid advances in vascular surgery began after World War II, beginning with the treatment of arterial lesions with endarterectomy by dos Santos in 1947 and with bypass using autogenous vein by Kunlin in 1951. Dubost first replaced an abdominal aortic aneurysm with an aortic homograft in 1951, and Voorhees and Blakemore used a graft made of synthetic fabric the following year. The successful treatment of carotid disease was first reported by Eastcott, Pickering, and Rob in 1954, when a symptomatic carotid bifurcation was resected and reanastomosed; 2 years later Cooley and his colleagues published a report of carotid endarterectomy for stenotic disease.

PATHOLOGY AND PATHOPHYSIOLOGY

Atherogenesis

Vascular injury and thrombus formation are the major events in the formation and progression of the atherosclerotic lesion. The most widely held view on the genesis of atherosclerosis, the "response-to-injury" hypothesis, was advanced by Ross in 1986. Three categories of vascular injury of increasing severity have been proposed (Table 20-1). In type I injury there is a functional alteration of the endothelial cell without morphologic changes. It has been hypothesized that the injury can occur from flow disturbances in certain parts of the arterial tree. Lipids accumulate in macrophages, and the lipid-laden "foam cells" may represent the earliest sign of atherosclerosis. Type II injuries begin with the release of toxic products from the macrophages. The subsequent adhesion of platelets at sites of injury and the release of a variety of growth factors results in the migration and proliferation of primitive smooth muscle cells and the development of the "fibrointimal" lesion. By the third decade of life some of these lesions become soft plaques with a cap of smooth muscle cells and collagen surrounding the lipid material. Type III injuries are characterized by fissures and disruption of the plaque with penetration into the media, exposed fibrillar collagen, increased platelet adherence and activation, thrombus formation, and extensive proliferation of smooth muscle cells. This mural thrombus deposited at the site of plaque disruption is important in the progression of the atherosclerotic plaque.

Platelets

Experimental injury models have characterized the time course of the platelet response. Phase I occurs with the deposition of thrombus composed of platelets and fibrin, beginning immediately after injury and completed within 24 h. Smooth muscle cells in the media begin to proliferate within 24 h. Phase II begins 4 days after injury and continues through day 14 and is characterized by migration of the smooth muscle cells into the intima. Phase III lasts from day 14 to 3 months and is marked by the process of intimal thickening and the accumulation of an extracellular matrix. Pigs lacking von Willebrand factor, a protein important in platelet adherence to injured vessel wall, do not develop spontaneous atherosclerosis. Furthermore, rabbits rendered thrombocytopenic do not develop the same degree of intimal thickening when subjected to balloon catheter injury as animals with normal platelet counts. These observations attest to the importance of the platelet in atherogenesis.

Macrophages

Macrophages are involved in the earliest stages of atherogenesis. These multipurpose cells facilitate the transport and oxidation of cholesterol, secrete a mitogenic growth factor that stimulates the proliferation of smooth muscle cells, and generate toxic products that produce endothelial damage. Most important, macrophages release proteases that digest extracellular matrix, which may be responsible for the plaque disruption that leads to thrombosis.

Plaque Disruption

The intact endothelium is nonreactive to platelets. Recent pathologic data in patients who died of cardiac events suggest that the occluded coronary artery resulted from recurrent episodes of plaque disruption that were serially covered with layered thrombus of varying ages in a repetitive process. The progression of a stenosis from a mild degree of diameter reduction to total occlusion can occur quite rapidly, resulting in distal ischemia in the form of a stroke, myocardial infarction, or lower-extremity gangrene. It has been found, at operation or in postmortem examination, that many of these patients have small to moderately sized lipid-laden plaques with disruption and that the severe stenosis or occlusion was the result mainly of superimposed thrombus. Angiographic studies in patients with acute myocardial infarction who have received thrombolytic therapy have shown that a considerable proportion of these patients have less than a 70 percent stenosis. The soft, lipid-rich plaque of mild to moderate severity is more prone to disruption than the fibrotic, calcified plaque because of its high fat content. Once a thrombus forms over an area of vascular injury, further clot deposition occurs. The luminal thrombus is thrombogenic itself, encouraging continued deposition. The process encroaches on the vessel lumen, resulting in a further increase in the degree of stenosis and an increase in the shear rate. This augmented shear rate stimulates further platelet activation and deposition, and the cycle repeats itself until occlusion occurs.

Tobacco Effect

Cigarette smoking has a well-documented effect on atherogenesis. The Framingham study documented an incidence of peripheral vascular disease in smokers of 0.65 percent, compared to 0.22 percent for nonsmokers. Other studies have shown that the risk for developing intermittent claudication was fifteen times higher in smokers and seven times higher in female smokers than in nonsmokers. The incidences of abdominal aortic aneurysms, amputations, stroke, and myocardial infarctions are all significantly higher in smokers.

The mechanisms by which cigarette smoking exerts its effect on atherogenesis are complex. Nicotine and carbon monoxide appear to be the most harmful constituents. In addition to the systemic effects of increased heart rate, increased blood pressure, and reduced myocardial oxygen delivery, these compounds exert adverse effects on vascular endothelium. Carbon monoxide causes increased permeability of the vessel wall to lipids. This is important because nicotine produces increased levels of circulating free fatty acids, which increase intracellular lipid deposition. Nicotine infused experimentally causes a significant increase in circulating carcasses of endothelial cells. This finding has been reproduced in human beings after smoking only two cigarettes. Nicotine also decreases cell synthesis of *prostacyclin* (*PGI₂*), the most potent inhibitor of platelet aggregation, and promotes the production of thromboxane A₂, which promotes platelet aggregation. Other deleterious effects of smoking include increased blood viscosity and fibrinogen and low-density lipoprotein (LDL) levels and decreased high-density lipoprotein (HDL) levels.

Cessation of smoking reduces the incidence of amputation and increases longevity in patients with peripheral vascular disease, improves walking distances in patients with claudication, and reduces the risk of stroke and myocardial infarction in the general population. The tobacco industry has responded to these data with the so-called low-yield cigarette. Filtered cigarettes reduce exposure to tar and are associated with a lower cancer mortality than unfiltered cigarettes. However, filtered cigarettes are not associated with any reduction in cardiovascular mortality rates. Therefore, the most important prophylactic advice that physicians can give patients is to refrain from smoking.

Prevention

Since lipid-rich plaques and subsequent thrombosis play a major role in atherogenesis, strategies for prevention have focused on the manipulation of lipid metabolism and platelet function.

Lipid Reduction

Evidence of regression of human atherosclerotic lesions is difficult to obtain because of the lack of reliable methods to serially quantitate the extent of the process. Angiographic studies of patients with coronary and femoral artery plaques suggest that some regression does occur with the use of lipid-lowering drugs. These studies have shown a small increase in residual lumen size and suggest that the regression is limited to the soft, lipid-rich plaque rather than the extensive, calcified fibrotic plaque. Thus, clinical usefulness appears to be confined to early, asymptomatic lesions without significant stenosis or occlusion. The National Cholesterol Education Program has proposed that persons without vascular disease maintain an LDL cholesterol level of 130 mg/dL or less. A patient with a level higher than 160 mg/dL should be treated with cholesterol-lowering drugs if diet and exercise do not reduce the levels to acceptable ranges.

Antiplatelet Therapy

The interaction between platelets and the vessel wall is dependent on the balance between thromboxane A₂ and *PGI₂*. Endothelial cells produce *PGI₂*, which is responsible for some of the thromboresistant properties of vascular endothelium. In addition, *PGI₂* enhances the activity of cholesterol ester hydrolase, suggesting a positive feedback between the *PGI₂* system and lipid accumulation in the vessel wall. *PGI₂* generation from atherosclerotic arterial tissue has been shown to be significantly lower than from normal arterial tissue. Platelets in patients with arterial thrombosis produce more thromboxane A₂ than normal.

Aspirin inhibits platelet activation induced by the release of thromboxane A₂ and is the most commonly used antiplatelet drug. Aspirin binds irreversibly to the active site of cyclooxygenase, inhibiting the conversion of arachidonic acid to thromboxane A₂ in the platelet and to *PGI₂* in the endothelium. Aspirin has theoretical limits as an antiplatelet agent since platelet aggregation is a complex mechanism that can also be initiated by adenosine diphosphate (ADP) and thrombin in the absence of thromboxane A₂. A single dose of aspirin leads to a platelet defect that lasts 7 days. The effect on the endothelium lasts for a shorter period, presumably because the endothelium can synthesize new enzyme, whereas the platelet cannot. Studies on human vascular fragments have revealed that aspirin doses of 40 mg/day inhibit both vascular wall *PGI₂* and thromboxane A₂ formation, with a greater effect on the latter. Fewer data are available on the efficacy of other antiplatelet agents that work via selective thromboxane A₂ inhibition, phosphodiesterase inhibition (dipyridamole and iloprost), thrombin inhibition (heparin, hirudin, and hirudin analogues), platelet membrane receptor inhibitors (ticlopidine 7E3), or angiotensin-converting enzyme (ACE) inhibitors (cilazapril).

Large clinical trials have evaluated the role of antiplatelet drugs in the prophylaxis of myocardial infarction and stroke. Overall, regardless of indication, antiplatelet therapy was most effective in the first year of treatment, there was a smaller but still significant benefit in year two, and there was no benefit in year three and beyond. A cumulative risk reduction of 25 percent was independent of disease categories, age, sex, blood pressure, and type of antiplatelet therapy. A separate analysis of the stroke cohort showed that antiplatelet therapy in high-risk patients resulted in a reduction of occlusive stroke occurrence, a small increase in the number of hemorrhagic strokes, and a substantial and significant net

reduction in total stroke occurrence. The effect of antiplatelet therapy on stroke in low-risk patients, i.e., as primary prevention, was unclear. In a cardiac cohort, antiplatelet therapy was shown to reduce the incidence of nonfatal myocardial infarction by 33 percent. Unlike in the stroke group, the cardiac effect was seen in both primary and secondary prevention.

Thrombogenesis

Thrombogenesis may be conceptualized as the interaction of three pathways: coagulation, platelet deposition, and thrombolysis. Coagulation pathways involve the clotting proteins in the plasma, terminating in the cleavage of fibrinogen and the deposition of fibrin matrix. Platelet pathways involve activating substances in the plasma and, as such, are intimately linked to the coagulation pathway. The end result of these interactions is the activation, attachment, and aggregation of platelets at the site of altered endothelial integrity. Thrombolytic pathways involve the cleavage of fibrin by plasmin to maintain luminal patency and tissue perfusion in the event of vessel thrombosis.

Coagulation

The coagulation pathways involve a cascade mechanism, with activation of clotting proteins through two pathways: the intrinsic pathway and the extrinsic pathway. Many of the clotting proteins have been assigned a Roman numeral by the International Committee on Nomenclature of Blood Clotting Factors.

There are two laboratory tests that are widely used to assess the adequacy of anticoagulation. The activated partial thromboplastin time (aPTT) reflects the potential activity of the intrinsic coagulation pathway, and the prothrombin time (PT) provides an index of the extrinsic system. The vitamin K–dependent factors II, VII, IX, and X are synthesized in the liver by gamma carboxylation–dependent mechanisms. Warfarin compounds inhibit the production of the vitamin K–dependent factors. The therapeutic control of warfarin anticoagulation is best monitored with the PT. Heparin achieves anticoagulation through its actions on antithrombin III, a potent inhibitor of factor X and other components, and therapeutic control is generally monitored with the aPTT.

The intrinsic coagulation pathway is initiated with the activation of Hageman factor (factor XII) by a variety of substances, including collagen, trypsin, and endotoxin. Activated factor XII in turn catalyzes the conversion of factor XI to its activated form, and the cascade continues in this fashion through factors IX and X. Activated factor X accelerates the conversion of prothrombin to thrombin, and thrombin cleaves fibrinopeptide A from fibrinogen to form fibrin monomer. Thrombin also activates factor XIII to cross-link the soluble fibrin monomers to form stabilized, insoluble fibrin clot.

The extrinsic pathway is initiated with the release of thromboplastin (tissue factor) from injured endothelial cells, catalyzing the activation of factor X. The coagulation cascade then proceeds through the common pathway to terminate in the formation of insoluble fibrin polymer.

There are several other factors that are important in coagulation mechanisms. Proteins C and S are vitamin K–dependent factors that inactivate factors V and VIII and thus function as natural anticoagulants. Calcium (factor IV) is instrumental in the activation of almost all the coagulation reactions, which explains the anticoagulant mechanism of calcium chelators such as ethylenediaminetetraacetic acid (EDTA) and citrate-phosphate-dextrose compounds. High-molecular-weight kininogen functions with prekallikrein to orient factors XII and XI on negatively charged surfaces to potentiate their activation.

Platelet Deposition

Platelets are membrane-bound cytoplasmic remnants of bone marrow megakaryocytes. Platelet function is essential in the sealing of vascular defects through the formation of a platelet plug. The intact endothelium normally conceals the adhesive glycoproteins (von Willebrand factor, fibronectin, and collagen) from the blood elements, thereby limiting platelet adhesion to sites of vessel injury. Endothelial damage results in platelet adhesion on the vessel wall, as von Willebrand factor binds to exposed subendothelial collagen and platelet membrane glycoproteins Ib and IIb/IIIa to form a bridge between collagen and the platelet.

Platelet activation follows adhesion, mediated by such agents as ADP, thromboxane A₂, collagen, epinephrine, and thrombin. Platelet arachidonic acid is released from membrane phospholipid and is metabolized by cyclooxygenase to prostaglandins G₂ and H₂, a process that is blocked by aspirin. These prostaglandin intermediaries are subsequently converted to the potent platelet-aggregating agent thromboxane A₂. Thromboxane A₂ produces platelet activation by interacting with a receptor located on the surface of platelets; thus it must exit from the platelet to bind to the same or a neighboring platelet in order to be effective.

Activated platelets undergo a process known as the release reaction, with migration of platelet granules to the cell membrane and release of their contents. The alpha granule contains fibrinogen, platelet factor 4, platelet- derived growth factor, von Willebrand factor, and fibronectin. The secretion of these agents from the platelet amplifies the process of activation, resulting in stimulation of neighboring platelets and initiation of platelet aggregation. It is important to note that the arachidonic acid pathway is not absolutely required for platelet activation. Although aspirin-treated platelets cannot generate thromboxane A₂, two agonists, thrombin and collagen, can cause the release of the contents of platelet storage granules even when the arachidonic acid pathway is blocked.

Platelet aggregation occurs through platelet-platelet cohesive attachment, principally by means of a mechanism involving fibrinogen and platelet glycoprotein IIb/IIIa. Glycoprotein IIb/IIIa is the platelet receptor most densely distributed on the cell membrane, with up to 50,000 molecules present on a stimulated platelet. Glycoprotein IIb/IIIa is unique in its ability to bind multiple ligands, including von Willebrand factor and fibrinogen, because of its binding site for the tripeptide arginine-glycine-aspartic acid (RGD, using the single-letter peptide nomenclature), a sequence that is present in all ligands that bind to the receptor. Through its RDG site, fibrinogen binds to the glycoprotein IIb/IIIa receptors of two platelets and forms a bridge between the two platelets, initiating the process of platelet- to-platelet attachment and the formation of a platelet plug.

Thrombolysis

Thrombolysis of formed clot involves the actions of plasmin, a proteolytic enzyme generated by the activation of plasminogen. Activators of plasminogen are present in the endothelium and the plasma (tissue plasminogen activator [t-PA]) and can be prepared from bacterial sources (streptokinase) or renal parenchymal cell culture techniques (urokinase). Factor XIIa and prekallikrein also activate plasminogen. Inhibition of plasmin occurs principally through the actions of a₂-antiplasmin. The binding of free plasma plasmin to antiplasmin is critical to the physiologic control of thrombolysis to prevent a systemic fibrinolytic state.

Plasmin lyses both fibrinogen and fibrin to produce proteolytic fragments X, Y, and A through E. Laboratory confirmation of thrombolysis is accomplished by measuring the level of generic fibrin degradation products or of *D dimer*, a unique fragment formed by the covalent binding of two fragment D moieties.

Pathophysiology of Arterial Obstruction

Peripheral arterial blood flow follows the physical principles of fluid dynamics as described in Bernoulli's principle, which characterizes the energy changes of liquids flowing in pipes, and Poiseuille's law, which describes the effects of viscosity on blood flow. A full explanation of these laws and the impact of the variables in the circulation is beyond the scope of this text. There are, however, a number of concepts that are necessary for understanding the basic pathophysiology of arterial occlusive disease.

Critical Arterial Stenosis

A critical stenosis may be defined as that degree of narrowing sufficient to produce a pressure drop. Pressure gradients occur as a result of energy losses and do not become evident until the cross-sectional area of a vessel is reduced by more than 75 percent. In practice, the precise measurement of an area reduction is difficult to obtain because most plaques are irregular. If the lesion is concentric, however, a 75 percent reduction in area corresponds to a 50 percent reduction in diameter. Energy losses also depend on the magnitude of blood flow across the stenosis, and blood flow is determined by the blood pressure and the resistance of the runoff bed. Thus in the coronary and carotid systems, where peripheral resistances are low, a critical stenosis may be reached with less luminal narrowing than in higher-resistance circuits such as the lower-extremity vessels. This phenomenon also explains how an iliac artery lesion that looks insignificant on arteriography can severely restrict a patient's activities when the resistance in the runoff bed is reduced during exercise.

The length of a stenosis is far less important than its diameter. Length figures in Poiseuille's equation in the first power, whereas radius is elevated to the fourth power. Doubling the length of a stenosis doubles the energy loss, but halving the radius increases the losses by a factor of sixteen. A considerable energy loss also occurs at the entrance and at the exit of a stenosis, where the losses are related to the fourth power of the diameter ratio of the normal to the stenotic segment. Thus separate stenoses of equal diameter are more significant than a single stenosis of the same diameter whose length equals the sum of the lengths of the two independent lesions.

Collateral Circulation

Flow reduction to the runoff bed distal to a stenosis depends on the collateral network, a group of preexisting pathways that enlarge as a stenosis develops in the main arterial supply. Collateral pathways may provide flow distal to an occlusion sufficient to preserve viability. Perfusion through collateral channels is never as efficient as through a patent artery, however, because the resistance of the collateral network always exceeds that of the major artery. Although arterial flow may be normal at rest, exercise will almost always bring out abnormalities in arterial flow.

Autoregulation

The blood vessels in skeletal muscles are innervated by vasoconstrictor and vasodilator nerve fibers, but these actions are superseded by locally produced metabolites. The term *autoregulation* refers to the ability of vascular beds to provide a constant blood flow regardless of perfusion pressures. This adaptability is the result of a myogenic response of the vessel wall to the local chemical environment. The mechanism fails and flow is not maintained when perfusion pressures fall below 20 to 30 mmHg for skeletal muscle and 50 to 60 mmHg for the brain.

Effect of Exercise

Resting blood flow is decreased in patients with ischemic rest pain but is normal in patients with intermittent claudication because of collateral circulation and autoregulation. With exercise, however, the intramuscular arterioles dilate, resistance falls, flow increases, and a pressure gradient develops across the arterial stenoses and collateral channels. A previously palpable pedal pulse may be lost as the distal perfusion pressure falls. Blood flow is insufficient to meet the local demands of the tissues. Metabolites accumulate, producing the pain characteristic of intermittent claudication.

DIAGNOSIS OF OCCLUSIVE ARTERIAL DISEASE

Clinical Manifestations

Atherosclerotic peripheral arterial disease is a generalized process. The patient who presents with symptoms of peripheral vascular disease is likely to have some degree of coronary artery disease as well. More than 50 percent of the mortality following arterial reconstruction of any type is attributable to cardiac events. Hertzner and colleagues found, in

coronary arteriograms they performed on 1000 consecutive patients undergoing elective vascular procedures, that over 90 percent of the patients had evidence of coronary artery disease. Triple-vessel disease was identified in 30 percent of these patients.

Despite the high prevalence of coronary disease in patients with peripheral vascular disease, routine coronary arteriography has not been found to be an efficient method of screening. The risk of perioperative cardiac morbidity can, however, be predicted with the Goldman cardiac risk index and other indices that stratify patients into risk groups on the basis of such parameters as age, type and setting of operation, general medical condition, cardiac history, and findings on electrocardiogram (ECG). The sensitivity of the Goldman index is not as high as originally suggested, and many patients at high risk for perioperative morbidity will escape detection. Recommendations for preoperative cardiac evaluation are discussed in the section on aneurysms but apply to all types of elective arterial reconstruction.

Acute Arterial Occlusion

Patients with acute arterial occlusion often seek treatment promptly because of the catastrophic nature of their symptoms. It is essential that the treating physician respond rapidly, because the process may quickly become irreversible. There is no fixed time after acute occlusion that ischemia is irreversible and reperfusion no longer indicated. The time interval depends on the preocclusive state, i.e., the status of the collateral circulation. Examination of the extremity is therefore the single most important determinant of urgency.

The P's of Acute Ischemia

The cardinal features of acute arterial ischemia each begin with the letter *p*. They are: pulselessness, pallor, poikilothermia (the tendency to drift toward ambient temperature), pain, paresthesia, and paralysis.

Once the diagnosis of acute arterial ischemia is made, the site of occlusion is localized by the absence of pulses on physical examination; for example, the presence of a palpable femoral groin pulse in the absence of a palpable popliteal pulse localizes the site of obstruction to the superficial femoral artery. Distal pulses are not palpable in patients with acute ischemia of an extremity, and they may or may not be detectable with Doppler ultrasonography. The presence of Doppler signals alone has little bearing on the presence or absence of ischemia. The degree of collateral circulation around the occlusion determines whether an audible signal is present and the urgency of revascularization. An objective determination of the adequacy of circulation can be made only by measuring the Doppler segmental pressures. An embolic occlusion in a previously normal extremity will be poorly collateralized, whereas a thrombotic occlusion of a stenotic vessel is usually well collateralized. Pallor is associated with decreased skin perfusion and often is accompanied by poikilothermia.

Pain is present in the vast majority of patients with acute ischemia. It is not a reliable criterion in an extremity without sensation, in a patient on mechanical ventilation, or in an unconscious patient. In these patients the other signs of acute ischemia must be relied upon.

The peripheral nerves are the most exquisitely sensitive tissues of the extremity. Therefore, the most important signs for evaluating the degree of ischemia are the neurologic signs, namely, paresthesia and paralysis. An extremity that is paralyzed from ischemia will certainly develop gangrene if left untreated. The earliest neurologic findings in the acutely ischemic lower extremity are in the distribution of the peroneal nerve and consist of hypesthesia in the first metatarsal space, inability to dorsiflex the great toe, and eventually a foot drop. Immediate revascularization is indicated when any of these neurologic signs is present.

Chronic Arterial Ischemia

Chronic arterial ischemia is almost always caused by atherosclerotic occlusive disease and is associated with tissue perfusion that is inadequate to meet the metabolic demands of the end organ. In some patients this occurs only when the demands are increased, such as during exercise, and produces the symptom of intermittent claudication, the most common complaint of patients with chronic arterial ischemia. Pain is experienced during exercise in the region of large muscle groups distal to an arterial occlusion or stenosis and gradually disappears as the activity ceases. Walking tolerance is diminished as the degree of arterial insufficiency increases. The symptom of intermittent claudication alone does not indicate that the extremity is immediately threatened. It does indicate, however, that the patient has a 10-year cardiovascular mortality rate up to fifteen times higher than those without peripheral arterial occlusive disease. This observation mandates careful assessment and treatment of underlying cardiovascular disease, including risk factor modification.

When metabolic needs cannot be met at rest, a state of critical ischemia is said to exist, and revascularization is indicated to avoid tissue loss. Symptoms may take the form of ischemic rest pain, nonhealing ulcerations, or frank gangrene. Ischemic rest pain is a constant burning type of pain that typically involves the distal foot, occurs when the foot is elevated, and is relieved when the foot is dependent. Rest pain can be distinguished from other types of severe foot pain because patients with other causes such as diabetic neuropathy do not have the same positional dependency.

Physical Examination

The diagnosis, site of occlusion, and extent of disease can almost always be determined by physical examination of the patient. A complete vascular examination involves inspection, auscultation, and palpation. The skin is inspected for color changes and integrity, the status of the nails, and the presence or absence of hair. The stethoscope is used over the neck, abdomen, and groin and the blood pressure is taken in both arms. Pulses are palpated in the upper extremity (brachial, radial, and ulnar) and the lower extremity (femoral, popliteal, posterior tibial, and dorsalis pedis). The abdomen is palpated for the presence of aortic enlargement. The legs are elevated and the color of the feet is noted. Normally perfused extremities do not blanch, whereas those with significant ischemia will blanch with elevation above the level of the heart. The legs are then lowered, and the time it takes for capillary filling to occur is noted. Finally, the legs are placed in a dependent position. Severely ischemic feet exhibit a cherry-red discoloration known as dependent rubor.

Diagnostic Vascular Testing

Noninvasive vascular diagnosis has emerged as an important tool for screening, preoperative assessment, and postoperative follow-up of patients with peripheral vascular disease. Improvements in instrumentation and techniques have resulted in the development of accurate methods of defining the anatomic and physiologic significance of lesions of the peripheral vascular system with little discomfort to the patient.

The Doppler ultrasonic flow detector is the most frequently used instrument in vascular diagnosis. It detects the frequency shift of ultrasound reflected off moving particles in the blood. The signal is processed in a variety of ways, ranging from an audible sound to a complex color-flow map. The combination of real-time B-mode scanning and Doppler spectral analysis, referred to as Duplex scanning, is the most accurate noninvasive method of detecting and following vascular lesions. In addition to providing an image of vascular lesions, it also permits an estimation of blood flow. The major disadvantages of Duplex scanning are the high cost of the equipment and the need for operator skill.

Plethysmography was once the mainstay of noninvasive diagnosis, but it has been replaced by the direct imaging modalities. Plethysmography detects changes in volume associated with cardiac contractions. Inferences concerning the vascular status are made from the character of the pulse wave. This modality is most often used for measuring segmental limb blood pressures and ophthalmic artery pressures.

Individual Testing Modalities

Doppler Segmental Blood Pressure Determination

In normal extremities the systolic blood pressure is slightly higher in the ankle than in the arm. This pressure differential is best expressed as a ratio, known as the ankle/brachial index (ABI). In normal volunteers the ABI averages 1.1; a value below 0.97 indicates some degree of arterial compromise. Patients with claudication usually have an ABI of 0.5 to 0.7. Patients with critical ischemia have an ABI of less than 0.5 and an absolute ankle pressure of less than 50 mmHg. When the ABI is reduced, segmental pressure measurements are performed by placing blood pressure cuffs on the high thigh, low thigh, midcalf, and ankle to identify the site of the lesion producing the pressure gradient. Patients with diabetes or end-stage renal disease may have vessels that are not compressible, rendering the test uninterpretable. In this situation, plethysmography is more accurate than pressure measurements. Additional information can be obtained by having the patient exercise, which exacerbates a pressure gradient in patients with vascular lesions but has minimal effects in those without.

Duplex Scanning

Duplex testing simultaneously evaluates the velocity of blood flow and displays this information in one of two ways. In the first, older method, the B-mode anatomic image is displayed on a screen and pulsed Doppler measurements are made as the operator positions a cursor at desired points within the vessel. The velocity of blood flow is interrogated within a small sample volume and expressed along an x-axis representing time. In the second, newer method, the machine continuously displays the velocity of particles in real time, expressed on the B-mode anatomic image in color, with slow or negative flow represented on one end of the color spectrum (e.g., blue) and rapid flow on the other (e.g., red). This technique is called a "color-flow" scan. The ease and accuracy of Duplex scanning has allowed the performance of carotid endarterectomy without preoperative arteriography in many institutions.

Arteriography

Significant advances in contrast media, catheter systems, and imaging techniques provide the essential information about arterial anatomy that makes definitive therapy possible. Today's ionic contrast agents use iodine attached to water-soluble carrier molecules. Many of their adverse side effects are related to a high concentration of sodium or meglumine. Contrast agents are excreted in the kidney, and patients with underlying renal dysfunction, particularly those with diabetes, are at higher risk for complications. Caution must be exercised when the serum creatinine concentration is greater than 2.0 mg/dL, and all patients must be well-hydrated before the study. The dye load is generally kept below 200 mL; normal individuals can tolerate volumes exceeding three times this amount. Neurotoxicity with seizures, cortical blindness, or frank stroke can develop, particularly with the concentrated (>60 percent weight/volume) sodium-containing agents. The newer agents have osmolalities of 600 to 900 mOsm, compared to 2000 mOsm in the conventional agents. The low-osmolality agents have clear advantages but cost considerably more than the conventional agents.

Idiosyncratic contrast reactions, with asthma, laryngeal edema or spasm, and cardiovascular collapse occur infrequently. These reactions are independent of dose. Although there is no reliable sensitivity test for these reactions, it is known that alcoholics, those allergic to iodine, and those with prior serious reactions are more likely to be affected. The absence of a reaction does not ensure that no reaction will occur with reexposure. Similarly, one reaction does not imply that subsequent exposure will result in another reaction. Since these reactions appear to be immune system-related, prophylactic steroids and antihistamines are administered to high-risk patients. Low-osmolality agents should be used in any patient who is at high risk for a contrast reaction.

Complications

The common femoral artery is the preferred site of arterial access. The brachial route can be safely employed in most individuals, but patients with small arteries (e.g., women) or calcific upper-extremity vessels (e.g., diabetics) have a higher rate of complications, such as localized hematoma formation and nerve palsy as well as arterial thrombosis. The percutaneous technique described by Seldinger is used by most angiographers. A guide wire is advanced under fluoroscopic guidance to the appropriate site, and a catheter is advanced over the wire. Iatrogenic damage from arterial catheterization may present as hemorrhage, dissection, thrombosis, embolus, false aneurysm, or arteriovenous fistula (Fig.

20-1). The reported incidence of these injuries varies from 1 to 2 percent for diagnostic procedures but is much higher when endovascular therapeutic modalities such as balloon angioplasty are performed.

Prompt diagnosis and intervention are necessary for successful treatment of these complications, which are increasing in incidence as the larger- catheter systems are used for endovascular manipulations. The critical risk factors are the size of the catheter and sheath, the site and method of puncture, and the duration of the procedure. Large, stiff catheter systems, axillary and brachial artery punctures, sheaths left in place for protracted periods, and use of anticoagulants or fibrinolytic agents are all associated with a higher incidence of complications. False aneurysms are more common in obese patients, in whom it is difficult to hold effective pressure over the cannulation site. Punctures of the profunda femoris and superficial femoral arteries are more frequently associated with false aneurysms. The thromboembolic complications require immediate operative intervention. Arteriovenous fistulae also require operative repair but are not as urgent. Femoral false aneurysms may be given a trial of compression with the Duplex probe, applying increasing pressure until flow is obliterated. If the patient is not on anticoagulation therapy, the aneurysm will frequently thrombose after 20 to 40 min of occlusion. Otherwise, operative repair is necessary.

Interpretation

The interpretation of arteriograms is a skill that must be gradually acquired. The older “cut films” clearly displayed bony landmarks; the newer digital subtraction images are more difficult for the novice to decipher. Since atherosclerotic plaques typically form at bifurcations, multiple views are necessary to view these bifurcations in profile. This is particularly important in viewing the bifurcations of the common carotid, femoral, and iliac arteries. Simultaneous biplane arteriography minimizes the chances of missing a significant lesion because of vessel overlap, but is not always available or possible. Several injections may be necessary while filming in different obliquities to open up bifurcations. Occasionally vessels do not fill and are considered occluded because the volume of contrast is inadequate or the exposure is too early after injection. If occlusion is diagnosed, contrast must opacify either the distal portion or a branch of the occluded vessel. This may require delayed filming. Digital recording techniques have improved visualization of distal vessels and reduced interpretive errors due to nonvisualization.

Other Procedures

Spiral Computed Tomography

Spiral computed tomography (CT) is a noninvasive modality that provides computer-reformatted images of a contrast-enhanced arterial lumen. The patient is advanced through a rotating gantry as contrast is administered intravenously. Rapid imaging produces a CT image similar to that obtained from conventional arteriography. Three-dimensional reconstructions of the data using shaded-surface or color display permit images that may be rotated and viewed from a variety of angles. Spiral CT has been most useful in the imaging of aortic aneurysms and plays a pivotal role as a preoperative tool in endovascular aortic reconstruction. Newer software provides reasonably good images of the carotid bifurcation and other vascular structures.

Magnetic Resonance Angiography

Magnetic resonance angiography (MRA) has been used as a less invasive, contrast-free alternative to conventional angiography and offers the added advantage of improved visualization of patent distal vessels when flow is minimal. The usual contraindications to magnetic resonance imaging exist, including the presence of metallic objects, such as intracranial vascular clips, electronic devices such as pacemakers, and claustrophobia. In addition, MRA requires immobilization for a relatively long period, especially when many vascular segments are interrogated. The cost of the test is considerable, and the most technologically advanced hardware is required to produce acceptable images. MRA has been successfully employed for the carotid bifurcation, the abdominal aorta, and the distal lower-extremity vasculature. Although Duplex ultrasonography still provides sufficient information at a considerable savings, MRA will have an increasing role in the planning of abdominal and lower-extremity arterial reconstructive procedures. MRA provides adequate visualization of venous structures inaccessible to ultrasound and is an excellent choice in the delineation of intracavitary venous disease, such as vena caval, renal, and mesenteric thrombosis.

ANEURYSMS

Classification

An aneurysm is an irreversible dilatation of an artery to at least one and one-half times its normal diameter. The dangers of these pulsatile masses whose “bright red blood ... spurted forth with much violence” have been known since the writings of Galen. Aneurysms may involve all layers of the arterial wall (true aneurysm) or only a portion of the vessel wall or surrounding tissue (false aneurysm). Aneurysms can be classified as nonspecific, traumatic, dissecting, mycotic, anastomotic, childhood, and those associated with pregnancy. Whatever the cause, once an aneurysm is formed it tends to enlarge and may ultimately produce serious and even lethal consequences.

Rupture occurs when the tangential stress at any point exceeds the tensile strength of the wall. Arterial wall strength is dependent on collagen, whose tensile strength is ordinarily far in excess of the wall tension. The collagen content of aneurysmal vessels is less than in atherosclerotic and normal vessels, however, placing a greater load on each fiber. The Laplace's law, which relates the tensile stress to wall pressure and radius, traditionally has been used to explain why large aneurysms rupture. The stress on the arterial wall is best expressed as pressure times radius divided by wall thickness, a modification of Laplace's law that is applicable only to thin- walled structures, in which the difference between inside and outside radius is negligible.

Degenerative Aneurysms

The most common type of aneurysm has been called atherosclerotic, but since the role of atherosclerosis in aneurysmal disease is unclear, the term “degenerative” is more appropriate. This view is supported by histologic evidence that demonstrates degeneration of the arterial wall. The intima usually is absent and replaced with compacted fibrin in multiple layers; the media has fragmented and reduced numbers of elastic lamellae; and, most important, there is focal loss of elastic tissue. Normal aortic tissue contains 12 percent elastin whereas aneurysmal aortic tissue has only 1 percent elastin. Biochemical data suggest that aneurysm pathogenesis may be related to a systemic connective tissue disorder. An imbalance between the two enzymes important in the metabolism of elastin, elastase (degradation) and a α_1 -antitrypsin (synthesis), has been identified in patients with aneurysms as compared to occlusive disease. This imbalance becomes even more pronounced in multiple aneurysms and ruptured aneurysms.

Traumatic Aneurysms

Many early descriptions of aneurysms dealt with traumatic or false aneurysms. Most traumatic aneurysms today are due to arterial catheterization or penetrating injuries. These lesions are characterized by a focal defect in the arterial wall, with the hemorrhage controlled by the surrounding tissues. With time, a fibrous capsule forms around the hematoma, but a definite risk of rupture is present because the surrounding tissues do not withstand arterial pressures and cannot contain the hemorrhage indefinitely.

Poststenotic Aneurysms

True aneurysms (involving all three layers of the normal arterial wall) can occur from the hemodynamic perturbations associated with an arterial stenosis. Aneurysms due to poststenotic dilatation are most often seen in thoracic outlet syndrome distal to a cervical rib, distal to coarctation of the aorta, and distal to aortic or pulmonary valvular stenoses (Fig. 20-2). These aneurysms do not have any preexisting defect but become dilated, possibly as a result of the increased lateral wall pressure suggested by Bernoulli's theorem. Once dilated, these arteries progressively enlarge according to Laplace's law.

Dissecting Aneurysms

The primary pathologic process in a dissecting aneurysm is a longitudinal splitting of the layers of the arterial wall. Whether the process begins with hemorrhage within the medial arterial layer or with a tear in the intima is unclear. The end result is a proximal intimal defect that allows blood to flow into a false channel that “dissects” between the intima and the inner two- thirds of the media. A site of reentry can occur distally, allowing blood to flow through the false channel. External rupture of the outer wall may occur, with exsanguinating hemorrhage. Hypertension is found in 75 percent of patients. Other, less common causes include Marfan's syndrome, Ehlers-Danlos syndrome, cystic medial necrosis, blunt trauma, and cannulation during cardiopulmonary bypass.

Mycotic Aneurysms

Mycotic aneurysms—i.e., aneurysms that are infected—can occur anywhere in the body as a consequence of either a blood-borne infection (intravascular) or an infection introduced from outside (extravascular). Blood-borne mycotic aneurysms can be further divided into preexisting aneurysms that become secondarily infected and mycotic aneurysms secondary to microbial arteritis. The classic type of this latter entity is the syphilitic aneurysm now rarely seen in the Western world, but formerly the cause of over 50 percent of all aneurysms. The only other bacteria with an affinity for arterial walls are *Salmonella* and *Staphylococcus*, which are now the most common organisms cultured from mycotic aneurysms. Infection of a preexisting aneurysm is unusual even though aortic aneurysms often grow bacteria from cultured intravascular thrombi.

The most common type of extravascular infected aneurysm follows a penetrating injury that contaminates the arterial puncture site and infects the resultant hematoma. Common in the era of bloodletting, the incidence of these aneurysms is increasing in intravenous drug users. Another form of extravascular infected aneurysm is the infected anastomotic aneurysm. Contamination can occur at the time of operation, from erosion of the graft material into the gastrointestinal tract, or from a contiguous hematoma secondarily infected from systemic sepsis.

Mycotic aneurysms should be suspected in patients with sepsis and with inflammatory changes around a pulsatile mass. Intravenous antibiotics are begun on the basis of results of blood cultures and clinical history. Since rupture carries a high morbidity and mortality, emergent operation is indicated after the necessary preparations.

Principles of Operation

The first operative decision in mycotic aneurysm is whether revascularization is necessary to prevent tissue loss. If so, a planned two- stage procedure is recommended. The initial stage creates an extraanatomic bypass in a remote operative field through uninfected tissues using autogenous tissue if possible. The second stage consists of resection of the infected arterial segment, debridement of surrounding tissues, and irrigation and drainage with a closed perfusion system of 0.1% povidone-iodine solution. If the clinical situation precludes a first-stage remote bypass or the need for revascularization is not certain, the infected aneurysm is approached directly. The same principles of wide excision, debridement, and irrigation are used. If revascularization is necessary, it is done through uninfected tissues immediately after the wounds are closed and all gowns, gloves, drapes, and instruments have been changed. Antibiotics are continued for many months, usually through a long-term indwelling intravenous-access catheter. Patients with *Salmonella* infections are placed on lifelong treatment.

Anastomotic Aneurysms

Since primary healing of a prosthetic anastomosis never occurs, anastomotic integrity depends solely on the strength of the suture line. Anastomotic aneurysms are the result of a separation between a graft and the host artery, forming a sac that becomes encapsulated with fibrous tissue. These false aneurysms contain no elements of the arterial wall. Most anastomotic aneurysms involve the common femoral artery after aortofemoral bypass.

Etiology

Any suture material that is degradable or easily broken can produce an anastomotic aneurysm. Silk has been abandoned as a vascular suture material because of its high incidence of late fatigue and subsequent anastomotic breakdown. Polypropylene and braided polyester are the most commonly used suture materials in vascular reconstructions and have not been associated with this complication.

In some instances the sutures may remain intact but an anastomotic false aneurysm occurs as they pull through the arterial wall. This may occur when placement of the sutures fails to incorporate sufficient amounts of arterial tissue, when there is excessive tension on the anastomosis from a graft that has been cut too short, and when there is degeneration of the artery. Infection of the graft and neighboring arterial wall is associated with anastomotic breakdown, especially when the organism is gram-negative. An artery that has undergone endarterectomy does not have the same tensile strength as the original artery, and care must be taken to place sutures such that they encompass enough tissue to prevent disruption.

Biologic grafts treated with formaldehyde or glutaraldehyde to prevent rejection are subject to aneurysmal degeneration. Some of the original prosthetic materials, such as Vinyon-N, nylon, and Orlon quickly lost tensile strength and were abandoned. Polyester has withstood the test of time and is the preferred material. Autogenous grafts may develop aneurysmal dilatation. Saphenous vein grafts have a 4 percent incidence of aneurysm formation when used in the extremities but a higher incidence when used in the aortorenal position, especially in children. For this reason, autogenous artery (usually hypogastric) is the conduit of choice for pediatric aortorenal bypass procedures.

Diagnosis

Patients with anastomotic false aneurysms usually present with a painless pulsatile groin mass. Rupture into surrounding tissues is unusual except when the aortic anastomosis is involved (10 percent). Duplex ultrasonography confirms the diagnosis in the case of peripheral false aneurysm, but computed tomography is the test of choice for an abdominal process. Arteriography is necessary to delineate the outflow and assist in the planning of a remedial operation. Anastomotic false aneurysms require repair because they may thrombose, embolize, or rupture.

Treatment

In the case of involvement of the femoral anastomosis, the graft often has retracted into the retroperitoneum. An interposition graft between the old graft limb and the femoral artery is required, and control of the graft limb should be obtained proximal to the inguinal ligament.

False aneurysms involving the proximal anastomosis of an aortic graft occur less frequently than false aneurysms at the distal anastomosis. Rupture into the peritoneal cavity or the duodenum may occur, and so repair is indicated. At operation, proximal aortic control is best obtained at the supraceliac level, approached through the lesser omentum after division of the crural fibers of the diaphragm (Fig. 20-3). In the absence of infection, the anastomosis is disconnected, the aorta debrided, and an end-to-end anastomosis is created just below the renal arteries with a new interposition graft. A flap of greater omentum is used to cover the new anastomoses.

When the aortic graft is infected, the graft should be removed and an extraanatomic bypass performed, routing the new graft through uninfected tissue planes. The procedure is performed in two stages. In stage one, bilateral axillobifemoral grafts are placed, with each distal anastomosis sewn at an uninfected site (distal superficial femoral or profunda femoris arteries). The aortofemoral graft limbs are then disconnected, the femoral defects are closed, and the ligated graft limbs are tucked beneath the inguinal ligament. In stage two, performed several days later, supraceliac aortic control is obtained and the proximal anastomosis is disconnected, removing the graft and oversewing the aortic stump. Aortic stump blowout, the major cause of postoperative mortality, occurs in a significant percentage of patients.

Aneurysms of Childhood

Aneurysms are rare in children. They are most often attributed to an underlying inherited disorder of connective tissue metabolism but can be acquired as a result of trauma or arteritis.

Infection

Infectious aneurysms are the most common pediatric aneurysms and usually involve the aorta. Bacterial endocarditis is the most common source of infection, and *Staphylococcus* and *Streptococcus* are the usual offending organisms. The aneurysms often develop in the aorta distal to a coarctation. Other predisposing conditions are umbilical artery catheters and bicuspid aortic valves. Prompt resection and reconstruction is indicated because of the high incidence of rupture.

Giant Cell Arteritis

This also affects the aorta and progresses to rupture and death. It is characterized pathologically by immune complex deposition in the vessel wall with complement fixation and neutrophil activation. This produces endothelial injury and transmural arterial ischemia from occlusion of the vasa vasorum. Degeneration and weakening of the vessel wall occur, with aneurysm formation and rupture.

Autoimmune Connective Tissue Disease

Children with these aneurysms exhibit the clinical features of an autoimmune process such as polyarteritis nodosa. These aneurysms are less than 3 mm in size and involve the arteries of the kidney, liver, and spleen. Rupture causes symptoms specific to the organ involved but may present as shock from intraperitoneal bleeding and the clinical scenario known as “abdominal apoplexy.” Ligation of the involved artery is indicated, with or without reconstruction.

Kawasaki Disease

Kawasaki disease or syndrome is also known as the mucocutaneous lymph node syndrome. The aneurysms occur in the axillary, brachial, iliac, and femoral segments. Coronary artery aneurysms occur in 20 to 30 percent of patients and are the most serious manifestation of this disease because of the risk of rupture and sudden death from pericardial tamponade.

Aneurysms Associated with Pregnancy

When aneurysms present during pregnancy, they often do so with rupture and shock, with a mortality rate of 65 percent. Splenic artery aneurysms are the most common, followed by aneurysms of the renal and iliac arteries. Pregnancy is also associated with aneurysmal dilatation in the aorta, presumably due to weakening of the arterial wall from the hemodynamic stresses of pregnancy and delivery. Matrix and elastic tissue abnormalities in the arterial wall have been described during pregnancy, making the vessel susceptible to rupture.

Abdominal Aortic and Iliac Artery Aneurysms

The infrarenal aorta is the most common site for the development of the nonspecific abdominal aortic aneurysm (AAA), the most common type of aneurysm presenting for treatment. Aortic aneurysms are typically fusiform in shape and usually arise below the origins of the renal arteries, extending a variable distance to and beyond the aortic bifurcation (Fig. 20-4). These aneurysms are found in 2 percent of the elderly population, and their incidence is increasing. Males predominate, with a ratio of 9:1. There is a definite familial tendency for the development of aneurysms. The tendency is sex-linked and autosomally inherited. The presence of a AAA in a female is almost always associated with aneurysms in family members. The estimated relative risk for first-degree relatives of affected individuals is 11.6 times greater than non-first degree relatives of similar age and sex. Screening with ultrasonography is indicated for relatives of patients with this lesion. Familial aneurysms tend to occur in individuals at a younger age and affect women more than does the noninherited variety.

Early attempts at repair consisted of wrapping the aneurysm with skin grafts or cellophane, injecting sclerosing agents around the aneurysm, intraluminal wiring, and endoaneurysmorrhaphy, a technique introduced by Matas at the turn of the century. The first successful resection was performed in 1951 by Dubost, who restored aortic continuity with a homograft. The modern “intrascaccular” (nonresective) method of repair was introduced by Creech and DeBakey in 1966. Current indications for operation include prevention of rupture, atheroembolization, associated occlusive disease, or pressure or erosion into contiguous structures.

Risk of Rupture

The decision to recommend resection must balance the immediate risk of operation against the risk of rupture, which is directly related to the size of the aneurysm. The natural history of the large aneurysm (> 6 cm) was defined by Estes over 30 years ago, and it is generally agreed that these aneurysms should be operated on unless severe comorbid conditions pose unacceptable operative risks or severely limit the patient's life expectancy. The classic study by Szilagyi and coworkers in 1966 clearly demonstrated that the repair of aneurysms larger than 6 cm in diameter prolonged patient survival. Although operations on smaller aneurysms did not prolong survival, in that report the operative mortality was 14 percent, therefore biasing the conclusions toward no operation. Application of current operative mortality rates (2 to 5 percent) would favor resection of even small aneurysms under the proper conditions. Crawford has reported a decline in mortality rates from 19.2 to 1.9 percent over a 25-year period despite a significant increase in operations on high-risk patients.

Less information is available about the natural history of small aneurysms (<6 cm), but it is clear that small aneurysms can rupture. Aneurysms grow an average of 0.4 cm/year, but growth occurs at an uneven rate. In the series by Szilagyi and coworkers, 19.5 percent of aneurysms smaller than 6 cm ruptured, and these data have been corroborated in more recent series showing a 20 percent risk for rupture in this group over a 5-year period. In a large autopsy study, Darling found a 25 percent rupture rate of aneurysms between 4 and 7 cm and a 9.5 percent rupture rate for aneurysms smaller than 4 cm.

Artery diameter is not a sufficiently accurate predictor of rupture to be used as the sole indication for operation. Cronenwett and coworkers analyzed 30 potential risk factors for a correlation with rupture and found that only diastolic hypertension, initial aneurysm size, and chronic obstructive pulmonary disease were important. The risk of rupture was only 2 percent when the diastolic pressure was less than 75 mmHg, the initial size was less than 3 cm, and there was no reduction in pulmonary function. By contrast, the risk of rupture was 100 percent when the diastolic blood pressure was higher than 105 mmHg, the initial size was greater than 5 cm, and the forced expiratory volume in 1 second (FEV₁) was less than 50 percent of the predicted value. Studies have tried to correlate the size of the aorta relative to the size of the patient and the risk for rupture. Ouriel and associates used computed tomography to measure the transverse diameter of the L3 vertebral body and calculated a ratio of aortic to vertebra size. Although aneurysms were observed to rupture with diameters as small as 3.5 cm, not one aneurysm ruptured with an aneurysm-to-vertebra ratio of less than 1.0. Surgeons should have an elective mortality rate well below 5 percent to justify operation in patients with small aneurysms.

Diagnostic Evaluation

Most aneurysms (75 percent) are discovered when still asymptomatic, either as a pulsatile mass on physical examination or unexpectedly during the course of an evaluation for an unrelated condition.

Physical Examination

This remains a reliable diagnostic method that is accurate in the vast majority of cases. The infrarenal aorta lies at the level of the umbilicus, so the mass should be felt around the navel. If the lateral borders converge below the costal margins, the aneurysm is probably infrarenal. Thin patients, particularly those with a lordotic spine, have a prominent aorta that can be mistaken for an aneurysm. An aneurysm can be distinguished from a tortuous aorta because the latter lies to the left and is moveable. An aneurysm can be felt to the right of the midline as well as to the left and feels expansile. Size can usually be estimated in a thin patient, and when the diagnosis of aneurysm is made in an obese patient by palpation, the aneurysm usually is larger than 6 cm.

Ultrasonography

The proven accuracy, safety, and low cost of ultrasonography make this the imaging procedure of choice for the diagnosis and follow-up of aneurysms (Fig. 20-5). The variability of the test when performed by the same technician is only 3 mm. Ultrasonography also can be used to evaluate blood flow in the renal and visceral arteries. It may not give sufficient information about the juxtarenal aorta or the iliac arteries and may be technically difficult in obese patients and in those with excessive intestinal gas.

Computed Tomography

This imaging technique provides an accurate characterization of the entire aorta. New software allows the data to be displayed in a three-dimensional image (Fig. 20-6). Its accuracy is not influenced by bowel gas or obesity, and it provides information about the character and thickness of the aortic wall, the level of the renal arteries relative to the proximal cuff, and the iliac arteries (Fig. 20-7). It is particularly helpful in evaluating patients with symptoms. As little as 10 mL of blood can be detected outside the lumen of the aorta. CT scanning is used in the postoperative period for the evaluation of suspected aortic graft complications.

Magnetic Resonance Imaging (MRI)

MRI can provide a much more detailed image than ultrasonography or CT. Three-dimensional presentations, showing the aneurysm's lumen and surface anatomy in four separate views, allow the surgeon to visualize the neck, renal arteries, and relationships to other periaortic structures. Unfortunately, progress in vascular imaging with this modality has been limited by lack of appropriate software support and high cost as well as the limitations of scanning obese patients and those with pacemakers.

Arteriography

Although not accurate in determining the size or the presence of an aneurysm, the information provided by arteriography is helpful in defining associated vascular anatomy, particularly the renal arteries (Fig. 20-8). Many vascular surgeons recommend routine arteriography before aneurysm resection, but the issue of whether the information learned is worth the cost, discomfort, and potential risk to the patient is unresolved. Arteriographic findings caused a modification of the procedure in 13 to 75 percent of patients, usually because of a pathologic condition of the iliac, mesenteric, or renal artery (Fig. 20-9). If arteriography is not used, a selective approach should be taken. Absolute indications for arteriography include: (1) symptoms of mesenteric ischemia, (2) hypertension or renal dysfunction, (3) horseshoe kidney, and (4) claudication or other signs and symptoms of coexistent lower-extremity occlusive disease.

Preoperative Cardiac Screening

Cardiac, pulmonary, hepatic, renal, and hematologic screening is essential in selecting patients for operation, deciding what imaging techniques are indicated, and selecting an appropriate incision. Patients with uncorrected coronary artery disease (CAD) undergoing aneurysm resection have an increased risk of myocardial infarction (5 to 10 percent), ischemia-related pulmonary edema (5 to 10 percent), and cardiac death within 30 days (5 to 10 percent). These risks contrast significantly with the 1 to 3 percent cardiac risk in patients without CAD. Preoperative cardiac revascularization has repeatedly been shown to reduce the operative risk in patients with CAD who undergo AAA repair.

Selecting patients at risk for cardiac events has been a difficult task. The Goldman cardiac risk index misses a large proportion of patients at risk. Exercise-induced ECG ST-segment depression misses a smaller but still significant group. Coronary arteriography is costly and carries some risk, and it is not appropriate as a screening procedure in an asymptomatic patient. Resting echocardiography, while missing a subset of patients with occult coronary disease, identifies most patients with impaired cardiac reserve. Invasive coronary studies are not always necessary to discover these high-risk patients because of recent advances in the areas of dobutamine stress echocardiography and dipyridamole-thallium myocardial scintigraphy. In a dobutamine stress echo, changes in myocardial wall motion are evaluated with increasing doses of dobutamine. The dipyridamole-thallium scan is probably the most widely used preoperative test to gauge myocardial risk. Interpretation of this test is based on selective regional uptake of isotope as a marker of perfusion. Normally perfused myocardium has a homogeneous uptake, old infarcts do not take up isotope even on delayed scans, and ischemic areas of viable myocardium exhibit delayed uptake, referred to as redistribution. Patients with redistribution are at increased risk (15 to 35 percent) for a postoperative cardiac event.

The following approach minimizes perioperative cardiac risk in patients undergoing aneurysm resection. Patients without a cardiac history and with a normal ECG (30 to 40 percent of AAA patients) can proceed to operation without further workup. Patients with clinically evident but stable cardiac disease (50 to 60 percent of AAA patients) should have a dobutamine echocardiogram or a dipyridamole-thallium study performed. An abnormal provocative test mandates coronary revascularization (coronary artery bypass or percutaneous angioplasty) before aneurysm repair when there is three-vessel disease or left main stenosis. For patients with clinically severe cardiac disease, catheterization and coronary revascularization can proceed without initial noninvasive testing. AAA repair can follow percutaneous coronary procedures almost immediately, and coronary artery bypass or valve repair can follow within 4 to 6 weeks. In the rare case of a symptomatic AAA and severe coronary disease, aneurysm resection has been performed concurrently with coronary artery bypass grafting.

Operative Management

Significant advances have been made in the anesthetic and medical management of these patients. Healthier patients can be admitted on the day of surgery, but higher-risk patients are brought into the hospital earlier for volume loading and preoperative cardiac optimization. A bowel preparation may be used. Epidural catheters are used to reduce the amount of anesthetic agent required, to allow for prompt extubation, and to control pain in the postoperative period. Swan-Ganz catheters or intraoperative transesophageal echocardiography are used in patients with a cardiac history. Patients are volume-loaded prior to clamping with crystalloid after vasodilatation with nitroglycerin. Blood is returned to the patient with an autotransfusion system to reduce the need for banked blood. Cefazolin or an antibiotic equivalently effective against *Staphylococcus aureus* is administered immediately before operation and continued until central venous access lines are removed.

Incision

The choice between a midline transperitoneal or a retroperitoneal approach to the infrarenal abdominal aorta is usually a matter of surgeon preference. Each offers advantages, and the preferable approach varies from patient to patient. The transperitoneal approach is obligatory when access to the right renal artery is required, when there is extensive disease of the right iliac system, and when access to abdominal contents is necessary. The retroperitoneal approach has advantages when there are extensive intraperitoneal adhesions, gastrointestinal or urinary tract stomata, severe chronic pulmonary disease, or the need for extensive suprarenal exposure.

Exposure

A thorough abdominal exploration is performed. The extent of the aneurysm and the status of the iliac arteries are determined by palpation, taking care to avoid extensive manipulation. The intestines are packed off to the right, retracted with a mechanical device fixed to the operating table. An incision is made in the retroperitoneum to the left of the base of the duodenum and carried proximally and distally over the right iliac artery (Fig. 20-10). The neck of the aneurysm is exposed below the left renal vein, which can be divided if necessary to the right of the adrenal vein to provide additional proximal exposure. Dissection over the aorta is confined to the anterior surface until the neck is clearly identified. Circumferential aortic control is infrequently necessary and is associated with an increased risk of atheroembolization. Instead, the anterior and lateral aspects of the normal aortic neck are cleared, palpating the bony spine to confirm adequate exposure for the application of an occluding clamp. If iliac artery aneurysms are present, the incision in the retroperitoneum is carried distally on the right, and the external iliac and hypogastric arteries are dissected free. The left iliac bifurcation is controlled after lateral mobilization of the sigmoid colon. The inferior mesenteric artery is examined for pulsation and size and is not ligated at this time.

Graft Insertion

The patient is given intravenous heparin and the activated clotting time is checked. The proximal aorta and iliac arteries are clamped, and the aneurysm is opened longitudinally. Intraluminal debris is evacuated, and bleeding from lumbar vessels is controlled with mattress sutures. If the inferior mesenteric artery is bleeding briskly, it is ligated at this point. The aortic cuff is examined. A well-defined sewing ring is usually present, and the graft can be sutured in place without transecting the aorta (Fig. 20-11), with the posterior wall of the anastomosis sewn from within the aneurysm sac. The distal portion of the aneurysm is inspected, and either a bifurcation or a straight tube prosthesis is selected. Anastomoses are sewn with continuous polypropylene suture. The proximal anastomosis is tested for leaking before distal-end suturing by removing the aortic clamp once and repositioning it on the graft. Any bleeding areas are reinforced with pledgeted sutures after the aorta is reclamped. The graft limbs must be passed under the ureters if the distal anastomoses are done beyond the iliac bifurcations. Just before the distal clamps are released, the iliac artery clamps and the graft clamp are removed to flush out any residual debris and evacuate air. The anesthesiologist is notified that clamp release is imminent, and necessary adjustments in pH are made. The proximal clamp is released slowly, with monitoring of the arterial pressure. Rapid release without adequate fluid and acid-base resuscitation may cause significant hypotension.

Prevention of Colon Ischemia

Visual inspection of the colon is notoriously inaccurate as a predictor of ischemia. If the inferior mesenteric artery is excessively large or the circulation to the sigmoid colon is questionable, Doppler ultrasonography or photoplethysmography can be used to assess flow. If a signal is present on the antimesenteric surface, the colon should be viable and no further action is indicated. Otherwise, measures must be taken to restore colonic blood flow by reimplanting the inferior mesenteric artery into the body of the graft using the Carrel patch technique. A cannula may be inserted into the inferior mesenteric artery orifice and a stump pressure measured. If the pressure is less than 40 mmHg, inferior mesenteric artery reimplantation is advocated. This is an uncommon occurrence, particularly when maintenance of perfusion to one hypogastric artery is possible.

Closure

Protamine sulfate is given to reverse the action of the heparin, and the activated clotting time is rechecked. When hemostasis is satisfactory, the wall of the aneurysm sac is sutured over the prosthetic graft to prevent adherence and subsequent erosion of the intestine into the graft or anastomoses. If this cannot be accomplished, a greater omentum flap can be brought in a retrocolic fashion and sutured to the surrounding tissues to cover the graft. The abdominal wound is then carefully closed, because there is a high incidence of late postoperative incisional hernias.

Endovascular Repair of Aortic Aneurysms

A new approach has been investigated for the repair of aortic aneurysms, using endovascular techniques to place a covered stent within an aneurysm, effectively excluding the aneurysm from the circulation. The components of an endovascular system include the graft itself, the stents that secure the graft to the arterial wall, and the delivery system used to thread the prosthesis to its appropriate anatomic location. Several configurations of endovascular grafts exist for use in the repair of aortic or aortoiliac aneurysms. The graft can

be tubular or bifurcated, but the anatomic configurations of most aneurysms require the use of grafts to the iliac arteries. Even when the aneurysm is limited in extent to the aorta, a bifurcated device usually is necessary because of the absence of a suitable distal cuff for adequate fixation and sealing of a tube graft. The graft material of most endovascular prostheses is polyester, and stents are made of stainless steel or nitinol and may be self-expanding or balloon- expandable. The fixation may be through barbs on the outer aspect of the stent, but some stents rely on friction alone. The delivery systems are cumbersome, but newer devices will be more “user friendly.” At present, the outer diameter of the introducers is from 18 to 24 French, allowing placement through most but not all femoral and iliac arteries with a groin cutdown.

The ultimate goal of all endovascular techniques for aortic aneurysm repair is to prevent the risk of aneurysm rupture through a less invasive procedure than standard open aneurysmectomy. The major problem of endovascular grafting is the leakage of blood into the aneurysm sac at the proximal or distal anastomosis. Immediate thrombosis of graft limbs also has occurred; this risk may be minimized by the placement of continuous stents along each iliac limb.

Controlled clinical studies are now in progress to compare the results of endovascular and open repair of aortic aneurysms. The results of these trials are be expected to allow governmental approval of appropriate devices, minimizing the morbidity and mortality of abdominal aortic aneurysm resection.

Ruptured Aortic Aneurysms

The mortality rate following rupture of an abdominal aortic aneurysm approaches 90 percent. Of patients who reach the hospital alive, the mortality rate averages 50 percent or more, depending on the status of the patient, the type of rupture, the experience of the surgical team, and the delay in controlling hemorrhage. Patients present with a variety of dramatic symptoms, ranging from generalized abdominal pain, severe back or flank pain, and/or circulatory collapse. Diagnosis may be delayed because this condition may mimic a perforated peptic ulcer, renal or biliary colic, or a ruptured intervertebral disc. The rupture may be contained in the retroperitoneal space, flow freely into the peritoneal cavity, or, in rare instances, flow into another anatomic structure such as the vena cava or duodenum.

Emergency Department Management

This is a true emergency, and prompt diagnosis, resuscitation, and control of hemorrhage are critical. Factors associated with an increased mortality are a history of cardiac disease, hypotension upon arrival, renal insufficiency, and inexperience of the surgical team.

The amount of time spent on diagnosis depends on the condition of the patient. When a patient presents in shock with a pulsatile mass and a known history of AAA, no further testing is needed and the patient is taken to the operating room immediately. An obese patient with normal vital signs and severe abdominal, back, or flank pain should undergo an emergency CT scan. If the patient is in shock and the diagnosis is in question, a portable ultrasound examination can be done in the emergency department. Operation should not be delayed in order to resuscitate the patient. Instead, the patient should be brought to the operating room while rapid infusions of fluid are continued and the patient is prepared for exploration.

Operating Room Management

The operating team should be scrubbed and the patient's abdomen prepared before anesthesia is induced. This is a period of significant risk for hypotension because of the vasodilatory effects of the anesthetic agents. If there is free intraperitoneal blood, aortic control is rapidly achieved at the diaphragm using a right-angled retractor to compress the supraceliac aorta against the vertebral column. This maneuver allows the anesthesiologist to resuscitate the patient. The retroperitoneum should not be opened before proximal aortic control has been achieved if the hematoma is contained. The right-angled retractor can be replaced with a well-placed supraceliac aortic clamp, or the retroperitoneal hematoma can be opened while an assistant holds pressure on the supraceliac aorta and the neck of the aneurysm can be carefully clamped below the renal arteries. The aneurysm is opened when iliac control is achieved. We prefer to leave a secure clamp on the supraceliac aorta and suture the proximal anastomosis without a clamp on the infrarenal aorta. This reduces the incidence of left renal vein injury. Heparin is given after the completion of the proximal anastomosis, the iliac arteries are back-bled and reclamped, and the distal end of the graft is sutured to the appropriate outflow vessel (distal aorta, iliac or femoral arteries).

Complications

Renal Failure

Renal failure (rise of >20 percent in blood urea nitrogen and/or creatinine) occurs in roughly 6 percent of elective aneurysm repairs and 75 percent of ruptured AAAs. Acute tubular necrosis can follow prolonged hypotension from excessive intraoperative blood loss or from declamping hypotension. It is probably the most common form of renal failure after resection of a ruptured AAA but is uncommon in elective repairs. Atheroembolization from the juxtarenal segment during aortic clamping is the most common cause of renal failure in elective repairs. This complication can be avoided by proper selection of a proximal clamp site. In cases in which degenerative changes in the cuff are present or the aneurysm is “juxtarenal” the proximal clamp should be placed on the supraceliac aorta to avoid renal damage (Fig. [20-12](#)).

Gastrointestinal Complications

Ischemic colitis involving the sigmoid colon is a dreaded complication of operations on the abdominal aorta with an incidence of 1 to 6 percent. The onset of bloody diarrhea, abdominal distention, leukocytosis, or signs of peritonitis should prompt immediate sigmoidoscopy. In most instances the injury is mucosal and the process is self-limited. When the muscularis is involved, a segmental stricture may occur and at a later date might require resection. In cases of full-thickness involvement, immediate resection of the involved intestine is required, with creation of an end colostomy and mucous fistula, before the prosthetic graft is contaminated.

Spinal Cord Ischemia

Paraplegia after operations on the infrarenal aorta is rare and is usually caused by pelvic devascularization; it is often accompanied by colon ischemia and necrosis of the skin of the buttocks. This is to be differentiated from the paraplegia that follows operations on the lower thoracic aorta after interruption of the greater medullary artery of Adamkiewicz (Fig. [20-13](#)). Although the complication is not entirely preventable, its incidence can be reduced by maneuvers that prevent atheroembolization into the hypogastric arteries and ensure the perfusion of at least one hypogastric artery and therefore of the iliolumbar and lateral sacral arteries. Reimplantation of the inferior mesenteric artery should be considered in patients who have occlusion of both hypogastric arteries.

Iliac Artery Aneurysms

Iliac artery aneurysms are usually dealt with in conjunction with operations for abdominal aneurysms but rarely can occur independently. The common iliac artery is involved in 90 percent of cases, with the remaining 10 percent involving the hypogastric arteries. The external iliac artery is never associated with degenerative aneurysmal disease. Iliac aneurysms are not often identified on physical examination, and imaging techniques are necessary for diagnosis. A hypogastric artery aneurysm may present as a pulsatile mass palpable on rectal examination. Isolated iliac aneurysms do rupture, and resection is indicated when the size exceeds 3 cm in diameter. Follow-up ultrasonography at 6-month intervals is recommended for patients who do not undergo operation.

Isolated unilateral iliac artery aneurysms can be approached through an ipsilateral retroperitoneal flank incision. Common iliac aneurysms are treated with interposition prosthetic grafts. Hypogastric aneurysms usually can be ligated proximally, and the outflow vessels can be controlled from within the sac. Bilateral aneurysms are best treated with concomitant replacement of the infrarenal aorta and insertion of a bifurcation graft (Fig. [20-14](#)).

Visceral Artery Aneurysms Renal Artery Aneurysms

Renal artery aneurysms are uncommon. They may present with rupture or distal embolization but are often identified as asymptomatic rings of calcification on an abdominal plain film. Women are more likely to be affected because of the association with fibromuscular dysplasia. The risk of rupture is greatest during pregnancy and for aneurysms larger than 1.5 cm in patients with hypertension. Direct repair with resection of the aneurysm and interposition bypass grafting is the procedure of choice for the extraparenchymal lesion. Renal salvage is possible in more than 90 percent of patients. Intraparenchymal lesions often require total or partial nephrectomy, especially if rupture has occurred. In some patients, an ex vivo repair can be performed for an intraparenchymal lesion prior to rupture.

Splanchnic Artery Aneurysms

See Chap. [33](#).

Peripheral Aneurysms Popliteal Artery Aneurysms

Popliteal artery aneurysms are the most common peripheral aneurysm (70 percent), but their incidence is low. They are often bilateral (50 to 75 percent); are associated with other aneurysms, particularly in the abdominal aorta (50 percent); and occur predominantly in males between the ages of 50 and 70 years. Retrospective studies have shown that roughly 60 percent of untreated patients develop the complications of thrombosis or distal embolization, and 20 percent require amputation. Some patients who suddenly develop symptoms present with inoperable situations because of occlusion of the distal arterial tree by embolic material. Early series reported excellent limb salvage rates after ligation, which attests to the importance of these embolic events (Fig. [20-15](#)).

Physical examination alone may not distinguish a popliteal artery aneurysm from arterial ectasia or a Baker's cyst. Imaging of the popliteal space with ultrasonography is necessary for diagnosis; arteriography should be performed to define the runoff bed (Fig. [20-16](#)).

Indications for operation are debated, but it is clear that long-term results are better when these aneurysms are treated before the onset of symptoms. Reports showing that aneurysms larger than 2 cm in diameter were more likely to cause complications than smaller ones have prompted many to advocate repair when the aneurysm reaches that critical size. However, the claim has also been made that thrombosis is more common in the smaller aneurysms. It appears that both statements may be valid—i.e., large aneurysms embolize and small aneurysms thrombose. In our view, then, repair is indicated in any patient with a popliteal aneurysm without serious comorbidity. Operative management can be done with excision and interposition bypass grafting through a posterior incision, or ligation and bypass around the popliteal space through medial incisions. Saphenous vein should be used in most instances.

Femoral Artery Aneurysms

Most of these uncommon aneurysms are found on physical examination. Diagnosis can be verified with ultrasonography. Most femoral artery aneurysms are degenerative in nature and associated with aortic dilatation, but mycotic and traumatic aneurysms also can occur. Most patients develop symptoms of an enlarging groin mass, local pain, venous obstruction, distal embolization, or thrombosis. All symptomatic aneurysms and asymptomatic aneurysms larger than 2.5 cm in diameter should be repaired. The operation of choice is resection with prosthetic graft replacement.

ACUTE ARTERIAL OCCLUSION

Acute occlusion of the arterial supply to an extremity often is an emergency. In patients without extensive collateral circulation, or in circumstances in which the occluded artery is the only vessel supplying the end organ, progression to irreversible ischemia may begin 6 h after the event. Prompt diagnosis and treatment in such situations is imperative, both to reestablish flow through the initial site of occlusion and to prevent propagation of the thrombotic process to the distal arterial tree and venous system.

Restoration of perfusion to the threatened limb is the principal objective at the time of initial evaluation. Nevertheless, establishing the nature of the occlusive process is an important secondary goal and should be attempted whenever feasible, since treatment options are affected by the cause. The differential diagnosis of acute arterial occlusion includes embolism, trauma, and thrombosis. Emboli may originate from the heart or from a more proximal artery; in either case there may be no prior or concurrent history of the underlying disorder. Injuries most often associated with arterial occlusion are posterior knee dislocation, long bone fractures, penetrating trauma, and catheter-related iatrogenic trauma. Thrombosis may occur in the presence of an atherosclerotic lesion or, less commonly, in the presence of an aneurysm.

Pathophysiology

Acute arterial occlusion results in anoxic ischemia of the tissues supplied by the involved arterial segment. Gangrene develops in approximately 50 percent of patients with this condition, depending on the site and length of occlusion and the presence of collaterals. The initial clinical features of pain, paresthesia, and paralysis reflect the greater susceptibility of nerves to ischemia in comparison to other structures. Striated muscle is only slightly less susceptible to ischemia, and therefore loss of nervous function heralds impending muscle necrosis. Tissue death typically begins 6 to 8 h after the embolic event, but this interval can vary markedly, depending on the presence of collateral circulation; in some instances tissue death may not occur at all. Thrombus does not propagate beyond branch points when flow through collaterals is sufficiently great, and the clinical findings are less serious. In the presence of stasis, however, thrombosis of the arterial and venous system distal to the initial occlusion does proceed, rendering attempts at operative and pharmacologic therapy futile.

Clinical Manifestations

The most significant features of acute arterial occlusion and the resultant ischemia can be summarized as the six *p*'s: pain, paralysis, paresthesia, pallor, pulselessness, and poikilothermia (Table [20-2](#)). Over 75 percent of patients with acute arterial occlusion experience pain as the presenting symptom. Pain may be absent from the clinical syndrome because of diabetic neuropathy, adequate collateral flow resulting in less severe ischemia, or rapid progression to advanced ischemia, with immediate anesthesia. Paresthesia and paralysis are the most critical features to evaluate in the patient with acute arterial occlusion. When present, these findings indicate anoxia of the sensory and motor nerve endings of the extremity, the structures most susceptible to ischemia. Sensory fibers are slightly more sensitive than motor fibers, hence the clinical observation that paresthesia usually precedes paralysis. Paralysis also may be the result of striated muscle necrosis in more advanced ischemia. The presence of paresthesia and paralysis is an ominous finding, indicating that the limb will almost certainly develop gangrene if the underlying condition is not alleviated within 6 to 8 h, and signals the need for rapid treatment. By contrast, the patient who does not exhibit these findings is at a much reduced risk of developing ischemic necrosis acutely.

Reduction in blood flow to an extremity results in a pale appearance of the limb and is frequently associated with the sixth *p*, poikilothermy, or coolness. Discolored, mottled skin that fails to blanch in response to digital pressure is an indicator of irreversible ischemia and is due to extravasation of blood into the dermis from ruptured capillaries. The level at which temperature and color changes occur can provide information regarding the level of the arterial occlusion; tissue ischemia usually develops one joint level below the segment of occluded artery. For example, an embolus occluding the origin of the superficial femoral artery produces ischemia distal to the knee joint.

The absence of pulses supports the diagnosis of acute arterial occlusion but does not prove it with certainty, since pulses may be absent chronically in the patient with peripheral vascular disease. The examination of the pulses also assists in locating the level of occlusion. In general, the occlusion can be localized to the segment of the arterial tree immediately proximal to the site of pulselessness. For instance, a patient with a palpable or exaggerated femoral pulse and an absent popliteal pulse can be assumed to have an occlusion of the superficial femoral artery.

In addition to the clinical findings, evaluation of the muscle turgor in the affected limb yields important information regarding the severity of ischemia and the degree to which the changes are reversible after reperfusion. The muscles are soft immediately after the onset of ischemia. With time, the muscles develop edema, which is associated with a doughy feeling on physical examination. Necrosis occurs at a more advanced stage, and the muscles feel stiff and hard; when this occurs, the ischemic changes are irreversible, regardless of therapy.

Diagnosis

In some instances the diagnosis of acute arterial occlusion can be made with relative certainty on the basis of the patient's history and physical examination, and no further workup is required before definitive treatment is initiated. An example of this is a patient who presents with the sudden onset of a cold, painful leg and is found to have an absent femoral pulse and atrial fibrillation. Groin exploration and femoral embolectomy without preoperative arteriography may be appropriate in this case. In the majority of patients, however, the presence or the precise location of an acute arterial occlusion may be uncertain, and further evaluation with arteriography is necessary to optimize treatment.

Acute arterial occlusion can be confused with nerve root compression, deep venous thrombosis, phlegmasia cerulea dolens, and infectious processes. Rarely, acute dissection of the thoracic aorta mimics multiple visceral and extremity emboli if these vessels are occluded by the false lumen. Noninvasive studies such as segmental arterial pressures can be helpful in establishing the diagnosis and localizing the site of occlusion. Arteriography is performed when additional information is required, for instance, to determine the appropriate sites of inflow and outflow for a bypass graft. If the ischemia is severe, however, the urgency of revascularization may preclude preoperative arteriography.

Arterial Embolism

Arterial emboli can be divided into two categories: emboli that lodge in large-diameter vessels, such as the common femoral artery, the vast majority of which are of cardiac origin; and atheroemboli to smaller vessels, such as branches of the digital arteries, that invariably originate from a plaque or thrombus in a more proximal vessel.

Cardiac Emboli Etiology

An embolus from the heart may be the first indication of a serious underlying cardiac disorder and may occur in three clinical settings: atrial fibrillation, myocardial infarction, and valvular disease. Appropriate evaluation and treatment of the cardiac disease, when necessary, must be undertaken concurrently with the therapy of the arterial occlusion, and the severity of the underlying cardiac condition must be taken into consideration when deciding on therapy for the arterial occlusion.

Cardiac emboli in the patient with mitral stenosis may result either from valvular vegetations or from mural thrombi that form in a dilated left atrium. Atrial fibrillation, occurring alone or in conjunction with mitral stenosis, predisposes to the formation of mural thrombi, which may then embolize to the peripheral circulation. Similarly, transmural myocardial infarction is associated with the formation of mural thrombus on the subendocardial surface overlying the infarcted ventricle, usually 2 to 3 weeks after the cardiac event. Rare sources of arterial emboli originating in the heart include bacterial endocarditis, atrial myxoma, prosthetic heart valves, and paradoxical embolization originating in the veins and passing through a patent foramen ovale. A diligent search must be made for the embolic source in all patients, including the use of electrocardiography, echocardiography, and Holter monitoring. Despite a thorough evaluation, however, the source cannot be identified in a significant percentage of patients.

Distribution

The two factors that influence the site in the arterial tree in which an embolus will lodge are patterns of blood flow and changes in vessel diameter. Consequently, 70 percent of cardiac emboli lodge in the arteries of the lower extremities, 13 percent in the arteries of the upper extremities, 10 percent in the cerebral circulation, and 5 to 10 percent in the visceral circulation. In addition, emboli generally lodge at arterial branch points, where the vessel diameter is abruptly reduced. Common sites in the lower extremity include the bifurcation of the abdominal aorta, the common iliac artery, the common femoral artery, and the popliteal artery (Fig. [20-17](#)). A corresponding pattern is seen in the arteries of the upper extremity, with involvement of the brachial artery at its bifurcation into the radial and ulnar arteries or, less commonly, at the takeoff of the profunda brachii artery. The ischemic insult is more severe when the bifurcation of a vessel is occluded, because collateral circulation cannot be supplied by a patent branch vessel.

Evaluation

Once the diagnosis of acute arterial occlusion has been established, it is desirable to differentiate between embolism and thrombosis as the underlying cause, since the treatment options may differ accordingly. A patient with embolic occlusion of an otherwise normal vessel (Fig. [20-18](#)) may only require thromboembolectomy under local anesthesia, whereas a patient with thrombosis of an atherosclerotic artery will likely require a bypass graft around the involved segment, under regional or general anesthesia. Clinical findings suggestive of an embolic cause include the presence of cardiac arrhythmia, myocardial ischemia, and valvular disease; the absence of factors predisposing to atherosclerosis; and the absence of prothrombotic hematologic disorders. The presence of normal pulses on the contralateral side is strongly indicative of an embolic source.

Arterial Atheroemboli Etiology

The vast majority of emboli that originate in arteries are fragments of an ulcerated atherosclerotic plaque that become dislodged and travel downstream, lodging in a more distal artery. These fragments may be composed of cholesterol crystals or of fibrin-platelet debris that deposits on the surface of the plaque and is then displaced into the circulation. The artery of origin, usually the aorta, iliac, or femoral vessels, is often some distance from the artery in which the embolus finally lodges, since these fragments are usually tiny and do not become trapped until they reach a small vessel (Fig. [20-19](#)). It is rare for large vessels such as the common femoral or popliteal arteries to become occluded by atheroemboli, except during operative manipulation of a diseased abdominal aorta. Fragments of intraluminal thrombi, such as those that are seen in association with aneurysms, also may dislodge and occlude distal vessels. This is much less common than atheroemboli but is typically seen in association with popliteal aneurysms. In either case, patients may experience repeated embolic episodes that are individually minor; the additive effects, however, may be quite significant.

Clinical Syndromes

The most common example of atheroemboli is the "blue-toe" syndrome. Patients with this syndrome develop severe ischemia of the toes and forefoot in the presence of palpable pulses. The superficial femoral artery, the popliteal artery, and at least one tibial vessel are usually patent in order for emboli to travel to the digital arteries; therefore, patients with blue-toe syndrome generally have palpable pedal pulses. This physical finding enables the clinician to differentiate this disorder from lower-extremity gangrene resulting from large-vessel atherosclerotic disease. Renal failure or mesenteric ischemia may accompany digital gangrene if the thoracic upper abdominal aorta is the source of emboli. If the digits of the upper extremity are involved, the subclavian artery usually is the source, with disease caused by proximal atherosclerosis or more distal poststenotic aneurysm formation secondary to thoracic outlet obstruction.

Evaluation

Atheroemboli should be suspected when the physical examination reveals digital ischemia in a patient with palpable pulses. The location of the embolic source can be estimated on the basis of the sites where the emboli have lodged. The source must be proximal to the site of the embolus; bilateral emboli to the feet indicate a source above the aortic bifurcation, and repeated unilateral emboli suggest a source distal to the bifurcation. A thorough arteriographic evaluation is required to confirm the diagnosis and to identify

precisely the offending lesion in preparation for operative intervention.

Arterial Thrombosis

Etiology

Spontaneous acute arterial thrombosis occurs most commonly in the presence of an underlying stenosis caused by atherosclerotic disease. The thrombotic event may be precipitated by plaque disruption and exposure of the thrombotic core, by hypoperfusion due to inadequate cardiac output, or from a critical reduction in flow across the involved arterial segment. Acute arterial occlusion also may occur because of an aneurysm, commonly popliteal, in which distal embolization of thrombotic material produces occlusion of the outflow tract, ultimately resulting in thrombosis of the aneurysm itself because of inadequate outflow. Hypercoagulable states, such as the antiphospholipid syndrome, resistance to activated protein C, and deficiencies in protein C, protein S, or antithrombin III also may cause acute arterial occlusion. These disorders should be suspected when the patient lacks the usual risk factors for atherosclerosis (Table 20-3). Rarely, acute arterial thrombosis develops as a result of repeated mild trauma, as occurs with a cervical rib compressing the subclavian artery or from occupational trauma such as the operation of a pneumatic tool, the vibrations of which cause injury to the digital arterial wall. Acute arterial thrombosis also may occur after diagnostic and therapeutic intraarterial procedures, such as cardiac catheterization and peripheral vascular arteriography. An intimal flap can be created by the catheter even with the most careful technique, resulting in thrombosis of the vessel.

Clinical Manifestations

The clinical picture of acute arterial thrombosis may range from no symptoms whatsoever to severe, limb-threatening ischemia, depending on whether or not an adequate collateral network is present and the size and location of the involved vessel. A patient with a chronic superficial femoral artery stenosis may progress to complete occlusion of the vessel and remain asymptomatic because blood flow is maintained by means of a well- established collateral network. By contrast, a young patient with a coagulation disorder may develop occlusion of the same vessel and have far more severe symptoms because of the absence of collateral circulation. In addition, occlusion of the superficial femoral artery is less serious than occlusion of the popliteal artery, because the profunda femoris artery can provide blood flow to the extremity when the superficial femoral artery is thrombosed, whereas the popliteal artery is the only major vessel to the foot at the level of the knee.

Diagnosis

The clinical findings in acute arterial thrombosis vary with the severity of ischemia but are essentially the same as those observed with acute arterial embolism and are described by the six *p*'s. In contrast to patients with straightforward peripheral embolism, all patients with suspected acute arterial thrombosis should undergo preoperative arteriography unless the degree of ischemia precludes the delay required for the procedure.

The nature of the acute arterial occlusion should be established as either thrombotic or embolic whenever possible. The distinction is often difficult to make, but certain findings on history, physical examination, and arteriography are helpful. A history of peripheral vascular disease (e.g., claudication) in the involved or contralateral limb suggests a thrombotic cause, as does a personal or family history of popliteal or aortic aneurysm. Similarly, physical findings indicative of arterial insufficiency, such as skin and nail changes or the absence of distal pulses in the uninvolved extremity, suggest thrombosis rather than embolus. Conversely, the absence of evidence of peripheral vascular disease on history or physical examination implies an embolic cause. Occasionally an embolus is visible arteriographically as a rounded structure at the most proximal site of occlusion. Often, however, this is not seen because of retrograde propagation of the thrombus. Differentiation between these two entities is very important; both require treatment, but the treatment may be very different. Patients with embolic disease have normal vessels, and operative or pharmacologic treatment is limited to elimination of the thromboembolus. By contrast, the vast majority of patients with thrombotic disease have underlying atherosclerotic occlusive disease and will require correction of the lesion responsible for the thrombotic event in addition to removal of the thrombus itself.

Operative Therapy

Operative therapy has been the standard of care for acute peripheral arterial occlusion since the 1950s. The introduction and refinement of thrombolytic agents, however, has led to confusion and controversy regarding the indications for use of these two effective therapeutic modalities. In general, thrombolytic therapy with *recombinant tissue plasminogen activator (rt-PA)* or, more frequently, urokinase is reserved for patients in whom the ischemic event is not so severe that reperfusion cannot be delayed for the time required to perform arteriography and effect clot lysis.

Arterial Embolus

Acute arterial embolism is a disorder that virtually always requires therapy. Most patients with emboli to the peripheral circulation have cardiac disease, and embolectomy can usually be performed through a small incision under local anesthesia. Embolectomy should not be avoided on the basis of its risk to the patient, as the development of extremity gangrene and amputation incurs a far greater physiologic and operative stress. Attempts should not be made to restore blood flow to nonviable limbs, as reperfusion may result in the return of toxic substances such as potassium, lactic acid, and myoglobin to the circulation. The risk of death under these circumstances is high, and amputation should be accomplished expeditiously.

The urgency with which blood flow must be restored to the ischemic limb can be estimated on the basis of clinical findings, and when possible, patients should be medically stabilized before even the simplest procedure. Systemic heparinization is accomplished as soon as the diagnosis is established and is continued throughout the operation. The incision is placed over the presumed site of the embolus. A groin incision is used if the thromboembolic process involves the femoral vessels, and a medial incision just below the knee is used for a popliteal embolus. When there is an embolic occlusion of the aortoiliac segment, the thromboembolus is removed through bilateral groin incisions. A transverse arteriotomy is performed after the patient has been heparinized and proximal and distal control of the artery has been obtained. The embolus is extracted through the arteriotomy by passing an embolectomy catheter proximally and distally to remove all propagated thrombus (Fig. 20-20). Additional incisions may be required to remove thrombus from distal vessels. For example, a below- knee incision would be used to expose the popliteal trifurcation in the case of retained anterior tibial thrombus, since this vessel is often difficult to catheterize from the femoral approach because of the acute angle at which it branches from the popliteal artery. A completion arteriogram should be obtained if there is concern about residual thrombus. Intravascular angiography has recently been used to ascertain the effectiveness of thrombectomy. Although usually reserved for trauma, fasciotomies may be performed if compartment swelling is anticipated.

Anticoagulation should be instituted during the early postoperative period. The majority of patients will require long-term oral anticoagulation with warfarin because of cardiac mural thrombi or valvular vegetation. The prognosis for limb salvage is determined by the preoperative condition of the extremity and the success of clot removal. Muscle necrosis and clot propagation are reduced when embolectomy can be accomplished within 6 h of the onset of ischemia. Beyond this interval, the preoperative finding of a soft, pliable calf muscle is predictive of a good outcome. The overall in- hospital mortality in patients with arterial emboli is 25 percent or more, with the majority of deaths occurring as a result of underlying cardiac disease.

Arterial Thrombosis

The operative choices are more complex when acute arterial occlusion develops as a result of thrombosis rather than embolization. Fortunately, time constraints are usually less stringent, as the degree of ischemia is rarely as severe. Operative therapy is generally a combination of the strategies of thrombus removal discussed in the section on arterial emboli and of bypass grafting discussed in the section on chronic arterial occlusion.

Thrombolytic Therapy

The use of thrombolytic agents in the treatment of acute arterial occlusion, introduced by Sherry and associates in the 1950s, has become increasingly popular. Many advances and technical refinements have been made over the years, including the use of catheter-directed administration of thrombolytic agents, coaxial catheter systems, and the use of safer, more effective agents. Thrombolytic agents may be used instead of or in addition to standard operative techniques, depending on the nature of the occlusion. Thrombolytic agents are unlikely to completely supplant operation in the treatment of acute arterial occlusion, since many patients have fixed arterial lesions that must be addressed in order to prevent rethrombosis. In addition, some thromboemboli are not susceptible to thrombolysis. In cases of severe ischemia, the time required to perform a diagnostic arteriogram and achieve pharmacologic clot removal may jeopardize limb salvage.

All thrombolytic agents work by activating the plasmin system, thus effecting lysis of fibrin and dissolution of the clot infrastructure. Streptokinase is a bacteria-derived compound that binds plasminogen, and it is this complex that converts plasminogen to form plasmin. Urokinase is derived from renal parenchymal cells, and *rt-PA* is derived from vascular endothelial cells; both of these agents activate plasminogen directly. Unlike streptokinase and urokinase, rt-PA has a greater affinity for plasminogen at the site of the thrombus than for circulating plasminogen but has not been proved to decrease the incidence of distant bleeding complications associated with local or systemic administration.

Thrombolytic therapy is indicated in any patient with an acute native arterial or graft occlusion and is most effective when administered within 2 weeks of thrombosis. There are a number of important contraindications to the use of lytic therapy, including a history of gastrointestinal or intracerebral lesions, pregnancy, and any contraindication to arteriography. As with catheter thromboembolectomy, blood flow should not be restored to nonviable limbs, since the consequences of reperfusion may be lethal.

The patient is given an antithrombotic agent such as aspirin or heparin once the decision to administer thrombolytic therapy has been made. This decreases thrombus formation around the catheter itself, reducing the incidence of embolization as the catheter is removed. In addition, platelet activation occurs during thrombolytic therapy and may result in the inhibition of clot lysis or early rethrombosis. Heparin or aspirin may prevent these platelet-associated complications.

Thrombolytic therapy is instituted with catheter placement proximal to the level of the occlusion, after which a diagnostic arteriogram is performed. The catheter is then advanced into the clot itself and is used to administer the thrombolytic agent. If the catheter cannot be advanced into the thrombus, attempts at lysis are rarely successful. A high dose of thrombolytic agent is usually administered for the first 4 h (e.g., 4000 IU/min urokinase), followed by a smaller dose (e.g., 2000 IU/min urokinase), which is continued up to 48 h. Arteriograms are repeated at regular intervals to assess the effectiveness of clot lysis. Therapy is discontinued before 48 h if satisfactory clot lysis has been obtained, if a complication occurs, if no significant lysis is achieved, or if the rate of lysis is inadequate for limb salvage. There are several treatment options after the thrombolytic agent is discontinued. No further treatment is needed if complete lysis is achieved and there is no arterial lesion, although most of these patients will require evaluation and treatment of the underlying cause of the occlusion. Patients with appropriate arterial or graft lesions may be treated by percutaneous balloon angioplasty. Other patients will require replacement or revision of a bypass graft to correct the underlying lesion.

Thrombolytic therapy has been postulated to be associated with a number of advantages over operative therapy in the treatment of acute arterial occlusion. In the case of thromboembolic disease involving an otherwise normal artery, for example, in a patient with a cardiac embolus or in a hypercoagulable state, operative intervention may be avoided entirely with successful thrombolysis. Benefit may also be conferred to the patient with an underlying arterial lesion, as the stenosis can be identified arteriographically after clot lysis. As a result, a vein graft may be revised using a patch angioplasty, rather than by replacing it with a new graft, and isolated lesions involving both arteries and grafts can often be treated by percutaneous methods. Finally, thrombolytic agents can lyse thrombi that are inaccessible by operative methods, such as those in the distal tibial vessels, potentially

improving the chances of limb salvage.

The risks and complications associated with the use of thrombolytic agents must be weighed against the potential benefits. Administration of any thrombolytic agent may result in a systemic lytic state, which predisposes patients with occult lesions to hemorrhage. The most common site is the gastrointestinal tract, but intracerebral hemorrhage, occurring in 0.5 to 1.0 percent of patients, is the most devastating. Bleeding may occur at the catheter site in approximately 15 percent of patients receiving thrombolytic agents, usually in the form of a hematoma that develops after catheter removal. Bleeding complications may be more frequent when the fibrinogen falls below 100 mg/dL, but this has not been demonstrated definitively. The treatment of hemorrhagic complications during thrombolytic therapy involves immediate discontinuation of the thrombolytic agents and appropriate medical or surgical management of the bleeding site. In cases of severe bleeding, particularly intracerebral, aminocaproic acid (Amicar) should be administered without delay. Streptokinase, derived from beta- hemolytic streptococcus, is antigenic and is associated with an allergic reaction in patients recently exposed to the bacteria or who have recently received streptokinase. The agent should not be given to these individuals. As clot lysis progresses, fragments from the dissolving clot frequently embolize to distal vessels, such as the tibial and digital arteries. When this occurs, the extremity becomes ischemic again, mimicking reocclusion of the vessel. Distal embolization of thrombus responds to continuation of thrombolytic therapy in almost all cases, preferably with advancement of the catheter into the involved segment.

The results of thrombolytic therapy for acute arterial occlusion have been encouraging, with most series reporting success rates of 60 to 80 percent, as determined by clot lysis and clinical improvement. Several prospective, randomized studies have compared initial thrombolysis with immediate operation for the treatment of peripheral arterial occlusion, including the Surgery versus Thrombolysis for Ischemia of the Lower Extremity (STILE) trial and the Thrombolysis or Peripheral Arterial Surgery (TOPAS) study. From the conclusions of these and other trials, several pertinent caveats may be drawn. First, the appropriate use of thrombolytic therapy often includes subsequent endovascular or open surgical intervention to correct the lesion responsible for the occlusive event. For instance, successful thrombolysis of a saphenous vein femoropopliteal bypass graft almost always uncovers a vein graft stenosis, which must be repaired to prevent rapid rethrombosis of the conduit. Second, patients with acute occlusion (e.g., less than 14 days) represent better candidates for thrombolysis than patients with chronic occlusion. Third, the results of thrombolysis appear to be better for graft occlusions than for occlusions of native arteries, and successful clot dissolution is almost never possible if the catheter cannot be guided into the substance of the thrombus. The best clinical results are achieved when these factors are considered, weighing the morbidity from bleeding complications and delays in reperfusion against the potential benefit of avoiding a larger, frequently urgent operation to restore adequate arterial perfusion.

CHRONIC ARTERIAL OCCLUSION

Atherosclerosis is the most common cause of occlusive disease of the arteries supplying the extremities. Although peripheral atherosclerosis is a diffuse process affecting the arterial tree, the most severe lesions tend to occur at discrete sites. Symptomatic peripheral arterial lesions are most frequently located in the infrarenal abdominal aorta, the iliac arteries, and the superficial femoral artery at the level of the adductor canal. The presence of arterial stenoses in other locations is usually indicative of coexistent disease processes such as diabetes mellitus (profunda femoris and tibial arteries) or inflammatory arterial processes (axillary arteries).

The incidence of symptomatic peripheral atherosclerotic occlusive disease is lower than that of coronary disease, but individuals with significant coronary disease frequently display manifestations of peripheral arterial disease as well. This finding is not surprising, given the common pathogenesis of the two diseases. Symptomatic peripheral arterial disease occurs predominantly in males over the age of 50 years. Additional risk factors that were documented in the Framingham study include systolic hypertension, cigarette smoking, hyperlipidemia, and diabetes mellitus (Table 20-4). The pathology observed in affected arteries is usually that of advanced atherosclerosis, with large cholesterol- and calcium-containing plaques encroaching on the lumen. As in all forms of symptomatic atherosclerotic occlusive disease, the pathophysiology is that of gradual plaque enlargement, with eventual thrombosis of the residual lumen. Arterial thrombi are initiated by the deposition of platelets on the plaque surface, attachment and aggregation of additional platelets, activation of the coagulation pathways, and deposition of fibrin, terminating in a platelet-fibrin thrombus and occlusion of the vessel. Thrombus formation on the exposed core of an ulcerated plaque can result in acute occlusion of highly stenotic lesions. This luminal thrombus provides the opportunity to recanalize chronic arterial occlusions with fibrinolytic agents, restoring a small channel in an apparently chronically occluded vessel.

Aortoiliac Occlusive Disease

The pathophysiology of the symptoms produced by atherosclerotic occlusive disease of the infrarenal aorta and iliac vessels was first described by Leriche, and the characteristic syndrome of claudication, impotence, and diminished femoral pulses bears his name. Three patterns of aortoiliac disease have been identified (Fig. 20-21). Localized disease confined to the aorta and the common iliac arteries is found in 10 percent of patients (type I). The disease extends into the external iliac arteries in approximately 25 percent of patients (type II). The occlusive process is multisegmental, involving the infrainguinal vessels, in the remaining 65 percent of patients (type III).

Clinical Manifestations

Aortoiliac occlusive disease typically is characterized by intermittent claudication involving the thigh, buttock, and calf, and sexual impotence in males. Symptoms may remain stable for years and may even improve as collateral vessels enlarge with exercise or cessation of smoking. Limb- threatening ischemia is rare with isolated aortoiliac occlusive disease; symptoms are limited to claudication in the absence of concurrent infrainguinal disease. Aortoiliac occlusive disease also may be manifested by atheromatous embolization from ulcerated plaques in the involved arterial segment. The emboli may lodge in the digital arteries of the toes, producing gangrene of one or more digits; as previously mentioned, this phenomenon has been termed the blue-toe syndrome. Microemboli also may travel to dermal vessels, resulting in a blue reticular cutaneous pattern. Biopsy examination of these lesions reveals cholesterol debris within the terminal arterioles.

Impotence represents a symptom complex of diverse causes. Satisfactory male sexual function requires adequate blood flow to the penis via the hypogastric and internal pudendal arteries, an intact pelvic parasympathetic system, and an appropriate hormonal and psychologic milieu. Impotence can be ascribed to a vascular cause in only a small minority of patients.

Diagnosis

An accurate history and physical examination can establish or exclude the diagnosis of aortoiliac occlusive disease with certainty in the vast majority of patients. The classic findings of thigh claudication and decreased groin pulses may be accompanied by lower abdominal and femoral bruits. The diagnosis is occasionally confused with that of lumbosacral nerve root compression caused by disc herniation or spinal stenosis. Patients with these disorders can be distinguished from those with claudication by the fact that their pain is temporally related to changes in position rather than ambulation.

The clinical suspicion of aortoiliac occlusive disease may be difficult to confirm using the noninvasive vascular laboratory. Arterial cuff pressures obtained at the proximal thigh level do not necessarily reflect pressures within the iliac or common femoral systems, since the cuff cannot be positioned high enough to occlude these vessels. A decreased proximal thigh cuff pressure may therefore represent occlusive disease of both the proximal superficial femoral and profunda femoris arteries rather than an inflow lesion of the aortoiliac segment. Common femoral Doppler waveforms have been of some use in the diagnosis of inflow disease. The normal waveform is triphasic, with a negative diastolic component. Loss of the triphasic waveform and the presence of a blunted, biphasic configuration occurs relatively early in aortoiliac occlusion and is evidence of inflow disease (Fig. 20-22). Percutaneous needle cannulation of the common femoral artery is an invasive but extremely precise means of identifying inflow occlusive disease. The presence of a pressure gradient between the brachial and the femoral vessels is indicative of significant aortic or iliac stenotic lesions. When a lesion is suspected but no gradient is present, papaverine administration may be helpful in identifying the stenosis. Papaverine is injected intraarterially at the time of cannulation to dilate the distal vasculature, thus reducing peripheral resistance and increasing blood flow across the aortoiliac segment. If a stenosis is present, this will result in a decreased pressure in the femoral artery. A decrease in the femoral-to-brachial pressure ratio of more than 15 percent after papaverine administration-induced hyperemia implies significant occlusive disease involving the aortoiliac segment. It also has been shown to be predictive of symptomatic relief with aortoiliac reconstruction alone, even in the presence of concurrent, uncorrected infrainguinal disease.

Vasculogenic impotence can be differentiated from impotence caused by other disorders using noninvasive laboratory techniques. The penile systolic pressure is measured using a small pneumatic cuff placed around the shaft of the penis and a photoplethysmographic probe on the glans. The penile/brachial index is defined as the penile systolic pressure divided by the brachial systolic pressure, and is approximately 1.0 in potent men. A penile/brachial index above 0.8 confirms the adequacy of penile arterial inflow and virtually excludes the diagnosis of vasculogenic impotence. The converse is not necessarily true; some men are potent despite indices of less than 0.6.

Preoperative standard or digital arteriography is required in patients undergoing elective operation for aortoiliac occlusive disease. The arterial tree should be visualized from the renal arteries to the feet, and the presence of concurrent disease in the inferior mesenteric, hypogastric, and profunda femoris vessels assessed. A transfemoral approach to arteriography is preferable for reasons of ease and safety, but the absence of palpable femoral pulses on either side may necessitate use of a brachial or translumbar route. Intravenous digital studies have been widely used in patients with aortoiliac occlusive disease and absent femoral pulses, but the resolution of this technique is poor because of factors such as bowel gas motion in the abdomen and sluggish blood flow in the extremities.

Treatment

The primary indications for aortoiliac reconstruction are threatened limb loss with rest pain, ulceration, or gangrene and atheroembolic phenomena such as the blue-toe syndrome. There is some disagreement on the advisability of operation for symptoms of claudication alone. In general, claudication of sufficient severity to restrict livelihood or lifestyle is considered an acceptable indication for aortoiliac reconstruction procedures in medically fit individuals.

Aortoiliac reconstruction can be accomplished by four methods: aortofemoral bypass, aortoiliac endarterectomy, extraanatomic bypass, and balloon catheter dilatation. The durability of aortofemoral bypass and aortoiliac endarterectomy is greater than that of extraanatomic bypass or balloon angioplasty. Extraanatomic procedures, typically in the form of axillofemoral or femorofemoral bypass, are usually reserved for medically unstable patients who are at high risk for perioperative complications with an intraabdominal procedure. Femorofemoral bypass has a higher patency rate than axillofemoral bypass, but it requires one normal iliac artery for inflow, and so it is not suitable for patients with bilateral disease. Femorofemoral bypass avoids any possibility of pelvic autonomic nerve injury and is therefore frequently employed as a primary reconstructive procedure in young male patients in whom maintenance of normal sexual function is an important issue. Balloon angioplasty is the least invasive method of restoring inflow to the lower extremity and has the same success rates as operative bypass in the treatment of isolated common iliac artery lesions.

Operations on the abdominal aorta are associated with a perioperative mortality of 2 to 10 percent, depending on the presence and severity of associated medical conditions. Complications are most often related to concomitant atherosclerosis involving the coronary and cerebral vasculature. The presence of chronic obstructive lung disease is associated with an increased risk of perioperative pulmonary complications, the most serious of which is the need for prolonged ventilatory support. The preoperative testing for aortoiliac procedures may include pulmonary function tests, dipyridamole-thallium myocardial imaging, dobutamine echocardiography, and coronary angiography, and the evaluation frequently uncovers unsuspected cardiac and pulmonary disease.

The perioperative technical complications of aortoiliac reconstruction are hemorrhage, thrombosis of the reconstruction, and distal embolization of atherosclerotic debris from aortic manipulation. Ischemia of the rectosigmoid colon may be a result of interruption of the inferior mesenteric and hypogastric arteries or from atheroembolization to the bowel.

Paraplegia is a rare complication that may be secondary to ligation or embolization of a prominent lumbar vessel providing blood flow to the spinal cord through the greater medullary artery of Adamkiewicz (arteria radicularis magna). Damage is usually limited to the anterior two-thirds of the spinal cord in the distribution of the anterior spinal artery, producing a loss of motor function with preservation of sensation. Sexual dysfunction is common after reconstruction of the abdominal aorta; retrograde ejaculation occurs in over 40 percent of patients, and impotence in 25 percent of patients. The frequency of postoperative sexual dysfunction may be decreased by minimizing periaortic dissection, thus preserving the sympathetic and parasympathetic nerves, and by maintaining blood flow through at least one hypogastric artery.

Aortofemoral Bypass

Aortofemoral bypass is the treatment of choice for symptomatic aortoiliac occlusive disease. A bifurcated prosthetic graft made of polyester or, less commonly, *expanded polytetrafluoroethylene (ePTFE)* is used to bypass the stenotic lesions. Although there are advantages and disadvantages to each of the prosthetic materials, this is of less concern in aortofemoral grafting, since patency is excellent with any conduit. The 5-year and 10- year patency rates of aortoiliac reconstructions are 90 and 75 percent, respectively.

Polyester grafts are constructed in either a knitted or woven configuration. The advantages of the knitted structure are the excellent tissue ingrowth that occurs through the wide interstices of the graft and the technical ease of handling at the time of operation. Prec clotting with the patient's blood is required, however, to avoid massive hemorrhage through the interstices when blood flow through the graft is initially established. Knitted grafts have been associated with degeneration and aneurysmal dilatation over time. With woven polyester grafts, prec clotting is infrequently required and graft dilatation is less common. Polyester grafts have been coated with albumin or collagen, eliminating the need for prec clotting but substantially increasing the cost of the grafts. Using *ePTFE* grafts for aortofemoral bypass also eliminates the need for prec clotting, since these grafts are impermeable to blood and aneurysmal degeneration has not yet been reported. Despite these potential advantages, the use of ePTFE grafts in the aortofemoral position has not gained widespread acceptance. The choice of graft material for aortoiliac reconstruction is determined by surgeon preference, since there is no difference in patency between the fabrics.

Once the graft has been selected and the patient has been systemically heparinized, the proximal anastomosis between the graft and the infrarenal aorta is created using either an end-to-end or a end-to-side technique (Fig. 20-23). The type of anastomosis chosen is determined by the need to maintain antegrade blood flow to the hypogastric vessels. An end-to-side proximal anastomosis is indicated when the external iliac arteries are stenotic or occluded bilaterally, since the arterial flow to the spinal cord and colon in this situation may be dependent on antegrade flow through the distal aorta to the hypogastric and inferior mesenteric arteries. The graft limbs are then delivered to the femoral vessels through tunnels that are created beside the external iliac arteries. The distal anastomoses should be placed at the common femoral level, through bilateral groin incisions. Unlike in patients with aneurysmal aortic disease, bypass to the iliac arteries is unwise in the presence of aortoiliac occlusive disease. Although aortoiliac bypass eliminates the need for groin incisions, the patency rate of this procedure is substantially lower than that of aortofemoral bypass. In addition, bypass to the femoral vessels provides the opportunity to correct proximal profunda femoris artery stenoses by placing the hood of the graft onto the profunda femoris artery beyond the area of narrowing.

Endarterectomy Procedures

Aortoiliac endarterectomy was introduced before the advent of prosthetic graft conduits. The aorta and iliac arteries are exposed and all branch vessels are controlled in preparation for clamping. The patient is heparinized, and occlusive vascular clamps are placed. The atherosclerotic plaque is removed through one or more arteriotomies, and the arteriotomies are subsequently closed, with or without a patch. Endarterectomy does eliminate the use of prosthetic material, but the procedure is not suitable for patients in whom the atherosclerotic process extends to the external iliac artery. Endarterectomy requires localized atherosclerotic disease with an appropriate ending of the plaque on relatively normal vessel. Early failure of the endarterectomized segment can be predicted in the absence of a suitable end point within 1 or 2 cm of the origin of the external iliac artery. Nevertheless, aortoiliac endarterectomy may be the procedure of choice for young patients with a long life expectancy and disease localized to the aorta and common iliac vessels. It achieves excellent patency rates and avoids some of the complications associated with prosthetic graft material.

Extraanatomic Bypass

Extraanatomic revascularization procedures include axillofemoral, femorofemoral, and obturator bypasses (Fig. 20-24). Axillofemoral and femorofemoral procedures are preferred when the patient's medical condition renders the risk of a major intraabdominal operation unacceptable. Other indications include lower-extremity ischemia in the presence of an infected aortic graft, reoperation for aortofemoral graft occlusion, and revascularization in a sexually active male to avoid the possibility of postoperative sexual dysfunction. Prosthetic graft materials are almost always employed. The axillary artery provides inflow for an axillofemoral graft; the graft is tunneled subcutaneously along the lateral trunk to the ipsilateral femoral artery in the case of an axillounifemoral bypass, or to both femoral arteries in an axillobifemoral bypass. The patency of axillounifemoral bypasses is poor, averaging well below 50 percent at 5 years. The flow through axillobifemoral bypass is twice that of the axillounifemoral graft, resulting in a significantly higher 5-year patency rate. Thus, axillobifemoral bypass is often employed for unilateral leg ischemia in order to improve graft longevity.

The best results of extraanatomic bypass procedures are achieved with femorofemoral bypass grafts; the 5-year patency rate of these grafts is between 50 and 75 percent. The procedure involves exposure of both femoral arteries in the groins and subcutaneous placement of a prosthetic conduit in the suprapubic region. A nonstenotic donor iliac vessel is required; the adequacy of this vessel can be evaluated with femoral artery pressure measurements made before and after papaverine injection. Concomitant iliac balloon angioplasty has been successfully employed in the presence of iliac stenotic disease to establish adequate inflow to the donor side. The obturator bypass is a rarely used but extremely effective procedure that is employed primarily to circumvent an infected graft in the groin area. The inflow site is the aorta, the iliac artery, or an aortofemoral graft limb within the abdomen. The graft is tunneled through the obturator foramen and joins the superficial femoral artery distal to the groin, thus completely avoiding the infected area.

Endovascular Procedures

Percutaneous transluminal angioplasty (PTA) was introduced in 1964 by Dotter as a method of reestablishing flow through stenotic vessels. A decade later, Gruntzig further refined the technique of balloon catheter dilatation of arterial stenoses. Today the procedure is initiated by the percutaneous cannulation of an accessible vessel, either the common femoral artery or the brachial artery. After a diagnostic arteriogram has been performed and pressure gradients have been measured, a balloon catheter is advanced across the arterial stenosis. The balloon is inflated to several atmospheres of pressure, and completion arteriograms and pressure measurements are obtained. Initially, the mechanism underlying balloon dilation was believed to be circumferential dilatation of all layers of the arterial wall with compression and remodeling of the plaque. Later studies suggested that this is not the case; rather, the plaque is cracked, forming fissures and false channels. The media is overstretched, causing destruction of its elastic components. The adventitia remains intact unless the artery has been overdilated. Rupture of the wall with hemorrhage or false aneurysm formation occurs if the artery is excessively overdilated, but this complication is rare.

The results of PTA have been satisfactory in relatively short nonoccluding lesions of the aortoiliac segment, with patency rates (60 to 90 percent at 5 years) approaching those of operative correction. However, balloon angioplasty procedures have generally been reserved for patients with less extensive atherosclerosis. Moreover, the long-term results of PTA have been determined with subjective methods rather than with the laboratory methods routinely used in evaluating the results of vascular surgical procedures. These two factors may have resulted in an overestimation of the success of PTA, rendering direct comparisons of patency rates between angioplasty and operative revascularization impossible.

A plethora of new devices designed to relieve arterial stenoses by relatively noninvasive means have been introduced over the past decade. In procedures collectively termed *endovascular interventions*, these devices are placed percutaneously or through femoral artery cutdowns. The devices have included: (1) laser-heated metallic probes designed to remodel plaque as the instrument is advanced through an atherosclerotic lesion; (2) catheters that vaporize atheromata by direct delivery of laser light; (3) "atherectomy" devices that remove plaque with rotational cutting blades; and (4) intraluminal metallic stents designed to maintain the artery in an expanded state and thus decrease the frequency of postprocedural restenosis.

Vascular endoscopes are frequently used to evaluate the results of these interventions, and these instruments have been helpful also in the assessment of operative procedures such as balloon catheter thrombectomy and valve disruption in in situ saphenous vein bypass. Intravascular ultrasound devices have been introduced to provide high- resolution images of atherosclerotic vessels both before and after endovascular manipulation. Unfortunately, the introduction phase of each new device has been associated with unwarranted optimism, prompting premature promotion of the device before long-term safety and efficacy have been demonstrated. To date, none of new devices has matched the results of standard operative revascularization procedures. It is hoped that the continued development, refinement, and objective long-term evaluation of newer endovascular techniques may someday provide less invasive means of restoring arterial flow to the compromised extremity.

Infrainguinal Occlusive Disease

The infrainguinal vessels are defined as those vessels distal to the inguinal ligament, including the common femoral, superficial and deep femoral, popliteal, and infrapopliteal arteries. The superficial femoral artery is the most common site of atherosclerotic obstruction in the lower extremity. The process usually is first evident at the level of the adductor canal, but recent studies suggest that this is not because the atherosclerotic process begins here. Rather, as Zarins and associates have shown, the deposition of atherosclerotic plaque occurs at an equal rate throughout the superficial femoral segment. The normal response to plaque deposition is dilatation of the vessel such that the lumen diameter is maintained. Zarins and associates have suggested that the most severe obliterative process occurs at the level of the adductor canal because this area of the vessel is least likely to display compensatory dilatation in response to plaque deposition.

The risk factors for infrainguinal atherosclerotic arterial disease are the same as those of atherosclerotic disease in general, namely, advanced age, systolic hypertension, cigarette smoking, hyperlipidemia, and diabetes mellitus.

Clinical Manifestations

Stenosis or occlusion limited to only one segment of the arterial tree seldom results in a threatened limb. The most frequent clinical scenario is that of solitary superficial femoral artery occlusion at the level of the adductor canal, producing calf claudication after the patient has walked several blocks. Critical ischemia manifested by rest pain, ulceration, or gangrene is seen only with the development of multisegmental disease involving the aortoiliac, deep femoral, or tibial vessels in addition to the superficial femoral artery. Rest pain is characteristically located in the forefoot or the medial arch and is most severe when the foot is elevated in bed at night. Nighttime calf pain usually is the result of benign conditions such as nocturnal cramping and should not be confused with pain from arterial ischemia.

The natural history of infrainguinal arterial occlusive disease depends on the severity of the patient's symptoms. Boyd prospectively followed more than 1400 patients with intermittent claudication and found that only 7 percent required amputation at 5-year follow-up. The results of the Framingham study were even more optimistic, with less than 2 percent of patients with claudication losing a limb over a 10-year interval. It must be kept in mind, however, that many of the patients in these studies eventually required arterial reconstructive procedures. Nevertheless, it is clear that symptoms of intermittent claudication alone are not a harbinger of limb loss, and operative procedures designed to relieve claudication must weigh this relatively small risk against the somewhat higher risks of perioperative morbidity and mortality. In contrast to intermittent claudication, symptoms of severe ischemia with pain at rest, ischemic ulceration, and gangrene are associated with a high risk of limb loss unless successful revascularization can be achieved. Over 50 percent of untreated patients with limb-threatening symptoms will require amputation within several months of presentation, and this risk increases as the severity of the ischemic process evolves from rest pain to tissue loss.

The life expectancy of patients with lower-extremity occlusive disease must be taken into account when considering options of operative revascularization. Given the systemic nature of the atherosclerotic process, it is not surprising that the survival of patients with lower-extremity occlusive disease is significantly decreased as a result of coexistent coronary disease, cerebrovascular disease, and an increased risk of lung cancer. Recent studies reliably suggest a 70 percent 5-year and a 50 percent 10-year survival rate in patients with chronic lower-extremity arterial disease; these rates on average are 20 to 30 percent below those of age-matched controls.

Diagnosis

The diagnosis of infrainguinal arterial occlusive disease is based on the characteristic clinical findings of intermittent calf claudication, rest pain in the foot, or tissue loss in the form of ischemic ulceration or gangrene. Claudication can be differentiated from other causes of leg pain on the basis of its reproducible relationship to ambulation. Patients with claudication develop pain after walking a distance that is constant for any given patient, and exhibit regression of the pain on cessation of activity, even if the standing position is maintained. By contrast, patients with leg pain secondary to disorders such as nerve root compression exhibit pain that is related to position rather than ambulation.

The physical examination is the most important tool in the diagnosis and localization of lower-extremity occlusive disease. Whereas the patient with aortoiliac occlusive disease manifests absent femoral pulses in the groin, patients with infrainguinal disease have normal femoral pulses and absent popliteal pulses (superficial femoral artery occlusion) or absent pedal pulses (tibial artery occlusive disease). The differential diagnosis of lower-extremity ulcerations can frequently be made on the basis of location and appearance. Ulcerations secondary to inadequate arterial blood flow are characteristically found on the lateral aspect of the ankle or on the foot and are pale and devoid of granulation tissue. Ulcerations resulting from venous stasis are located above the medial malleolus, are pink with abundant granulation tissue, and are accompanied by other stigmata of venous disease such as brown discoloration of the pretibial skin.

The noninvasive laboratory examination of the patient with infrainguinal occlusive disease parallels the physical findings. The common femoral Doppler waveforms have the normal triphasic configuration. Proximal-thigh Doppler segmental pressures may be normal if the superficial or profunda femoris artery is free of disease, but will be decreased if the proximal portions of both vessels are affected. The distal superficial femoral artery is the most frequent site for occlusive disease. The most common Doppler segmental findings, therefore, consist of a proximal-thigh cuff pressure equal to or greater than the brachial pressure and a gradient of at least 30 mmHg between the proximal and distal thigh pressures.

Vessel wall calcification, a change characteristically seen in diabetic patients, renders the arterial wall incompressible. Doppler cuff pressures are unobtainable in this case, since flow through the calcific, rigid vessels cannot be obliterated, even with cuff pressures exceeding 300 mmHg. Pulse volume recordings (PVRs) play an important role in the evaluation of the diabetic with incompressible vessels. The PVR does not require compression of the vessels; rather, it relies on volume changes in the leg between systole and diastole. The presence of blunted PVR tracings between two levels in the extremity is indicative of an interval stenosis. Despite the common misperception that diabetic arterial disease is an entity involving the small arteries and arterioles, the larger vessels are actually the most common location of arterial lesions even in diabetics, while the smaller, digital arteries are frequently spared. This feature of diabetic arterial disease makes it possible to use toe pressures to detect arterial insufficiency when the large vessels are incompressible. Toe pressures are obtained by placing a small pneumatic cuff around the base of the digit and a photoplethysmographic sensor distal to the cuff. The systolic pressure at the toe level is normally somewhat less than the brachial pressure, with normal toe indices (toe pressure divided by brachial pressure) averaging 0.70 or more. Patients with claudication usually have toe indices in the range of 0.40, and patients with rest pain, ulceration, or gangrene have toe indices of 0.10 to 0.20.

Arteriography is the most precise method of documenting the location and severity of infrainguinal arterial occlusive disease. Transfemoral standard or intraarterial digital studies offer the highest-quality images. The small caliber and slow flow characteristic of diseased lower-extremity vessels diminishes the resolution of intravenous digital studies and renders this form of arteriography inappropriate in most patients with infrainguinal occlusive disease. Duplex ultrasonography has been used as a less invasive alternative to contrast arteriography in patients with lower-extremity occlusive disease, but its lack of precision makes it a less adequate method of preoperative evaluation.

Treatment

The indications for infrainguinal arterial reconstruction, similar to those for aortoiliac reconstruction, include severe claudication and limb-threatening ischemia manifested by rest pain, ulceration, or gangrene. Nonoperative management is indicated in patients with mild claudication, and treatment regimens include cessation of tobacco use and institution of a daily program of regular exercise. Vasodilators, including calcium channel blockers, have been used with limited success.

Pentoxifylline is an oral agent that reduces blood viscosity, theoretically increasing blood flow through stenotic arteries and small collateral vessels. In a randomized, double-blind, multicenter trial, pentoxifylline was found to reduce symptoms of claudication in 45 percent of patients, as compared to 23 percent with placebo. The therapeutic benefit from pentoxifylline has been unpredictable in actual clinical usage, with many patients having little or no relief of claudication. The documented improvement in microcirculatory flow, however, has led to interest in using pentoxifylline to improve tissue perfusion in patients with limb-threatening ischemia. Encouraging results have been reported in patients with arterial ulcerations, although randomized, blind studies have yet to be completed.

Three factors are required for successful infrainguinal revascularization: adequate inflow, adequate outflow, and a suitable conduit. As in the case of aortoiliac disease, endarterectomy and bypass grafting both may be used in the management of arterial insufficiency. For patients requiring infrainguinal reconstruction, however, the concerns of inadequate inflow, inadequate outflow, and type of conduit assume much greater importance. These three issues are rarely of concern in aortoiliac revascularization but are fundamental in more distal reconstructions.

Endarterectomy

Infrainguinal endarterectomy was introduced more than four decades ago, before the development of adequate bypass conduit materials. The segment of artery containing the atherosclerotic lesion is exposed, the artery is opened, and the atheroma is removed using a spatula or mechanical stripper. The arteriotomy or arteriotomies are then closed primarily or with a venous or prosthetic patch. The long-term success of the procedure appears to depend on the size of the artery and the length of the lesion. As with any endarterectomy procedure, the atherosclerotic lesion must terminate distally in an area of normal arterial wall. If care is not taken to ensure that this is the case, dissection of blood beneath the shelf of the atheroma may result in the creation of a flap and acute postoperative thrombosis of the reconstruction. Endarterectomy is generally limited to the treatment of short lesions of the superficial femoral artery at the adductor canal and localized stenoses at the origin of the deep femoral artery. Popliteal and tibial artery endarterectomy procedures are rarely successful, because the atherosclerotic disease usually is diffuse in these vessels.

Bypass Procedures

Bypass procedures are described in terms of the inflow and outflow sites. Thus in a *femoropopliteal bypass* the site of inflow is the femoral artery and the site of outflow is the popliteal artery. The two most commonly performed infrainguinal reconstructions are the femoropopliteal and femorotibial bypasses (Fig. [20-25](#)). Popliteal-to-tibial and popliteal-to-pedal bypasses are important but are performed less commonly, as are more distal reconstructions to the arteries of the foot.

The common femoral artery is the standard site of inflow for infrainguinal bypass procedures. The superficial femoral or popliteal arteries also may be used for inflow in the absence of significant proximal stenoses, and these alternative sources of inflow become particularly important when the length of vein available for the bypass is limited. The choice of inflow and outflow sites is relatively straightforward when a single segment of the arterial tree is involved with an occlusive lesion. For example, a femoropopliteal bypass is appropriate when the occlusive process is limited to the superficial femoral artery between its origin and the knee joint. The presence of multisegmental arterial occlusion, however, requires that a decision be made regarding which levels of occlusion should be bypassed. In general this decision is based on the magnitude of the ischemic process. The presence of extensive tissue loss in the foot mandates bypassing all levels of occlusion. For example, a patient with a superficial femoral artery occlusion, an open popliteal artery, and short proximal occlusions of all three tibial arteries will be well served by a femoropopliteal bypass to the "isolated" popliteal segment if the symptoms are of claudication or even rest pain. By contrast, a patient with the same arteriographic configuration but with a large necrotic ulceration over the dorsum of the foot will require bypass to an open distal tibial artery.

The selection of conduit material is of paramount importance in infrainguinal arterial reconstruction, as the long-term patency rate is highly dependent on two primary factors: the site of outflow and the type of bypass material. The graft choices are autogenous vein and prosthetic graft materials (Table [20-5](#)). Among the prosthetic materials, *ePTFE* is the most commonly used, and polyester is infrequently employed. Bovine heterografts have been abandoned because of the propensity for late aneurysm formation, and human umbilical vein biografts have decreased in popularity because they are cumbersome to sew and also are associated with late aneurysm formation. Saphenous vein is the best infrainguinal bypass conduit material. The long-term patency rate of saphenous vein grafts is better than that of prosthetic grafts, and there are fewer postoperative infections than with prosthetic material. The difference in patency is of sufficient magnitude that vein should be used for all infrainguinal bypasses, even if this requires use of the lesser saphenous or cephalic veins. Possible exceptions to this rule may occur when the bypass graft does not cross the knee joint, as in a femoral-to-supragenicular popliteal artery bypass. In this instance, the patency rates of venous and prosthetic bypass grafts do not differ substantially, and use of a prosthetic graft preserves the vein for use in the event of graft failure or for use in a subsequent coronary artery bypass.

There are two general techniques used in autogenous vein bypass procedures, differentiated by the orientation of the vein in relation to the direction of blood flow and the alignment of the venous valves. Advantages and disadvantages have been demonstrated with each method, but most studies have failed to document significant differences in long-term patency rates when precise operative technique has been employed. The *reversed* bypass technique, the first method of infrainguinal vein bypass employed, excises the vein in its entirety and reverses it such that the caudal end is anastomosed proximally and the cranial end distally. The reversed orientation allows blood flow to proceed in the natural direction allowed by the valves, but the small end of the vein is anastomosed to the larger (inflow) artery, and the large end of the vein is anastomosed to the smaller (outflow) artery (Fig. [20-26](#)).

In situ vein bypasses are a second technique, first performed in the 1960s. The vein is left in its usual orientation, disrupting the venous valves to allow blood to flow from the cranial end of the vein to the caudal end. The large and small ends of the in situ venous conduit are connected to the arterial vessels of corresponding size, and venous vasa vasorum are preserved if the vein is left in its natural bed. A variety of instruments have been devised to assist in the disruption of the valves; some are passed through side branches of the vein and others through the caudal cut end of the conduit. An alternative technique, representing a modification of the in situ method, involves excising the vein, disrupting the valves, and connecting the conduit in an "excised, nonreversed" fashion. This technique interrupts the potentially advantageous venous vasa vasorum but decreases the risk of vein injury during the valve-cutting process, since the vein may be straightened and shortened in an "accordion" fashion onto a semirigid valve-cutting instrument.

Endovascular Procedures

Infrainguinal endovascular procedures include percutaneous transluminal balloon angioplasty, laser angioplasty, and catheter atherectomy. None of the devices has matched the success of endovascular procedures in the aortoiliac segment. Balloon angioplasty with or without the use of a stent may have a role in selected patients with superficial femoral disease, but the current results only justify the routine use of this technique for iliac lesions. Laser angioplasty and catheter atherectomy must be considered experimental techniques, and assessment of their safety and efficacy awaits the results of randomized, controlled trials that adhere to uniform reporting standards.

The Diabetic Foot

Diabetes mellitus is the risk factor associated with the highest rate of limb- threatening ischemia of the lower extremity; the incidence of gangrene in diabetics is more than fifty times greater than in nondiabetics. The anatomic distribution of arterial lesions in diabetics differs from that of the nondiabetic population, but the underlying histopathology is similar. Diabetic arterial disease is more common in the distal profunda femoris artery, the distal popliteal and tibial arteries, and the digital arteries of the foot. The aortoiliac segment is usually spared. The widespread belief that diabetic arterial disease is primarily localized to the small vessels is erroneous. Although diabetics do develop a microangiopathy characterized by thickening of the intima and basement membrane, the primary pathologic process is that of typical large-vessel atherosclerosis. The media of muscular arteries is often heavily calcified in diabetics, rendering the vessels incompressible and difficult to occlude with standard vascular clamps.

Diabetics are predisposed to ulceration and gangrene of the foot with relatively rapid progression to limb loss. A variety of complex and interrelated factors are responsible for this. Diabetic neuropathy produces motor and sensory deficits in the foot. The loss of efferent motor fibers results in atrophy of the intrinsic muscles, a characteristic high-arched deformity of the foot, and markedly increased pressure on the metatarsal heads. Sensory loss compounds the problem, because the patient is unaware of pressure-induced skin necrosis and minor injuries. Arterial disease accelerates the process, and the presence of apparently minimal external pressure may lead to the development of extensive tissue damage.

The impression that infection is more common in diabetic patients has never been proved, but when infections do occur, they are often very aggressive and are associated with a high incidence of gangrene and limb loss. Infection may develop from seemingly trivial skin defects and can quickly spread along musculofascial planes to involve the tendon sheaths and muscles of the foot and leg. The bacteriology of foot sepsis in the diabetic patient is multimicrobial, with gram-negative, gram-positive, and anaerobic organisms acting in synergy. The bacteria most commonly cultured in diabetic foot ulceration are *Peptococcus* (80 percent), *Proteus* (55 percent), and *Bacteroides* (45 percent), with an average of five or more species per specimen. The organism populations cultured from superficial sites usually differ from those deep within the wound. The inaccuracy of bacterial identification obtained from superficial wound cultures emphasizes the importance of broad-spectrum antibiotic therapy in the treatment of diabetic foot infections. The time-honored approach is to begin treatment with an aminoglycoside, clindamycin, and ampicillin, and instituting a more specific therapy when results are obtained from deep- wound cultures taken at the time of operation. Newer cephalosporins and penicillin derivatives have been used successfully and are particularly useful for patients with renal insufficiency.

The classic diabetic foot ulcer, the “mal perforant” plantar ulcer, is located over the metatarsal heads on the plantar aspect of the foot. This is a *neurotrophic ulceration*, as it results from the sensory neuropathy of the diabetic. The ulcer begins to form beneath a callous and may eventually erode into the bone, producing a secondary osteomyelitis.

Treatment

The treatment of diabetic foot ulcers differs from the treatment of ulcers in nondiabetic patients in a number of ways. The exclusion of significant underlying arterial disease is of primary importance. The presence of calcific, incompressible vessels may falsely elevate Doppler pressure measurements, and a normal ankle/brachial index does not eliminate the possibility of arterial disease in the diabetic patient. Toe pressure determination is an accurate means of evaluating arterial insufficiency in the diabetic, since digital artery medial calcinosis is rare. Arterial disease is unlikely when the toe/brachial pressure ratio exceeds 0.70. Arteriography is frequently required to exclude the possibility of reconstructible arterial disease of the infrapopliteal, tibial, and pedal vessels. Revascularization procedures are indicated whenever ulceration occurs in the presence of significant arterial disease. In the absence of arterial disease, local wound care, avoidance of repetitive trauma, and the use of specially fitted shoes are appropriate. Amputation may be necessary when the disease is limited to the most distal vessels or when gangrenous changes have progressed despite apparently adequate arterial supply. Minor resections, including digital amputations, transmetatarsal amputations, or other, case-specific resections involving variable portions of the foot, may be appropriate in the diabetic. The presence of a palpable dorsalis pedis or posterior tibial pulse is an excellent predictor of healing, and debridement and amputation should be conservative under these circumstances. Myocutaneous free flaps have been of value in covering seemingly insurmountable exposure of subcutaneous tissue and bone. Latissimus dorsi or rectus abdominis muscle has been utilized, basing the inflow on a patent tibial artery or vein bypass graft.

Sepsis resulting from diabetic foot infection demands urgent treatment. Control of hyperglycemia, drainage of purulent collections, debridement, and rapid institution of broad-spectrum antibiotic therapy are the important initial interventions. Arterial reconstructive procedures are performed after the septic process has been controlled, minimizing the risk of graft infection by avoidance of prosthetic conduits.

Upper-Extremity Occlusive Disease

The causes of chronic upper-extremity ischemia are multiple, and although atherosclerosis is the usual cause, nonatherosclerotic causes are also common. Vasospastic and inflammatory arteritides affect the upper extremity more often than the lower extremity, whereas the reverse is true of atherosclerotic disease. The characteristic symptom of chronic upper- extremity ischemia is arm claudication. Rest pain and tissue loss are unusual because of the extensive collateral network about the shoulder and elbow. When digital gangrene does occur, it is usually a result of microemboli originating from atherosclerotic lesions of the subclavian artery.

Diagnosis

The diagnosis of upper-extremity arterial occlusive disease begins with palpation of the axillary, radial, ulnar, and brachial pulses. The most common site of atherosclerotic disease of the upper extremity is the origin of the subclavian arteries proximal to the vertebral artery, a lesion that may produce a supraclavicular bruit. More distal lesions generally are not atherosclerotic in origin, and disorders such as giant cell arteritis and thoracic outlet syndrome should be considered. Proximal subclavian lesions may produce the arteriographic finding of flow reversal in the vertebral artery, since this vessel may serve as a major source of collateral blood flow to the arm. Serial images performed after injection of contrast material into the contralateral nonstenotic subclavian artery will reveal normal antegrade flow in the uninvolved vertebral artery, with retrograde flow in the vertebral artery on the side of the subclavian stenosis. This arteriographic finding is associated with the clinical picture known as the *subclavian steal syndrome* and is characterized by nonhemispheric cerebrovascular symptoms in association with mild arm claudication. This symptom complex develops as a result of a reduction in posterior cerebral arterial flow secondary to the “steal” of blood as it flows in a retrograde direction through the vertebral artery into the subclavian artery. The isolated arteriographic finding of retrograde vertebral artery flow is rarely associated with cerebral symptoms, since the collateral circulation through the circle of Willis usually is sufficient to compensate for the amount of blood diverted to the arm. The presence of neurologic symptoms in the subclavian steal syndrome should suggest the presence of a coexistent carotid stenosis, limiting the compensatory capacity of the cerebral circulation.

Treatment

The indications for operation in upper-extremity arterial occlusive disease include incapacitating arm claudication, emboli to the hand or posterior cerebral circulation, and symptomatic subclavian steal syndrome in the absence of a coexistent carotid lesion. Significant carotid lesions should be corrected first in the subclavian steal syndrome, reserving direct subclavian reconstruction for patients without carotid lesions or in whom carotid endarterectomy fails to alleviate cerebral symptoms.

Endarterectomy of the subclavian artery has decreased in popularity because of the fragility of the subclavian artery and the intrathoracic nature of the procedure. Extrathoracic approaches to the subclavian artery are more commonly used, employing a transverse supraclavicular incision to expose the common carotid artery for inflow and the distal subclavian artery for outflow. A carotid–subclavian artery bypass is generally performed with a polyester or *ePTFE* conduit; this is one of the few situations in which prosthetic material may have a better patency than vein (Fig. 20-27). Operative alternatives include transposition of the end of the subclavian artery onto the side of the common carotid artery (Fig. 20-28), and axillary–to–axillary artery bypass. Treatment of multiple aortic arch vessel involvement requires a median sternotomy and placement of bypass grafts from the ascending aorta to one or more vessels in the neck. Embolic disease is treated by exclusion of the involved segment of artery from the circulation, usually with a standard carotid–subclavian artery bypass, and ligation of the subclavian artery proximal to the origin of the vertebral artery. Lesions of the axillary and brachial arteries are treated with saphenous vein bypass grafts around the occluding lesion; the patency of these grafts is excellent. Cervical sympathectomy has been used in the treatment of nonreconstructible disease such as digital artery occlusion, but the results have been discouraging.

MESENTERIC AND RENAL ARTERY OCCLUSIVE DISEASE

Occlusive disease involving the arteries to the viscera is associated with the life-threatening complications of intestinal infarction and renal insufficiency. The process may develop suddenly, as in the case of embolization to the vessels, or the process may evolve insidiously, as in the case of progressive encroachment on the arterial lumen by an atherosclerotic plaque. The relative rarity of these disorders, along with the inaccessibility of the organs to physical examination, renders the diagnosis elusive.

Acute Mesenteric Ischemia

Acute mesenteric ischemia is associated with astonishingly high mortality rates, around 75 percent in most series, despite advances over the past few decades in operative technique and perioperative management. There are two basic explanations for these dismal results. First, the rarity of the disease renders prompt diagnosis difficult, and an improved outcome can be achieved only with rapid identification of the problem and restoration of blood flow before the onset of irreversible intestinal gangrene. Second, acute mesenteric ischemia usually develops in elderly, medically compromised patients who do not tolerate the physiologic insult well.

There are four basic causes of acute mesenteric ischemia; superior mesenteric artery embolization, superior mesenteric artery thrombosis, nonocclusive mesenteric ischemia, and acute mesenteric venous thrombosis. Emboli generally originate from the heart and occur in patients with atrial fibrillation or after myocardial infarction. Thrombosis occurs in the presence of underlying mesenteric atherosclerotic disease, as a critical stenosis progresses to occlusion. Nonocclusive mesenteric ischemia develops in patients with low-cardiac-output states, especially in the presence of digoxin or vasopressors. The pathophysiology of nonocclusive ischemia is that of mesenteric artery vasoconstriction, which is frequently manifested as segmental spasm of the secondary and tertiary branches of the superior mesenteric artery. The precise mechanisms causing nonocclusive ischemia are unclear, but the process probably involves a vicious circle in which vasoconstriction is induced in the mesenteric artery bed by a remote stimulus such as cardiogenic shock; local hypoxia follows, and the resultant bowel ischemia leads to bacterial transudation, worsening shock, and sympathetic stimulation, and perpetuation of the mesenteric vasoconstriction. Venous thrombosis may be secondary to infection or dehydration.

Changes discernible in electron microscopy occur in the intestinal mucosa after only 10 min of ischemia, and in light microscopy after 1 h. Hemorrhagic necrosis develops later, with sloughing of the mucosa, edema of the bowel wall, and hemorrhage into the lumen. The bowel wall becomes permeable to the luminal bacteria once the mucosa is shed. Peritonitis ensues from transudation of microflora across the intestinal wall, and septicemia and bacteremia develop as the organisms enter the portal circulation and overload the filtering capacities of the liver. Massive fluid shifts into the bowel wall and peritoneum follow, resulting in hemoconcentration, oliguria, and hypotension. Serum levels of lactate dehydrogenase (LDH), serum glutamic oxaloacetic transaminase (SGOT), serum glutamic pyruvic transaminase (SGPT), and creatine phosphokinase (CPK) become markedly elevated with the death of intestinal cells.

Diagnosis

The classic presentation of acute mesenteric ischemia is the sudden onset of abdominal pain out of proportion to the physical findings. The patients complain of periumbilical pain as a result of the small-intestinal ischemia and spasm. The spasm results in gastrointestinal emptying, with emesis and bloody diarrhea. Laboratory changes occur later, with leukocytosis, elevation of the hematocrit, lactate, LDH, SGOT, SGPT, and CPK levels, acidosis, and hyperkalemia.

Arteriography has played a crucial role in the early diagnosis of acute intestinal ischemia. Differentiation of the three forms of mesenteric artery occlusion can be made with lateral views on a transfemoral aortic injection. Mesenteric emboli generally lodge at the orifice of the middle colic artery, generating the characteristic arteriographic picture of a normal-appearing proximal superior mesenteric artery terminating in a “meniscus sign” several centimeters from its origin on the aorta. By contrast, mesenteric thromboses occur at the level of the most proximal superior mesenteric artery, before the middle colic takeoff. A tapering termination of the superior mesenteric artery is seen within 1 or 2 cm of its origin, and the development of collateral circulation may be evidence of a long-standing stenotic lesion. Nonocclusive mesenteric ischemia produces the characteristic arteriographic finding of segmental mesenteric vasospasm with a relatively normal-appearing main superior mesenteric artery trunk.

The differentiation of superior mesenteric embolus, thrombosis, and nonocclusive ischemia often can be made at the time of laparotomy on the basis of the distribution of the ischemic process. Emboli lodge at the origin of the middle colic artery, distal to the first few jejunal branches of the superior mesenteric artery. Continued perfusion of these jejunal branches spares the proximal jejunum from the gangrenous process. By contrast, thrombotic occlusion occurs at the origin of the superior mesenteric artery, proximal to all the jejunal branches. The gangrenous process runs from the ligament of Treitz to the mid–transverse colon, without sparing of the proximal jejunum. Finally, nonocclusive mesenteric ischemia involves the branches of the superior mesenteric vessel in a segmental fashion and is associated with a patchy appearance of alternating pink and dusky bowel.

Treatment

The treatment of acute mesenteric artery occlusion varies with the cause of the process. It is useful to obtain a preoperative arteriogram so that appropriate management can be planned accordingly. This is not always possible, since the diagnosis of mesenteric ischemia may not have been made before laparotomy; also, some patients present in a moribund state, and the delay required for arteriography may be ill-advised.

Mesenteric Embolus

The basic goal of operation in superior mesenteric artery embolus is the rapid restoration of arterial perfusion with removal of the embolus from the vessel. Usually a long midline abdominal incision is used. The transverse colon is lifted superiorly, and the small bowel is reflected toward the right upper quadrant. The superior mesenteric artery is approached at the root of the small-bowel mesentery, usually as it emerges from beneath the pancreas to cross over the junction of the third and fourth portions of the duodenum. A transverse arteriotomy is made in the vessel, and a balloon catheter is inserted proximally and distally to remove the embolus and propagated thrombus. An assessment of intestinal viability must be made after perfusion has been restored, and segments that are obviously nonviable must be resected. Numerous technical aids have been employed to predict viable from nonviable bowel, including intraoperative intravenous fluorescein injection and inspection with a Wood's lamp and Doppler assessment of antimesenteric intestinal arterial pulsations. A second-look procedure is necessary in many patients and is usually scheduled 24 to 48 h after embolectomy. The rationale for second-look procedures is that the precise extent of intestinal viability may not be evident immediately after reperfusion, and additional resection of ischemic bowel may be necessary. An important tenet of second-look laparotomy is that the decision to proceed with the additional operation must be made at the time of the initial laparotomy. Subsequent analysis of the patient's postoperative course should not alter the decision, since the early postoperative status of these patients bears no correlation with the presence or absence of residual nonviable bowel.

Mesenteric Thrombosis

The therapy of mesenteric thrombosis differs from that of embolization because of the nature of the superior mesenteric artery. The vessel itself is normal in embolic disease, and simple removal of the thromboembolus is all that is necessary. By contrast, the thrombotic process occurs in a severely atherosclerotic proximal superior mesenteric artery. Therefore, patients with thrombotic disease require placement of a bypass graft to the superior mesenteric artery distal to the occlusive process to restore adequate mesenteric flow. The aorta or the iliac artery may be used as the origin of the bypass graft. There are two advantages to using the supraceliac, infradiaphragmatic aorta rather than the infrarenal aorta as the origin for the graft. First, this segment usually is soft and does not present the problems associated with clamping of the frequently calcific infrarenal aorta. Second, the use of the more proximal aorta allows placement of an antegrade graft, which is less prone to kinking when the small bowel is returned to its normal location after construction of the anastomoses. Saphenous vein is usually the graft material of choice in patients with acute mesenteric ischemia; prosthetic materials should be avoided because of the risk of bacterial seeding from transudation or during intestinal resection.

The principles behind the resection of nonviable bowel are the same for patients with thrombotic and embolic mesenteric ischemia. Doppler and fluorescein techniques can be worthwhile, and a second-look procedure may be advisable. Not infrequently, the length of viable intestine is so short as to be unable to sustain life. The treatment of these patients must be individualized on the basis of an understanding of the patient's wishes with respect to permanent intravenous hyperalimentation. Closure of the abdomen without revascularization or resection of bowel may be the most appropriate management for many of these patients.

Thrombolytic therapy is a potential consideration in patients with acute mesenteric ischemia, with intraarterial delivery of the agent into the thrombus at the time of arteriography. Successful lysis of the central core of acute thrombus will return the mesenteric circulation to its chronic, stable state, and subsequent operative revascularization or even balloon angioplasty of the stenotic superior mesenteric artery can be electively undertaken. Percutaneous intraarterial thrombolysis, however, does not provide the opportunity to inspect the bowel for viability after reperfusion. In addition, many hours may be lost during the time required for clot lysis, and attempts at operative revascularization will be significantly delayed if thrombolysis is unsuccessful. Thrombolytic interventions in acute mesenteric ischemia must therefore be considered investigational at present and should be reserved for selected patients.

Nonocclusive Mesenteric Ischemia

The therapy of nonocclusive mesenteric ischemia is primarily nonoperative. A metabolic cause of the problem should be sought and corrected whenever possible. Intraarterial vasodilating agents such as tolazoline or papaverine can be administered directly into the superior mesenteric artery at the time of arteriography, and the effects of the agents can be objectively documented with serial arteriograms. Resection of nonviable bowel may be required when the patient's condition permits.

Chronic Mesenteric Ischemia

Chronic mesenteric ischemia occurs as the result of atherosclerotic occlusive disease of the superior mesenteric and celiac vessels. Less frequent causes include fibromuscular dysplasia, radiation arteritis, autoimmune arteritides, and a secondary mesenteric arteritis after aortic coarctation repair. A disorder known as the “celiac band syndrome” has been described in which the celiac axis is compressed by the median arcuate ligament. The significance of this finding as a cause of abdominal symptoms is doubtful, because a large percentage of normal individuals manifest a moderate celiac stenosis on lateral aortography.

The mesenteric atherosclerotic process usually begins with stenotic plaque formation at the origins of the visceral arteries as they exit the aorta. In contrast to acute mesenteric occlusion, which produces intestinal infarction with superior mesenteric artery occlusion alone, the chronic form of the disease is not associated with symptoms until both the celiac axis and the superior mesenteric arteries are significantly stenosed. The classic symptom complex of postprandial abdominal pain and weight loss is almost uniformly present and is in keeping with the alternate name of “intestinal angina” used to describe the disorder.

The anatomy of the collateral arterial circulation is an important feature in chronic mesenteric ischemia (Fig. 20-29). The celiac and superior mesenteric artery circulations communicate through the gastroduodenal artery and pancreatic branches, respectively. Significant stenotic disease of one system results in dilatation of the gastroduodenal artery to accommodate the increased collateral flow. The inferior mesenteric artery communicates with the superior mesenteric system through two routes. First, a clinically insignificant pathway exists through the left colic artery, the marginal artery of Drummond, and the middle colic artery. Second, a hemodynamically important pathway may evolve through a “meandering mesenteric artery.” This artery is not present in the absence of mesenteric occlusive disease; it develops as a tortuous vessel running through the medial aspect of the left mesocolon to provide anastomotic collateral flow between the proximal inferior mesenteric artery and the superior mesenteric artery. A meandering mesenteric artery can be distinguished arteriographically from a normal marginal artery of Drummond by its more medial location, tortuosity, and shorter route between the inferior and superior mesenteric vessels.

Diagnosis

The most common presentation in patients with chronic mesenteric ischemia is postprandial abdominal pain and weight loss. The pain occurs 20 min to 1 h after a meal, and its intensity is correlated with the amount of food ingested. The relationship between food and pain is so striking that patients decrease the size and frequency of meals, giving rise to the term “food fear.” This conscious restriction of food intake is the cause of the patients' weight loss, rather than a malabsorption syndrome secondary to ischemic bowel injury.

Physical examination in chronic mesenteric ischemia is remarkable only for the obvious cachectic appearance of the patient and the frequent finding of a midabdominal bruit. Occult blood in the stool is unusual and implies a different diagnosis or a more acute ischemic process. Remote stigmata of diffuse atherosclerotic disease are usually present, including carotid and femoral bruits and the absence of palpable lower-extremity pulses. Laboratory tests generally are not helpful. Absorption studies such as fecal fat analysis and urinary excretion of orally administered d-xylose are of limited value. Upper and lower gastrointestinal barium contrast studies usually are normal in patients with chronic mesenteric ischemia, although some patients manifest decreased motility and mucosal edema. Duplex scanning of the celiac and superior mesenteric origins has been reported to be a useful screening method in patients with suspected mesenteric ischemia, but arteriography is necessary to confirm the diagnosis. Lateral aortography reveals stenosis or occlusion of both the celiac axis and the superior mesenteric artery, frequently in association with a large meandering mesenteric vessel on the anteroposterior views. Mesenteric ischemia so rarely occurs in the presence of a solitary visceral arterial occlusion that the accuracy of the diagnosis is called into question when only one visceral vessel is involved.

Treatment

Revascularization is indicated in all patients with symptomatic mesenteric ischemia in order to prevent the development of catastrophic bowel infarction. Nevertheless, rather impressive visceral artery disease may appear arteriographically in the absence of symptoms. The coexistence of abdominal pain and arteriographic findings does not necessarily imply a causative relationship, and other diagnoses such as occult malignancy or inflammatory bowel disease should be excluded in all patients, irrespective of the arteriographic findings.

A transperitoneal approach to operative revascularization provides the opportunity to thoroughly explore the abdomen to exclude nonvascular causes for the patient's symptoms. Historically, endarterectomy was the first method used in visceral arterial revascularization. Endarterectomy is most easily accomplished through a lateral or posterior aortotomy, removing a plug of atheroma from the celiac trunk and the superior mesenteric artery through a “transaortic” exposure. The danger in this technique lies in the inability to adequately visualize the termination of the endarterectomy on the visceral vessel, risking the inadvertent creation of an intimal flap and early thrombosis of the reconstruction.

Bypass procedures have gained widespread acceptance in visceral arterial reconstruction. As in any arterial reconstructive procedure, the surgeon has three decisions to make: the

inflow site, the outflow site, and the type of bypass graft material. The infrarenal aorta, the supraceliac aorta, and the iliac artery are all satisfactory choices as the site of inflow. The supraceliac aorta has the advantage of a much lower incidence of atherosclerotic change than the more distal sites, thereby decreasing the danger of iatrogenic embolization secondary to the clamping of a diseased vessel. Supraceliac inflow also allows the bypass graft to be placed in an antegrade fashion, avoiding the kinking tendency associated with retrograde grafts from an infrarenal location. Both the celiac and superior mesenteric systems should be revascularized if possible, to provide a margin of safety if one reconstruction fails. Anastomosis of one end of the graft to the side of the common hepatic artery will revascularize the celiac system and a similar anastomosis to the superior mesenteric artery will revascularize that arterial bed. Satisfactory results have been achieved with both saphenous vein and prosthetic conduits. The use of a bifurcated polyester graft from the supraceliac aorta is a commonly performed operation for mesenteric ischemia. This procedure is an antegrade reconstruction that preserves the saphenous vein and reduces the number of vascular anastomoses needed to be performed.

Endovascular treatment of mesenteric artery stenoses, using percutaneous balloon angioplasty with or without placement of a stent, is gaining acceptance as a viable treatment option. The procedure must still be considered experimental, pending the availability of long-term clinical follow- up data.

Renovascular Disease

The proximal renal arteries are a common location for the development of atherosclerotic lesions. Two important disorders occur as a result of significant renal artery disease: renovascular hypertension and renal insufficiency. Renovascular hypertension has emerged as the leading cause of surgically correctable hypertension, stimulated by the work of Goldblatt in 1934. In the years that followed, the recognition that renal function improved after renal artery reconstruction fostered interest in the salvage of functioning renal parenchyma, irrespective of the presence or absence of hypertension.

Renovascular Hypertension

Hypertension secondary to renal artery disease is thought to affect 5 to 10 percent of the hypertensive population. Renovascular disease tends to produce a marked elevation in the systolic and diastolic pressures, causing the prevalence of a renovascular etiology to be negligible in the subpopulation of patients with mild to moderate hypertension. Age is also an important correlate of renovascular causes of hypertension. Severe hypertension in young children and elderly adults has the highest probability of a renovascular cause, while hypertension in young and middle-aged adults is usually essential hypertension.

Etiology

There are two basic causes of renovascular hypertension: atherosclerosis and fibromuscular dysplasia. The ratio of atherosclerotic to fibromuscular cases is roughly 21. Atherosclerosis typically occurs at the renal artery ostia, more commonly on the left than on the right. Severe atherosclerosis of the abdominal aorta frequently coexists with renal artery disease.

Fibromuscular dysplasia of the renal artery is an idiopathic disorder encompassing a variety of histopathologic subgroups, all producing stenotic lesions of the intima, media, or adventitia. The most common variety consists of medial fibroplasia with alternating stenoses and small aneurysms, producing the characteristic string-of-beads appearance on arteriography (Fig. 20-30). Fibromuscular dysplasia most commonly occurs in young, multiparous women. In contrast to atherosclerotic disease, the right renal artery is involved more frequently than the left and the lesions frequently occur at the midportion of the renal arteries rather than proximally.

Pathophysiology

Richard Bright was the first to call attention to the relationship between hypertension and renal disease when he observed an association between “hardness of the pulse” and scarred, shrunken kidneys in 1836. Goldblatt defined the cause of the process in his classic canine experiments reported in 1934. Unilateral renal artery constriction produced ipsilateral renal atrophy and systemic hypertension.

The renin-angiotensin-aldosterone system has since been defined as the critical hormonal pathway responsible for the maintenance of the normotensive state. Systemic hypertension may develop as a result of overfilling of the arterial system, arteriolar vasoconstriction, or a combination of the two. Renal artery stenosis produces a low renal perfusion pressure, a compensatory increase in unilateral renin secretion, increased angiotensin-II formation, and elevated blood pressure secondary to vasoconstriction and hyperaldosteronemia-induced volume overload. A normal contralateral kidney can partially compensate for the hyper-renin state by increasing natriuresis. This compensatory response does not occur in the presence of a diseased or absent contralateral kidney or in the presence of contralateral renal artery stenosis.

Two forms of Goldblatt hypertension provide experimental corollaries to renovascular hypertensives with unilateral versus bilateral renal disease. In the “two kidney, one clip” Goldblatt model a single renal artery is clamped and the opposite kidney is left undisturbed. In the “one kidney, one clip” model one renal artery is clamped and the contralateral kidney is removed. Renin levels are elevated indefinitely in the two-kidney model, and the administration of angiotensin-II inhibitors results in decreases in blood pressure both initially and in the established phase. By contrast, renin levels are only transiently elevated in the one-kidney model. Elevated renin secretion is soon suppressed by volume expansion from sodium and water retention, and the established phase of hypertension is maintained by volume expansion and not by renin-mediated vasoconstriction. Predictably, angiotensin-II blockade is ineffective in ameliorating the hypertensive state in the established phase of the one-kidney model.

Diagnosis

Although renovascular hypertension is more common when the hypertension is severe, is of recent onset, is associated with an abdominal bruit, and occurs very early or very late in life, none of these characteristics is sufficiently predictive to substantiate or exclude the diagnosis. It is important to rule out such diagnoses as pheochromocytoma or other adrenal tumors with urinary 17-hydroxy ketosteroid and catecholamine levels.

A widely used but highly inaccurate screening test for renovascular hypertension is the intravenous pyelogram. Findings consistent with renovascular causes include a delay in the appearance of contrast on one side, a difference of 1.5 cm or more in the length of the two kidneys, defects in the renal parenchymal outline consistent with segmental infarction, and ureteral notching from compression secondary to dilated collateral vessels. As many as 30 percent of patients with renovascular hypertension will demonstrate no abnormal findings on pyelography, with a significantly higher false-negative rate in pediatric patients and in patients with bilateral disease.

Radioisotopic renal nuclear scans have been employed in the assessment of renovascular artery disease using iodine-131 hippurate as an indirect measure of renal plasma flow and technetium-99 chelated diethylenetriamine pentaacetic acid (DTPA) as an index of glomerular filtration rate. The diagnosis of renal artery disease is suggested by an asymmetry between appearance and excretion of radiopharmaceutical by the two kidneys. Unfortunately, the nuclear medical techniques are associated with a high incidence of false-negative and false-positive results.

Hypersecretion of renin from a kidney supplied by a stenotic renal artery is the hallmark of surgically curable hypertension. Theoretically, peripheral renin levels should correlate well with renin hypersecretion, since the clearance of renin from the blood remains a constant proportion of the arterial renin concentration. In practice, many patients with curable hypertension exhibit normal peripheral renin determinations. Angiotensin-II blockade provides a means of increasing the sensitivity of peripheral renin determination. The administration of the oral angiotensin-converting enzyme captopril results in an increase in plasma renin activity to a markedly greater extent in patients with renovascular hypertension than in patients with essential hypertension. The single-dose captopril test accurately discriminates between renovascular and essential hypertension, although the test does not establish which kidney is responsible for the problem.

Split renal function tests were among the first diagnostic tools used to predict whether patients would benefit from renovascular reconstructive procedures. These evaluations specifically identify which kidney is responsible in the majority of instances. The ureters are individually catheterized and urine is collected. Split function tests are considered suggestive of a renovascular cause for hypertension when one kidney demonstrates a 40 percent reduction in urine volume, a 50 percent increase in creatinine concentration, or a 100 percent increase in para- aminohippuric acid concentration.

Selective renal vein sampling has proved to be a valuable method of determining the presence of a renovascular cause for hypertension. Renal vein renin ratios greater than 11.5 between the two kidneys are correlated with a renovascular cause for hypertension and predict a satisfactory response from renovascular reconstruction. Recent appreciation that the kidneys do not remove renin from the arterial blood has fostered the use of subtraction methods for improving the accuracy of renal vein renin testing. Contralateral suppression of renin secretion results in negligible differences between renal arterial and venous renin concentrations on the uninvolved side, and the renin concentration in the inferior vena cava has been shown to approximate the renal arterial renin concentration. In practice, renovascular hypertension should be suspected when the difference between caval and renal vein renin levels is near zero on the uninvolved side and when the renal vein renin increment is 50 percent higher than the caval level on the involved side.

Intravenous digital subtraction angiography, a minimally invasive procedure that may be performed on an outpatient basis, has been used as a screening test for the identification of renal artery stenoses. In this procedure dye is injected through an antecubital or femoral vein into the central venous circulation. A computer-subtracted image is obtained as the dye enters the abdominal aorta and the renal arteries, and the test is frequently combined with selective renal vein renin sampling. The disadvantages of intravenous digital angiography include its poor resolution, its dependence on adequate cardiac output, and the large amount of dye necessary to obtain adequate images. Intraarterial contrast studies are the most reliable method of delineating renal artery lesions. Conventional or intraarterial digital techniques provide satisfactory images, and oblique views may be useful to more clearly define proximal stenoses. Intraarterial arteriography is a prerequisite for operative correction of renal artery disease, and many centers perform this procedure regardless of the results of prior screening tests.

Diagnostic Tests Predictive of Outcome

Renal artery disease may produce significant deterioration in renal excretory function in addition to hypertension. The development of an elevated serum creatinine level implies significant bilateral renal disease. Renal artery reconstruction can be expected to result in retrieval of function in the azotemic or nonazotemic patient, with objective improvements in glomerular filtration rate and increases in the size of the kidneys after the procedure. A successful surgical outcome is correlated with a preoperative glomerular filtration rate of less than 20 mL/min. Interestingly, the greatest increase in renal size after revascularization occurs with small kidneys.

Treatment

The indications for reconstructive procedures in patients with renovascular disease include documented renovascular hypertension and chronic renal insufficiency secondary to renal artery lesions. Operative therapy has been shown to be safer than long-term medical management of renovascular hypertension, with significantly greater relief of hypertension and survival rates, whether the cause is fibrodysplastic or atherosclerotic. Moreover, long-term therapy with angiotensin-converting enzyme inhibition may be associated with substantial decreases in renal excretory function. Therefore, pharmacologic management of renovascular hypertension is contraindicated in all but the most medically compromised patients. Thrombolytic therapy has been employed in acute renal artery occlusion, such as occurs with emboli to the kidneys. The thrombolytic agent is infused into the renal artery thromboembolus at the time of arteriography. The results of thrombolytic therapy in acute renal ischemia remain anecdotal, and therapy must be individualized.

There are four choices in the treatment of chronic symptomatic renal artery disease: unilateral nephrectomy, percutaneous transluminal angioplasty, renal artery bypass, and renal endarterectomy. Nephrectomy is indicated in patients with significant renovascular hypertension when the involved kidney is the source of renin production but is so severely damaged from chronic ischemia that the prospects for retrieval of renal function are remote. Renal size and the presence of viable glomeruli on biopsy examination have been used to predict the likelihood of improving renal function with revascularization, with nephrectomy reserved for unilateral renal artery disease in a small kidney with minimal residual excretory function. Balloon angioplasty has been used successfully in renovascular hypertension, but the best results are achieved with lesions distal to the renal artery ostia. Relative contraindications to balloon angioplasty include lesions involving renal artery bifurcations and bilateral renal artery stenoses. The most favorable results with balloon dilation occur in patients with fibromuscular dysplasia; for this patient group the procedure is the initial intervention of choice, although its long-term benefits remain undefined. Placement of an intraluminal stent has been helpful in the treatment of ostial lesions and in other renal lesions refractory to simple balloon angioplasty. Data on the long-term results of renal stenting are unavailable.

Renal artery bypass and endarterectomy are the two main operative procedures performed to correct renal artery stenoses and occlusions. Exposure of the renal arteries is most easily accomplished by reflecting the left colon medially in the case of left renal artery reconstructions and by reflecting the right colon and duodenum medially in the case of right renal reconstructions. Both renal arteries may be exposed with mobilization of the right colon and ileum and reflection of these structures cranially. The choice of the type of renal reconstruction depends on the status of the abdominal aorta. An aortorenal bypass using autogenous saphenous vein is the procedure of choice when the aorta is relatively spared from atherosclerotic change and clamping will not produce injury or distal embolization. Prosthetic grafts with *ePTFE* or polyester are acceptable alternatives to saphenous vein. Saphenous vein should be avoided in children, because it is prone to the development of aneurysmal change. Hypogastric artery is the best choice for aortorenal grafting in the pediatric patient. In the presence of aortic atherosclerotic disease, saphenous vein bypass from the hepatic artery to the right renal artery or splenic artery bypass to the left renal artery the most appropriate alternative. Both procedures avoid the embolic and hemodynamic consequences of aortic clamping. Splenorenal grafts are performed by transecting the splenic artery and constructing an anastomosis of one end of the splenic artery to one end or the side of the left renal artery; collateral flow from the short gastric vessels obviates the need for splenectomy. Distal renal artery lesions may be difficult to expose and revascularize with the kidney in situ; in such cases removal of the kidney, ex vivo bench reconstruction, and autotransplantation has provided an excellent alternative.

Renal endarterectomy is appropriate for atherosclerotic lesions but is not applicable in fibrodysplastic disease. The procedure may be accomplished through a transaortic exposure, endarterectomizing the renal orifices through an aortic incision. This procedure is useful in bilateral renal artery lesions, and a transverse aortotomy across the aorta into both renal arteries avoids the problems associated with a blind ending of the distal extent of the endarterectomy.

EXTRACRANIAL CEREBROVASCULAR DISEASE

The term “carotid” originates from the Greek *karotides*, which in turn is from *karoun*, to stupefy. The term was first used by Galen, who found that compression of these vessels produced a soporific state. Thomas Willis, a seventeenth-century English physician, clearly defined the carotid and vertebral artery supply to the brain in his work *Cebri Anatome*. Despite extensive study of the anatomy of the cerebral vasculature, the relationship between carotid artery disease and ischemic stroke was not appreciated until the early twentieth century. In 1913 Ramsay Hunt described the relationship between cerebral softening and occlusive lesions involving the main arteries to the brain. He urged that evaluation of the carotid arteries in the neck become a routine part of the physical examination. These early observations were followed by advances in cerebral angiography by Moniz in the 1920s, the elucidation of the pathophysiology of transient neurologic deficits by C. Miller Fisher in 1950, and the first operative carotid artery reconstruction by Eastcott, Pickering, and Rob in 1953. The first major series describing carotid artery reconstructions was reported by Lyons and Galbraith in 1957, and carotid endarterectomy soon became established as a safe procedure, although its role in improving the natural history of ischemic stroke continues to elicit much controversy.

Nomenclature

Focal cerebral ischemic disease, or stroke, results from insufficient blood flow to the affected portion of the brain. Stroke may be classified in a number of ways, including the anatomic location of the ischemic cerebral insult, the location of the causative arterial lesion, the pathogenesis, and the time sequence. It is helpful to categorize stroke with regard to two of these parameters in particular: location of the cerebral defect, and time course of the event (Table [20-6](#)).

Location

Neurologic deficits may be divided into those that are focal and those that are diffuse. Focal deficits are those that may be specifically localized to a discrete area of the brain. It is useful to classify these deficits into anterior or *hemispheric* symptoms and posterior or *vertebrobasilar* symptoms, since hemispheric symptoms are frequently caused by emboli from the carotid circulation and vertebrobasilar symptoms originate from either flow-limiting or embolic lesions of the aortic arch vessels, the vertebral arteries, or the basilar artery. Symptoms not referable to the carotid territory are often referred to as *nonhemispheric* symptoms; these include true vertebrobasilar symptoms as well as the more poorly defined global symptoms of dizziness and syncope. The confluence of both vertebral arteries to form the basilar artery provides a margin of safety with respect to hindbrain ischemia. Disease in one vertebral artery does not produce cerebral hypoperfusion unless the contralateral vertebral artery is also diseased or atrophic. Unilateral vertebral artery lesions can, however, be associated with nonhemispheric symptoms as a result of embolization to the ipsilateral or the contralateral hindbrain.

The *subclavian steal syndrome* is a variant of nonhemispheric ischemia in which a subclavian artery stenosis proximal to the origin of the vertebral artery results in retrograde vertebral artery perfusion as a source of collateral blood flow to the arm (see earlier section, [Upper-Extremity Occlusive Disease](#)). The arteriographic finding of retrograde vertebral artery flow is common in the presence of a subclavian lesion, but cerebral symptoms are rare unless concomitant carotid artery disease exists. A small proportion of cerebral infarctions occur as a result of cardiac emboli, which may travel to the anterior or the posterior cerebrum. *Amaurosis fugax* is defined as transient monocular blindness; persistence of the deficit indicates that the ischemic process has progressed to infarction, and the patient has suffered a *retinal stroke*. The cause of monocular visual symptoms usually is an embolus arising from an atherosclerotic plaque at the carotid bifurcation and traveling through the ophthalmic artery to the terminal arterioles of the retina. The embolus is visible on funduscopic examination in about 10 percent of patients with amaurosis fugax, appearing as bright intraarteriolar bodies. Although C. Miller Fisher was the first to report this finding in patients with amaurosis fugax, the clinical sign has been given the eponym *Hollenhorst plaque*, after Hollenhorst's report of 27 patients with this finding.

Time Course

Neurologic deficits can be grouped into three categories on the basis of the duration of the signs and symptoms. An incident of transient symptoms resolving completely within 24 h is termed a *transient ischemic attack (TIA)*, although most TIAs actually resolve within minutes rather than hours. When the frequency of TIAs is greater than two or three per day, the term *crescendo TIAs* is applied. When symptoms last longer than 24 h but resolve within 3 weeks, the condition is known as *reversible ischemic neurologic deficits* (RINDs). When a deficit lasts longer than 3 weeks, it is considered a *completed stroke*. These definitions are based only on clinical findings. Imaging techniques such as CT and MRI have documented cerebral infarction in a significant proportion of patients with transient symptoms, blurring the distinction between TIA and stroke.

There are two types of stroke that represent an unstable clinical situation; the *stroke-in-evolution* and the *waxing-and-waning neurologic deficit*. In patients with stroke-in-evolution the neurologic deficit worsens through a series of discrete exacerbations. Waxing-and-waning deficits fluctuate between mild and severe neurologic compromise, usually over a period of several hours. Stroke-in-evolution and waxing-and-waning deficits occur in the presence of critical carotid stenoses, usually in excess of 90 percent diameter reduction. The pathophysiology may be repeated episodes of embolization or recurrent, borderline cerebral ischemia from low arterial blood flow. In either case, urgent carotid endarterectomy is indicated before irreversible cerebral infarction develops.

Pathophysiology

Carotid and subclavian artery atherosclerotic lesions usually occur in predictable, focal sites. Atherosclerotic disease involving the carotid artery is almost always limited to the carotid bifurcation and ends several centimeters distal to the origin of the internal carotid artery. This makes it possible to perform endarterectomy procedures rather than bypasses for carotid bifurcation disease, removing all the atherosclerotic plaque and ending the endarterectomy distally on an uninvolved segment of the internal carotid artery. Subclavian atherosclerosis tends to occur at the origin of these vessels, with left-sided stenoses prevailing by a ratio of 41.

The particular susceptibility of the carotid bifurcation to atherosclerotic change is most likely due to hemodynamic conditions at this location. Initially it was assumed that the atherosclerotic propensity of the carotid bifurcation was a result of local turbulence and elevated blood flow rate. Investigators speculated that high wall shear stresses produced endothelial damage, predisposing to atherosclerotic degeneration. Zarins and associates refuted this theory and offered convincing evidence that plaque formation is accelerated within areas of low flow velocity and inhibited in areas with high flow velocity and elevated shear stress. They evaluated transverse light microscopic sections of postmortem human carotid bifurcation specimens. Atherosclerotic plaque was most prominent along the outer, posterior aspect of the proximal internal carotid artery. This finding was correlated with hemodynamic observations in transparent models of human carotid bifurcations. The areas of cadaver artery most susceptible to plaque formation corresponded to regions of low velocity and wall shear stress in the models, while the zones that were relatively free of plaque formation corresponded to regions of high flow and high shear stress.

Mechanism of TIAs

There are two possible mechanisms for the development of TIAs: emboli to the intracranial arteries, and temporary cerebral hypoperfusion. The older literature commonly attributed TIAs to transient decreases in systemic factors such as blood pressure and cardiac output. This theory predicts that TIAs would occur only in patients with hemodynamically significant stenoses. Recent studies, however, have failed to demonstrate a good correlation between the severity of carotid stenosis and prognosis or clinical manifestations of hemispheric, carotid-territory TIAs. Further, cardiac disease was less common in patients with carotid TIAs than in the nonhemispheric TIA group. These observations support the concept of an embolic cause for carotid-territory TIAs. Nonhemispheric TIAs are likely to be embolic when the symptoms are focal, but global symptoms are more likely to be caused by transient hemodynamic compromise in the presence of posterior circulation disease.

Pathologic specimens of carotid artery lesions from symptomatic patients frequently contain irregular luminal surfaces with exposed subintimal structures. Areas of the vessel lacking the normal intimal layer are termed *ulcers*. Aggregates of platelets and fibrin as well as cholesterol crystals are commonly observed in carotid artery ulcers (Fig. [20-31](#)).

There are at least three possible causes of embolization from a carotid bifurcation plaque. First, fragments of the cholesterol-calcium plaque may break off and be discharged into the lumen. Second, roughened, thrombogenic subintimal structures may be exposed to the flowing blood, and platelet thrombi may be formed that are easily detached. A third mechanism involves intramural carotid artery hemorrhage. The pathophysiologic mechanisms underlying this process are incompletely understood, but the presence of acute or recent intramural hemorrhage has been reported in over 90 percent of symptomatic patients, compared with less than 30 percent of asymptomatic patients. It is possible that intramural hemorrhage results in rapid, unpredictable progression of a moderate asymptomatic carotid artery lesion to a high-grade symptomatic stenosis, with eventual rupture of the intramural process and discharge of the plaque contents into the arterial lumen.

There are several unusual causes of carotid artery pathologic conditions that may be associated with cerebrovascular symptoms. The artery may be elongated, tortuous, or kinked;

these anatomic abnormalities are rarely associated with symptoms. Radiation therapy may induce a symptomatic carotid artery injury similar in appearance to an atherosclerotic lesion. The internal carotid artery may undergo spontaneous dissection, resulting in neurologic symptoms and thrombosis of the vessel. Treatment of carotid artery dissection is nonoperative, with anticoagulation therapy and control of hypertension.

Fibromuscular dysplasia is the most common nonatherosclerotic lesion of the carotid artery. The cause of the process is unknown, but it tends to involve long arteries with few branches. The vast majority of affected individuals are females, and the disease is usually bilateral. Four histologic types have been described: intimal fibroplasia, medial hyperplasia, medial fibroplasia, and perimedial dysplasia. Medial hyperplasia is the most frequently encountered variety. Operative intervention is indicated in symptomatic patients, usually in the form of intraluminal dilatation of the involved segment, with an open or a percutaneous approach. Prophylactic operation in the asymptomatic patient is not recommended, since no objective data exist on the natural history of the process.

Patients with repeated *TIA*s often experience symptoms in exactly the same anatomic region of the brain with each episode. For example, a patient with crescendo *TIA*s involving right arm monoplegia will frequently exhibit repeated identical episodes without other deficits such as aphasia or left-sided amaurosis fugax. It is reasonable to wonder why the pattern of neurologic dysfunction is frequently reproduced identically, since one would expect carotid embolization to occur in a random fashion throughout the distribution supplied by that carotid vessel. This apparent paradox is answered by the fluid dynamics associated with laminar flow. Embolic material originating from a point source in the arterial system tends to travel to the same terminal arterial branch, as illustrated by Millikan. He serially injected small metal pellets through a needle placed in the internal carotid artery of monkeys. Subsequent examination of the animals' brains revealed that the pellets stacked up one on the other within the same cortical branch (Fig. [20-32](#)).

Diagnosis

The most valuable tool in the diagnosis of carotid artery disease is a careful history and physical examination. A thorough neurologic evaluation is of great importance. The diagnosis of carotid bifurcation disease is facilitated by the relatively superficial location of the vessel, rendering it accessible to study. The cervical carotid pulse is usually normal in patients with carotid bifurcation disease, since the common carotid artery is the only palpable vessel in the neck and is rarely diseased. The superficial temporal artery pulse provides some indication of the status of the external carotid artery, but the internal carotid artery pulsation is inaccessible to digital palpation. Auscultation has been the most widely employed method of assessing the carotid bifurcation. Carotid bifurcation bruits may be heard just anterior to the sternocleidomastoid muscle, near the angle of the mandible. Auscultation in the supraclavicular fossa may reveal bruits originating from stenotic lesions of the subclavian artery. Carotid and subclavian artery bruits must be differentiated from one another as well as from murmurs due to cardiac lesions such as aortic stenosis, which may be transmitted along the great vessels into the neck. Bruits do not develop until the stenosis is large enough to reduce the diameter by approximately 50 percent. Bruits may be absent in extremely severe lesions because of the extreme reduction of flow across the stenosis.

Oculoplethysmography

The development of noninvasive carotid artery testing modalities has provided more precise information regarding the nature and severity of the lesion. Oculoplethysmography (OPG) is a test designed to measure ophthalmic artery pressure indirectly through the use of plastic suction cups applied to the sclera. The ophthalmic artery pressure is a good indicator of internal carotid artery pressure, and abnormalities in this measurement can identify hemodynamically significant carotid artery stenoses. Stenoses of the common carotid artery and the ophthalmic artery must be excluded, as these lesions will produce an abnormal OPG in the absence of significant carotid bifurcation disease.

Duplex Ultrasonography

The most useful noninvasive test is the Duplex ultrasound carotid artery examination. This test combines pulsed-Doppler measurements with a B- mode ultrasound image to provide data on the velocity of blood flow and the anatomic profile of the carotid bifurcation, respectively. The Duplex scan is an extraordinarily accurate noninvasive means of identifying carotid artery stenoses. The subclavian and vertebral arteries also may be evaluated, but the accuracy in identifying disease is much lower in these vessels. The technical difficulty of the test has been largely overcome with the introduction of the color-flow Doppler. This instrument is a modified Duplex scanner that provides a color image in which the velocity and direction of blood flow are keyed to the color of the image at all points within the vessel so that arterial flow can be seen as red and venous as blue. Standard Duplex scans cannot assess the cerebral arterial circulation beyond the first few centimeters of the internal carotid artery. A transcranial Doppler has been developed to evaluate the middle cerebral artery and other intracranial vessels, using a low-frequency Doppler signal to penetrate the thin bone of the temporal and occipital regions. The test is most easily performed when coupled with an ultrasound image (transcranial Duplex), allowing the technologist to visualize the intracranial vessels on the color- flow ultrasound image, measuring blood velocities by placing the pulsed- Doppler within the exact anatomic area of interrogation.

Magnetic Resonance Imaging

Magnetic resonance imaging has been used as a noninvasive means of imaging the carotid arteries, displaying the image as a two-dimensional scan of the carotid system analogous to conventional arteriography. The procedure is called magnetic resonance angiography (MRA). The resolution of early MRA was insufficient to accurately define the location and extent of carotid artery disease, but technical improvements in software and hardware have markedly improved the image quality of the scans.

Arteriography

Arteriography remains the gold standard for the diagnosis of extracranial arterial disease. Intraarterial standard and digital-subtraction arteriograms provide the best images and usually are performed via the transfemoral approach, with views obtained in the anteroposterior and lateral planes (Fig. [20-33](#)). Most images are of sufficient quality to allow precise determination of the severity of the carotid artery lesion, expressed as the percent diameter reduction in comparison to the normal artery distal to the stenotic process (Fig. [20-34](#)). Intravenous digital images have been disappointing, with images less precise than those obtained with Duplex ultrasonography.

The major problem with contrast arteriography of the carotid system has been a significant incidence of stroke associated with the procedure. Some studies have documented a rate of cerebral infarction of 1 percent. Because of this, along with the increasing accuracy of Duplex ultrasonography, many surgeons have forgone preoperative contrast arteriography in the majority of patients undergoing carotid endarterectomy.

Treatment

The therapy of cerebrovascular occlusive disease is best considered by dividing patients into two subgroups: those with posterior circulation lesions involving the subclavian or vertebral arteries, and those with anterior circulation lesions involving the carotid arteries.

The indications for operation are relatively straightforward and noncontroversial in patients with subclavian and vertebral artery disease; operative correction is recommended for symptomatic lesions that are surgically accessible. Proximal subclavian artery lesions are best treated with a carotid-to-subclavian artery bypass, using end-of-graft to side-of-artery anastomoses at each end. Alternatively, a subclavian-to-carotid artery transposition may be performed, with division of the subclavian artery proximal to the origin of the vertebral artery and construction of an end-of-subclavian to side-of-carotid artery anastomosis. Embolic symptoms caused by lesions of the subclavian artery require exclusion of the embolic source from the arterial stream. Ligation of the subclavian artery proximal to the vertebral artery origin is performed in the case of a carotid-to-subclavian artery bypass. Disease involving the vertebral arteries may also be responsible for posterior cerebral symptoms. Vertebral artery lesions usually occur at the origin of the vessel and end within the first centimeter of its course. The stenotic lesion may be endarterectomized and the arteriotomy closed with a patch graft. More commonly, the vertebral artery is transected beyond the stenosis, the stump is ligated, and the distal end is anastomosed to the side of the common carotid artery.

The medical management of symptomatic carotid artery disease includes reduction of risk factors such as smoking and hypercholesterolemia, and the use of antiplatelet agents. Current recommendations include a regimen of one to four 325-mg aspirin tablets daily, although recent studies have suggested that doses as low as 30 mg/day produce similar antiplatelet effects. Ticlopidine and other antiplatelet agents also have been used with success. Dipyridamole was once widely employed, both alone and in combination with aspirin, but the drug has not been shown to be more effective than aspirin alone. Oral anticoagulation therapy with warfarin has been recommended for patients who continue to experience symptoms while they are taking antiplatelet agents.

A number of well-designed, randomized trials of antiplatelet therapy for symptomatic carotid artery disease have been undertaken. Although all the studies revealed a trend toward reduction of the risk of subsequent stroke, none documented a statistically significant benefit of aspirin compared with placebo. It was only when the end points of mortality, myocardial infarction, *TIA*, and stroke were combined that a significant advantage was achieved with antiplatelet therapy. The effects of aspirin on the incidence of myocardial infarction and on mortality were the most important benefits; the effects on cerebrovascular disease were less striking. A metaanalysis of these results suggested a 15 percent reduction in stroke risk with aspirin and a 40 percent reduction in myocardial infarction and deaths.

The benefit of carotid endarterectomy for patients with symptomatic cerebrovascular disease has recently been established (Table [20-7](#)). The North American Symptomatic Carotid Endarterectomy Trial (NASCET) and the European Carotid Surgery Trial (ECST) have documented a significant reduction in cerebrovascular events following the procedure compared with patients managed medically. The NASCET study randomized patients with symptomatic carotid artery disease to either carotid endarterectomy or optimal medical care including antiplatelet therapy (generally 1300 mg aspirin daily). In the subset of over 600 patients with stenoses that reduced artery diameter by 70 to 99 percent, a 26 percent incidence of ipsilateral stroke was observed in the medically managed group at 2-year follow-up, compared with a 9 percent incidence in the surgical group, a difference that was statistically significant and prompted early termination of the study. On the basis of these results, patients with symptomatic high-grade carotid artery stenoses should undergo carotid endarterectomy by a surgeon with the low rate of perioperative morbidity that has characterized the centers in the NASCET trial. Patients with minimal stenoses (less than 30 percent diameter reduction) are best managed medically, reserving operative considerations for recurrent symptoms on antiplatelet therapy or for disease progression. The management of the group of patients with moderate stenoses (30 to 69 percent) has not been settled and awaits the results of ongoing randomized trials.

The appropriate treatment of asymptomatic carotid artery stenosis has become less controversial with the availability of data from the Asymptomatic Carotid Artery Study (ACAS). This trial documented an improvement in clinical results when patients with carotid artery lesions with reductions of 60 percent or more in vessel diameter underwent surgical repair. Carotid endarterectomy appeared to halve the risk of cerebral events at 5-year follow-up, although the overall rate of stroke was relatively low. Moreover, the differences were most prominent in men. Despite these limitations, the ACAS data would appear to suggest that surgical candidates with severe carotid artery lesions should undergo repair whether or not they have symptoms.

Patients with nonhemispheric cerebral symptoms appear to benefit from correction of carotid artery lesions, if the symptoms are classic and the stenosis is significant (greater than 60 percent diameter reduction). Patients with nonclassic symptoms such as dizziness or syncope are rarely improved by carotid artery reconstruction. These patients should be managed as though the lesions were asymptomatic, reserving operation for the healthy patient with a relatively long life expectancy and a high- grade carotid artery lesion.

The timing of operation is a critical consideration in the treatment of carotid artery disease. Patients with unstable neurologic symptoms such as crescendo *TIA*s or a waxing-and-waning deficit invariably have severe disease and should undergo urgent arteriography and operation. Because of the high risk of cerebral infarction from recurrent embolization or carotid artery thrombosis, patients with these symptoms should be started on anticoagulation therapy at the time of presentation. Patients with a completed stroke

should undergo carotid endarterectomy a minimum of 4 weeks after the event, as the risk of perioperative intracerebral bleeding is high immediately after a stroke. Emergency carotid endarterectomy is indicated for an acute carotid artery thrombosis if the procedure can be performed within a few hours after the event. It is unusual to have the opportunity to restore carotid artery perfusion so expeditiously; the possibility arises in early postoperative carotid artery occlusion or when occlusion occurs during cerebral arteriography. Operation is contraindicated for an occlusion that is several hours old, since the danger of intracerebral edema and hemorrhage from reperfusion outweighs the potential benefits of revascularization.

Carotid endarterectomy can be performed under local or general anesthesia. The advantage of local anesthesia is that the surgical team can be aware of the patient's mental status at all times, most importantly when the carotid artery is clamped. Changes in level of consciousness or motor deficits at the time of interruption of blood flow signal the need for placement of an intraoperative shunt. General anesthesia confers the advantages of ventilatory support and increased cerebral blood flow with the use of halogenated agents, but alternative methods of assessing the adequacy of cerebral perfusion must be implemented.

The exposure of the carotid bifurcation is accomplished with a vertical incision made along the anterior border of the sternocleidomastoid muscle (Fig. 20-35). The dissection is carried down through the platysma muscle, and the internal jugular vein is mobilized laterally after ligation and division of the facial vein. The carotid bifurcation is exposed along with an adequate length of the common and internal carotid arteries to allow clamping of these structures above and below the plaque. The hypoglossal nerve can be mobilized and reflected cranial for additional exposure. Exposure of the distal internal carotid artery can be achieved by dividing the posterior belly of the digastric muscle when needed; further proximal exposure can be accomplished by dividing the omohyoid muscle. The patient is systemically heparinized, the carotid vessels are clamped, and a longitudinal arteriotomy is started on the common carotid artery and carried onto the internal carotid artery, with the distal termination on normal vessel. The plaque is removed using an endarterectomy spatula, and every attempt is made to achieve a smooth, tapering distal end point. The termination of the endarterectomy should have a normal adherent intima and media to ensure that the resumption of forward blood flow will not create an occlusive flap. The arteriotomy may be closed primarily with nonabsorbable suture. Alternatively, a vein or prosthetic patch may be employed if the vessel is small.

In a small proportion of patients the collateral cerebral circulation is inadequate to compensate during the period of carotid artery cross-clamping. The use of an indwelling shunt intraoperatively reduces the incidence of intraoperative cerebral infarction in these patients. Some have advocated the routine use of a shunt because it is difficult to predict which patients will not tolerate the period of relative cerebral ischemia. There are disadvantages to the use of a shunt; the shunt can injure the intima at the time of placement, and it impedes visualization of the end point of the endarterectomy. For these reasons, intraoperative electroencephalographic (EEG) monitoring and internal carotid artery stump pressure determination have been employed in an effort to identify patients at greatest risk of stroke with temporary carotid interruption. EEG leads are placed on the patient's scalp preoperatively; any slowing of the EEG waveform is indicative of cerebral ischemia and necessitates the insertion of a shunt. Carotid artery stump pressures below 25 mmHg after clamping also are predictive of an increased risk of cerebral infarction during clamping and mandate the placement of a shunt. Use of either EEG or stump pressure determination allows the surgeon to limit the use of indwelling shunts to less than one in five patients undergoing carotid procedures.

The complications of carotid endarterectomy include cranial nerve damage resulting from nerve division, excessive traction, or perineural dissection. The most frequently injured nerves are the vagus, the hypoglossal, the glossopharyngeal, and the marginal mandibular branch of the facial nerve. Revascularization of a severely stenotic carotid artery may result in the "hyperperfusion syndrome," characterized by headache, seizures, and, occasionally, intracranial bleeding. This phenomenon is believed to result from reperfusion of a chronically ischemic tissue bed in which the arterioles are maximally dilated. With the sudden restoration of blood flow, the area becomes markedly edematous and the clinical picture described above develops. Beta blockers may reduce the severity of symptoms and should be instituted in high-risk patients with the first signs of the syndrome, usually a headache.

The most dreaded complication of carotid endarterectomy is perioperative stroke. Perioperative strokes can occur as a result of inadequate cerebral perfusion during the clamp period, embolization of debris from the plaque at the time of operation, or early postoperative thrombosis. Refinements in operative technique have reduced the perioperative stroke rate to less than 5 percent in most institutions and to less than 2 percent in centers with the greatest experience.

The incidence of perioperative myocardial events is distressingly high; they are at least as common as perioperative neurologic events in patients undergoing carotid endarterectomy. Preoperative cardiac screening procedures include myocardial imaging using thallium or echocardiography, often supplemented with exercise, dipyridamole, or dobutamine. In a study performed at the Cleveland Clinic, routine coronary arteriography was used as a screening test in a series of 1000 patients in whom elective vascular procedures were planned; 295 of these patients presented because of pathologic cerebrovascular conditions. Significant coronary lesions were detected in 33 percent of patients with symptoms of cardiac disease and in 17 percent of patients without clinical manifestations of cardiac disease. The 75 percent 5-year survival rate observed in patients who underwent staged cardiac and carotid artery revascularization procedures was surpassed only by that of the group of patients with normal or minimally diseased coronary arteries (91 percent). Given these results, it seems reasonable to use one of the minimally invasive screening tests, such as dipyridamole-thallium studies, before carotid endarterectomy. Coronary arteriography is reserved for patients with overt cardiac symptoms or a positive screening test, and coronary artery bypass is used for patients with severe arteriographic coronary artery disease.

NONATHEROSCLEROTIC DISORDERS

Thoracic Outlet Syndrome

The thoracic outlet is the space through which the subclavian artery and vein and the brachial plexus pass from the neck into the upper extremity. Its anatomic boundaries are the chest wall, the scalene muscles, the clavicle, and, potentially, a variety of anomalous, sometimes compressive structures such as fibrous bands or cervical ribs (Fig. 20-36). Proper treatment of *thoracic outlet syndrome (TOS)* requires a detailed history and physical examination, appropriate diagnostic tests, and understanding of the intricate anatomic relationships in this area.

Arterial Component

The arterial complications of *TOS* are caused by a bony cervical rib or an anomaly of the first rib. Patients may present with an asymptomatic pulsatile cervical mass, or, more often, with upper-extremity ischemia ranging from unilateral Raynaud's phenomenon to acute ischemia with absent pulses. Symptoms are caused by atheroemboli from a poststenotic dilatation or true aneurysm (rarely thrombosis) of the subclavian artery (Fig. 20-37). The evaluation should include cervical x-rays, noninvasive vascular testing, and arteriography when appropriate.

Treatment requires removal of the embolic source, resection of the bony anomaly, and reperfusion, if possible, of the ischemic extremity. The subclavian artery aneurysm is best approached through a supraclavicular incision with or without removal of the medial half of the clavicle. The aneurysm is resected and replaced with an interposition saphenous vein graft. The cervical and first ribs can be excised through this approach, or the patient can be repositioned and the operation completed through the axilla. If the artery is dilated but not aneurysmal, resection is not indicated, because bony decompression is sufficient to prevent further atheroemboli in most instances. In this situation the transaxillary approach is preferred, because it simplifies removal of the first rib. Any distal embolic material that must be removed should be approached through separate arteriotomies in the arm. Sympathectomy can be a useful adjunct when distal emboli are irretrievable.

Venous Component

Venous obstruction of the upper extremity is caused by a narrowing of the costoclavicular space between the medial aspect of the first rib and the clavicle. This is the site where the axillary vein passes over the rib and under the clavicle to join the internal jugular vein. Both hyperabduction of the arm and hyperextension of the shoulders can narrow this space, causing venous obstruction. Venous obstruction takes one of three forms: intermittent obstruction, acute thrombosis, and postthrombotic intermittent obstruction.

Intermittent Obstruction

These patients present with arm swelling, cyanosis, and pain when the arm is abducted or the shoulders hyperextended. The diagnosis can be made by phlebograms with the arm in both the relaxed and symptomatic positions. A positive examination shows a beaklike appearance in the vein proximal to the first rib when the arm is stressed and a normal venous anatomy with the arm in a neutral position. Venous pressure measurements also can be made in the same arm positions. Treatment consists of transaxillary first rib resection or medial subtotal claviclectomy.

Acute Thrombosis

Patients usually are young and healthy and present with the sudden onset of arm pain, swelling, and cyanosis. The problem often follows some repetitive activity, such as throwing a ball, house painting, paper hanging, swimming, or rowing a boat, and has been called the Paget-Schroetter syndrome or "effort thrombosis." Phlebograms show a complete obstruction of the subclavian vein, often with thrombus distally in the axillary vein.

Treatment options include elevation and heparinization, venous thrombectomy, and local thrombolysis; the latter has gained much support in recent years. Once the diagnosis is confirmed, a coaxial catheter is inserted into the basilic vein and placed directly into the thrombus. Thrombolytic infusion is begun and its progress monitored with serial venography. If successful thrombolysis uncovers an underlying stenotic lesion, decompression of the thoracic outlet should be undertaken (usually clavicular resection), with simultaneous surgical repair of the venous lesion (patch angioplasty or venous bypass). The results of percutaneous balloon dilation of these lesions have been unsatisfactory. Stents should almost never be used in this group of patients, because they may be compressed between the first rib and clavicle. The patient is discharged on warfarin, which is continued for 3 months.

Postthrombotic Intermittent Obstruction

Patients with acute obstruction and unsuccessful clot removal, whether chemical or mechanical, have a 50 percent chance of developing residual symptoms of venous obstruction. Venograms usually demonstrate an occluded vein with large collateral vessels around the first rib (Fig. 20-38). Hyperabduction of the arm results in compression of these collateral veins. Either first rib resection or a medial claviclectomy will relieve these symptoms in some patients. Direct repair of the chronically occluded subclavian vein may be preferable and can be achieved by mobilizing the internal jugular vein and turning it down for anastomosis into the divided patent axillary vein.

Neurologic Component

The subjective nature of the symptoms and the lack of objective diagnostic criteria make the management of the neurologic component of *TOS* potentially very difficult. Some would restrict this diagnosis to only those patients with symptoms and signs limited to the T1 nerve root (ulnar nerve), while others would broaden it to include any neurologic symptoms of the neck, upper back, and upper extremity. These symptoms are exacerbated by elevation and abduction of the arm. Trauma may precipitate the symptoms in a susceptible individual.

An accurate and complete history is important, because the diagnosis is often one of exclusion. Most elements of the differential diagnosis have reliable diagnostic tests. All these

patients should have cervical spine films to identify any bony anomalies and to rule out cervical disc protrusion or spondylitis. Nerve conduction studies are indicated to rule out carpal tunnel syndrome and ulnar nerve compression at the elbow. Orthopaedic and neurologic consultations may be necessary to rule out specific pathologic conditions of the shoulder, multiple sclerosis, and spinal cord tumors. Physical examination includes blood pressure measurements in both arms. The hands are examined for signs of atrophy of the ulnar nerve–innervated interosseus muscles and the median nerve–innervated thenar muscles (Fig. 20-39). Percussion over the median nerve (Tinel's test) and rapid wrist flexion (Phalen's test) are performed to further evaluate the median nerve. A complete neurologic examination of the neck and upper extremity is performed, including the application of pressure in the supraclavicular space over the brachial plexus. The traditional Adson's test is totally unreliable in detecting brachial plexus compression and is of no use in making the diagnosis of neurologic TOS. The elevated arm stress test (EAST) described by Roos has the patient raise the arm to 90 degrees and open and close the hands for 3 min. This may reproduce the patient's symptoms but is unfortunately also positive in 90 percent of patients with carpal tunnel syndrome.

When neurologic TOS is the considered diagnosis and treatment is indicated, a conservative approach should always be followed. Patients with severe pain and cervical muscle spasm are initially treated with physical therapy directed at relieving the muscle spasm. Peets' shoulder strengthening exercises are started as the pain subsides. Methods of opening the costoclavicular space by hunching the shoulders upward and forward are used when the patient first feels symptoms recurring. Indications for operation include failed physical therapy, intractable pain, and/or progressive neurologic dysfunction.

Technique of First Rib Resection

Clagett suggested in 1962 that the first rib was the “common denominator” in the various compression syndromes of the thoracic outlet and recommended its resection in appropriate cases. In 1966 Roos described the technique of transaxillary first rib resection, which, because of its cosmetic appeal, simplicity, and safety has become the most frequently performed operation for neurologic TOS.

With the patient in the lateral decubitus position with the arm elevated, a skin incision is made in the axillary hair line between the pectoralis major and latissimus dorsi muscles. The first rib is reached by blunt dissection in the axillary tunnel, taking care to avoid the intercostobrachial nerve. The subclavian artery and vein are identified and the subclavius muscle tendon divided. The anterior scalene muscle can now be identified and divided at the point where it inserts on the first rib anterior to the artery. At this time a digital search for anomalous bands is performed. These may originate from the C7 transverse process, from an incomplete cervical rib, from attachment to two places on the first rib, or from the middle scalene muscle. After any bands have been divided, the middle scalene muscle and the intercostal muscle attachments are pushed off the first rib. When all muscle fibers are cleared and the T1 nerve root is visualized and protected, the rib is divided and removed. The wound is irrigated with saline to detect any pneumothorax, which, if present can be treated by inserting a small chest tube into the pleural space. The tube can be removed in the recovery room if the lung is fully expanded and there is no air leak.

A number of brachial plexus injuries have been reported with this approach, and there are proponents of a supraclavicular approach that avoids any traction on the brachial plexus.

Popliteal Artery Disease

There are two causes of nonatherosclerotic popliteal artery disease that characteristically produce symptoms of calf claudication in young males: popliteal entrapment syndrome and adventitial cystic disease of the popliteal artery.

Popliteal Entrapment Syndrome

Calf claudication in a young male should suggest the diagnosis of popliteal entrapment syndrome. The popliteal artery normally traverses the popliteal fossa between the two heads of the gastrocnemius muscle, but the artery courses medial to the medial head of the gastrocnemius or popliteus muscle in patients with the disorder (Fig. 20-40). The popliteal vein accompanies the artery and is entrapped in just over 10 percent of the cases. Popliteal entrapment is infrequently encountered in women, with a male-to-female ratio of 151. Bilateral involvement has been observed in one-quarter of the cases.

The findings on physical examination are dependent on whether the popliteal artery is occluded at the time of presentation. The majority of patients present with thrombosis of the vessel and absent pedal pulses. The pedal pulses are present in only one-third of the patients with compressed but nonoccluded vessels. Passive dorsiflexion and active plantar flexion of the foot tense the gastrocnemius muscle, which compresses the artery, obliterating the pedal pulses. Femoral arteriography with and without plantar flexion of the foot may be instrumental in confirming the diagnosis. Three arteriographic abnormalities of the popliteal artery have been described: medial deviation, occlusion, and poststenotic dilatation. A CT scan of the popliteal fossa can be indispensable if the popliteal artery is occluded, defining an abnormal position of the artery with respect to its surrounding muscles.

Operative intervention is indicated in both symptomatic and asymptomatic patients. A posterior approach to the artery is preferred if the artery is patent and extensive proximal and distal exposure is not necessary. Lysis of the constricting head of the gastrocnemius muscle is all that is required in these instances. Complete resection of the medial head of the gastrocnemius muscle is well tolerated, with minimal alteration of function.

Intraarterial thrombolytic agents or balloon catheter thrombectomy is used in the recently occluded popliteal artery, followed by lysis of the compressing band if the artery appears otherwise normal. Chronic occlusions necessitate a bypass of the involved segment, with exposure of the proximal popliteal or superficial femoral artery and the distal popliteal or tibial artery. A medial approach offers advantages in this subgroup, as it facilitates exposure of the inflow and outflow vessels and provides access to the greater saphenous vein.

Adventitial Cystic Disease of the Popliteal Artery

Adventitial cystic disease of the popliteal artery is a rare disorder characterized by cystic degeneration of the adventitia of the artery, producing extrinsic compression of the lumen at the level of the knee joint. The cause of the process is unclear, but the two most widely accepted theories invoke a pathogenesis similar to simple ganglia of the wrist, with formation of the cystic cavities as a result of developmental rests of mucin- secreting cells within the adventitia, or abnormal connections between the synovial space of the knee and the wall of the popliteal artery. The cyst contents consist of a viscous material high in hyaluronic acid, similar to the fluid found in ganglia. The fluid is colorless in most instances but may take on the appearance of currant jelly if hemorrhage into the cyst has occurred. The cyst may be unilocular or septate.

Popliteal adventitial cystic disease generally produces symptoms of calf claudication in middle-aged men. Arteriography reveals curvilinear stenotic lesions in two-thirds of the patients; popliteal artery occlusions are observed in the remaining one-third of the cases. Ultrasonography and CT are useful in delineating the cyst and its relationship to the arterial lumen.

Although percutaneous cyst aspiration under ultrasound or CT guidance has been used in the treatment of this disorder, the long-term results have been discouraging, with rapid reaccumulation of the cyst fluid. Operative exploration, incision into the cyst, and evacuation of the contents is the most appropriate treatment if the stenotic process has not progressed to occlusion. Autogenous vein bypass is necessary in the presence of popliteal artery occlusion, with the level of the proximal and distal anastomoses determined by the extent of propagated clot.

Vasospastic Disorders

Arterial vasospasm in the extremities occurs in a variety of clinical situations, including immunologic disorders such as scleroderma and systemic lupus erythematosus, the thoracic outlet syndrome, mechanical (vibratory) or cold-induced small-vessel injury, and the use of drugs such as ergotamine and oral contraceptives.

Raynaud's Syndrome

Raynaud's syndrome is the prototypic symptom complex associated with peripheral vasospasm. “Raynaud's disease” and “Raynaud's phenomenon” are older terms used to describe a primary benign disorder and a secondary more virulent process, respectively. Improvements in immunologic testing have resulted in a decrease in the frequency of the primary classification, and today over 50 percent of patients with severe Raynaud's symptoms have documented autoimmune disease. Use of the term “Raynaud's syndrome” may be more appropriate than the artificial separation of the entity into primary and secondary forms.

Raynaud's syndrome is characterized by episodic cutaneous color changes consisting of sequential pallor, cyanosis, and rubor. It is observed most frequently in the digits of the upper extremities but also may affect the toes. The explanation for the white, blue, and red color changes lies in the pathophysiology of the vasospasm. Initially, cold exposure or emotional stress precipitates intense small-artery vasospasm, with digital blanching and numbness. A cyanotic hue develops as partial arterial perfusion is restored and the cutaneous venules become filled with desaturated blood. Finally, the vasospasm resolves and the digits turn bright red as a result of reactive hyperemia.

Between 70 and 90 percent of patients with Raynaud's syndrome are female, almost all younger than age 40. Raynaud symptoms affect the majority of the general population at one time or another, but only a small percentage have symptoms that are severe enough to cause them to seek treatment. Complications include sclerodactylia (atrophy of skin and loss of elasticity, resembling the changes seen in scleroderma), recurrent paronychia infection, and digital ulceration and gangrene. Amputation of the fingers or toes is occasionally necessary, but fortunately the process almost never progresses to involve structures proximal to the digits.

The treatment of Raynaud's syndrome is initially conservative, with avoidance of tobacco, cold exposure, and drugs that have been reported to exacerbate the symptoms, such as oral contraceptives, beta blockers, and ergotamine. Pharmacologic therapy with vasodilating calcium channel blocking agents such as nifedipine has provided partial resolution of symptoms in many patients and represents the mainstay of treatment of the disorder. Cervical sympathectomy has been used in the treatment of severe Raynaud's syndrome, but results have not been gratifying.

Acrocyanosis

Acrocyanosis is a vasospastic disorder occurring almost exclusively in women. It is characterized by persistent edema, coolness, and cyanosis of the hands, lower legs, and feet. The pathophysiology is that of cutaneous arteriolar vasospasm. The disease is not as dependent on the temperature of the environment as Raynaud's syndrome, and the process never progresses to tissue loss. Vasodilating agents have been beneficial in acrocyanosis, as has avoidance of cold.

Livedo Reticularis

Livedo reticularis is a condition characterized by constant cyanotic mottling of the skin of the lower legs and feet. The upper extremity occasionally is involved. The changes are always present but become more prominent with exposure to cold. Livedo reticularis occurs as a result of the random spasm of cutaneous arterioles in association with the secondary dilatation of venules to produce a reticulated pattern. The disorder occasionally occurs in association with such disorders as systemic lupus erythematosus and periarteritis nodosa. Most patients have no associated diseases and the process is one of cosmetic concern. Avoidance of cold is the only treatment indicated.

Causalgia

Causalgia (“burning pain”), also known as “posttraumatic reflex sympathetic dystrophy,” is a painful disorder that develops after incomplete nerve transection. Vasomotor dysfunction is almost always present. The most frequent causes are penetrating missile injuries, fractures, and crush injury. Patients complain of burning pain in the peripheral portions of the extremity, and the symptoms are not limited to the area supplied by the injured nerve. A characteristic feature of the cutaneous dysesthesia is that its intensity is such that the patient cannot tolerate contact of the affected area with clothing or bed sheets. The pain results in limitation of motion and chronic disability. The vasospasm is associated with cyanosis, edema, hyperhidrosis, and coolness of the extremity. Surgical sympathectomy is the treatment of choice and is successful in relieving pain in the vast majority of patients.

Inflammatory Arteritis

The term “inflammatory arteritis” refers to an arterial inflammatory response arising from a group of diseases of unknown or immunologic cause. Many of the arteritides are associated with inflammatory changes in the veins as well, and the term “vasculitides” may be more appropriate when describing these disorders. The vasculitides can be classified by the size of the vessels involved and by the coexistent clinical features of the disease process. An immunologic mechanism has been defined in almost all the vasculitides, and most are treated with systemic steroids or cytotoxic agents.

Giant Cell Arteritides

Two disorders comprise the giant cell arteritides: temporal arteritis and Takayasu's arteritis. These diseases are identical in histologic appearance and are associated with similar laboratory abnormalities. Anatomically, however, Takayasu's arteritis involves the aorta and its branch vessels near their origin, whereas extracranial lesions occur in only 9 percent of patients with temporal arteritis. Takayasu's arteritis occurs almost exclusively in female patients in their teens and twenties. The female predominance is less in temporal arteritis, averaging 31, and patients are typically over the age of 50 years.

Temporal arteritis classically begins with a prodromal phase of malaise, myalgias, headache, and low-grade fever. These symptoms are followed by a second, quiescent phase. A palpable, tender temporal artery may be found in some patients. Jaw claudication is a frequent complaint, and an abnormal temporal artery biopsy is found in about 50 percent of patients. The association between polymyalgia rheumatica and temporal arteritis is so striking that some have advocated temporal artery biopsies in all patients with polymyalgia rheumatica to rule out occult arteritis. An elevated erythrocyte sedimentation rate is a uniform finding in temporal arteritis, and the diagnosis should be questioned if the sedimentation rate is normal. Ocular complications in the form of unilateral or bilateral visual loss occur as a result of ischemic optic neuritis. Upper-extremity claudication may occur in the rare cases of extracranial involvement. When extracranial manifestations develop, the characteristic arteriographic finding is that of bilateral smooth tapering or occlusion of the axillary and brachial arteries. The ocular and peripheral arterial complications of temporal arteritis do not generally develop until several months after the onset of symptoms, and the disease is self-limited, with cessation of the process after a period of several years.

Steroid therapy should be initiated early to prevent sudden visual loss from ophthalmic artery thrombosis. Steroids frequently result in remarkable resolution of stenotic lesions over a period of several months, and the erythrocyte sedimentation rate should be monitored as an indicator of the efficacy of treatment. Operative revascularization is sometimes necessary for occluded extracranial arterial lesions, but reconstructions tend to thrombose if undertaken during the active phase of the disease and should be delayed until inflammation has been adequately suppressed with steroids.

Takayasu's arteritis, also known as “pulseless disease,” is a rare disorder associated with stenoses and aneurysms of the aortic branch vessels. Like temporal arteritis, acute and chronic phases occur. Systemic symptoms and laboratory abnormalities during the acute phase parallel those of temporal arteritis, except for associated findings of erythema nodosum and arthralgias with synovial changes typical of rheumatoid arthritis in some patients. Diagnostic arteriography should include the entire aorta and its branches. Steroid therapy is instituted initially, and the erythrocyte sedimentation rate is monitored as an index of the response to treatment. Cytotoxic agents have been of some benefit. Operative reconstructive procedures are reserved for lesions unresponsive to medical therapy, but they should be delayed beyond the acute disease phase if possible.

Buerger's Disease

Buerger's disease, also known as thromboangiitis obliterans, is an inflammatory vasculopathy occurring in medium-sized and small arteries of young male smokers. The disease is exceedingly rare in females and is not observed in nonsmokers. The lesions occur in the upper and lower extremities and in superficial veins as well as the arteries. The entity was first described by Winiwarer in 1879 and later by Buerger in 1908. Initially Buerger's disease was thought to occur exclusively in the Jewish population, but subsequent studies have shown this to be false, with other populations having about the same incidence.

The cause of Buerger's disease remains obscure. There is evidence for an autoimmune pathogenesis for the disease, with increases in complement factors and collagen antibody levels. The disease also has been linked to the presence of certain human lymphocyte antigens (HLAs). Smoking is the most important risk factor in the development and progression of the disease process, although the mechanism behind its effects are unknown.

The histopathologic features of Buerger's disease are those of a panangiitis, involving all layers of the vessel wall. Lymphocytes and fibroblasts infiltrate the media and adventitia of the artery in the early stages of the disease. The occluding thrombus is involved with an inflammatory process as well, with multinucleated giant cells and leukocytes giving the appearance of microabscesses within the clot. The late lesion of Buerger's disease is characterized by an occluded, contracted artery with a marked fibrotic reaction in the adventitia, media, and intima. The lesions tend to occur in a localized, segmental fashion, with normal vessel segments interposed between involved segments. The vein and adjacent nerve may be tightly bound to the artery in this dense fibrotic process. The tibial arteries and the vessels of the foot are the predominant sites of involvement in the lower extremity. Approximately 30 percent of patients with Buerger's disease have involvement of the upper extremities, and these lesions occur principally in the vessels of the forearm and hand.

The clinical presentation of patients with Buerger's disease is distinct from that of patients with atherosclerotic disease. Involvement of the smaller arteries may produce symptoms of rest pain and gangrene without antecedent claudication. Necrotic lesions commonly develop at the tips of the fingers and toes. Recurrent superficial thrombophlebitis may develop in the upper or lower extremity. Therapy is directed against the inciting effects of tobacco, with complete arrest of the process once smoking has been abandoned. Vascular reconstructive operations frequently are not feasible, because the involvement of the small vessels of the extremity makes it difficult to locate suitable outflow sites for bypass grafts. Surgical sympathectomy has been used with some success. Digital amputations frequently are necessary, but major amputations usually are avoidable because the larger vessels are not involved.

Periarteritis Nodosa

Periarteritis nodosa is an inflammatory process involving the small and medium-sized arteries of all organs. Occasionally digital artery involvement produces Raynaud's symptoms and may progress to ulceration or gangrene. The disease affects males more frequently than females, in a ratio of 21. Individuals of any age may be affected, but most patients are middle-aged at the time of diagnosis. The pathologic process is that of inflammation progressing to occlusion or aneurysm formation. Renal and gastrointestinal complications occur in the early stages of the disease and include renal failure, intestinal perforation, and intraabdominal hemorrhage. Late mortality occurs as a result of cerebral and cardiovascular events. Steroids have increased the 5-year survival rate to over 50 percent and adjuvant cyclophosphamide therapy has been used with success in severe cases. Operation is reserved for the treatment of hemorrhagic or gangrenous complications.

Hypersensitivity Angiitis

Hypersensitivity angiitis encompasses a diverse group of disorders involving the smaller arteries, with basement membrane thickening, fragmentation of elastic fibers, and swelling of the collagenous structures, terminating in vascular occlusion. With the exception of scleroderma, all the causative mechanisms involve antigen exposure and the formation of antigen-antibody complexes that damage the small vessels. The antigens include hepatitis B virus, tumor antigens, and drugs. The primary clinical manifestation of hypersensitivity angiitis is digital artery occlusion with digital ischemia, Raynaud's symptoms, ulceration, and gangrene. Vasodilators such as calcium channel blockers and guanethidine have been helpful despite the lack of demonstrable vasospasm. The results of surgical sympathectomy have been discouraging.

Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is an autoimmune disease with antibodies directed against DNA and other cellular constituents. A prominent component of SLE is an arteritis involving the medium-size vessels of the skin, intestine, kidney, lungs, and heart. Larger arteries and veins are affected with a thrombotic process distinct from the arteritis, secondary to the presence of a circulating substance known as the “lupus anticoagulant.” This substance is not specific to SLE and is also found in other vasculitides. Although an elevated partial thromboplastin time is associated with the lupus anticoagulant, a hemorrhagic propensity does not develop. Thrombosis of the arteries of the upper and lower extremity, the carotid arteries, and the coronary arteries may occur, as may venous thrombosis involving the inferior vena cava, the upper and lower extremity veins, and the retinal veins. Steroid and cytotoxic therapy is directed against the arteritis; long-term anticoagulation therapy with warfarin is directed against the thrombotic diathesis.

Inherited Connective Tissue Disorders

Marfan's syndrome and Ehlers-Danlos syndrome are two relatively rare autosomal dominant inherited diseases of connective tissue. They are not vasculitides, as such. The major clinical features are those of aneurysm formation and dissection of the aorta and its major branches. These complications often produce exsanguinating hemorrhage and rapid death in affected patients in the second through the fifth decade of life.

Marfan's syndrome is characterized by a defect in collagen cross-linking. Mitral valve prolapse, ascending aortic aneurysm, and aortic dissection are the primary cardiovascular manifestations. Therapy is directed at lowering blood pressure with the use of beta blockers and electively resecting aortic aneurysms when the aortic diameter reaches 6 cm or more.

The Ehlers-Danlos syndrome comprises a group of at least ten disease entities identified by defects in the conversion of procollagen to collagen. Hyperelasticity of the skin, spontaneous rupture of large arteries, aortic aneurysm, and aortic dissection are the major clinical manifestations. Arteriography and bypass operations are associated with a substantial risk of hemorrhage, and ligation is the procedure of choice when rupture of a vessel is encountered in Ehlers-Danlos patients.

Antiphospholipid Syndrome

The antiphospholipid syndrome (APS) is a hypercoagulable state characterized by the clinical features of thrombosis, recurrent fetal loss, and thrombocytopenia occurring in association with antiphospholipid antibodies. The two methods currently in use to detect antiphospholipid antibodies are the lupus anticoagulant and anticardiolipin antibody tests, and either or both of these antibodies may be present in patients with APS. Other laboratory abnormalities frequently encountered in patients with APS include thrombocytopenia, elevation of the erythrocyte sedimentation rate, and prolongation of the partial thromboplastin time. The clinical features of APS may be seen in patients with antiphospholipid antibodies and no other concomitant disease process; this is known as primary APS. Alternatively, antiphospholipid antibodies may occur in association with systemic lupus erythematosus or other autoimmune, infectious, malignant, and inflammatory disorders; this is defined as secondary APS. Thrombosis of the venous system is more common than

thrombosis of the arterial system in patients with APS, but the latter is far more devastating. In addition, arterial thrombosis due to primary APS may be difficult to distinguish from atherosclerosis, since circulation to the extremities, brain, and myocardium is frequently affected in both disorders, and the resultant ischemic symptoms are identical. Certain clinical features should raise suspicion of the diagnosis, however. In comparison with the atherosclerotic patient population, patients with arterial manifestations of APS are more often female, are significantly younger, and more frequently have upper-extremity involvement. Bypass of arterial lesions in patients with APS is associated with a very high incidence of early graft thrombosis (75 to 80 percent), although there is some evidence that preoperative and postoperative treatment with steroids, cytotoxic agents, and warfarin may improve the results of vascular reconstruction.

FROSTBITE

Several forms of cold injury have been described, including acute pernio or chilblains, chronic pernio, trench foot, and frostbite. Acute and chronic pernio represent focal, relatively mild injury of the skin and subcutaneous tissues resulting from cold exposure of moderate severity. The lesions tend to heal rapidly and are seldom a problem for the surgeon. Trench foot is principally a military injury produced by prolonged exposure to a cold, damp environment. Trench foot often occurs with temperatures above freezing, but with prolonged immobility. Immersion foot is the seagoing counterpart of trench foot.

Frostbite is the cold injury most frequently encountered in civilian practice. Frostbite occurs with exposure of tissues to subfreezing temperatures for a period of several hours. Shorter-term exposure to subzero temperatures results in a different form of frostbite, commonly occurring in airplanes at high altitudes and characterized by the term “high-altitude” frostbite. It has been demonstrated experimentally that cold-induced injury to mammalian tissues begins when the tissue temperature reaches 10°C. At -5°C cells lose the ability to recover from the freezing process. Observations in the Korean War revealed that frostbite characteristically occurred with exposure to ambient temperatures of -7°C for 7 to 18 h.

Several factors influence the injurious effect of cold exposure, including humidity and wind. Clinical observations have revealed that the most severe cold injuries occur in patients with prolonged contact with moisture or metallic surfaces, both of which function as efficient heat conductors. Wind also accelerates heat loss, presumably through increased evaporation of sweat and disruption of the radiant heat around the body. Coexistent peripheral arterial occlusive disease may contribute to the rapid development of cold-induced tissue injury. Chronic arterial disease should be excluded as a contributing factor whenever severe frostbite is observed in an adult civilian patient.

The pathophysiologic features of frostbite depend on the degree of cold-induced injury. Initially vasoconstriction occurs on exposure to the cold. The histologic findings in mild frostbite consist of a low-grade vasculitis; the process progresses to an intense inflammatory reaction of the intima with severe frostbite. The capillary endothelium becomes permeable, and the resultant extravascular fluid accumulation produces soft-tissue edema. Thrombi form in the terminal arterioles and capillaries, and irreversible tissue necrosis develops. It is not known whether the fundamental cold injury occurs as a result of direct freezing with disruption of cell membranes or from ischemic necrosis secondary to widespread thrombosis of the arterioles and capillaries.

Frostbite may be classified into four degrees of severity, analogous to the classification of burn injury. First-degree injury consists of edema and redness without necrosis; blistering becomes evident in second-degree injury; necrosis of skin constitutes third-degree injury; gangrene develops in fourth-degree injury, necessitating amputation of the affected extremity. A simpler categorization of frostbite divides injuries into superficial and deep classifications. Superficial frostbite involves the skin and superficial subcutaneous tissue; deep frostbite involves the deeper subcutaneous tissue, muscle, and even bone.

Treatment

The treatment of frostbite begins with rapid warming of the injured tissue. The involved body part should be immersed in warm water, with a temperature in the range of 40 to 44°C. Complete rewarming generally requires about 20 min. Once warm, the injured extremity should be elevated to minimize the formation of edema, and antibiotics and antitetanus therapy are instituted. The extent of gangrene is difficult to assess early in the course of frostbite. The degree of irreversible injury is often much less than initially feared, because the skin may be involved to a much greater extent than the subcutaneous tissue. For this reason, amputation should be delayed several weeks, until the precise extent of gangrene can be accurately determined.

Experimental and clinical evidence has demonstrated a beneficial effect from early sympathectomy in the treatment of severe frostbite. Operative sympathectomy should be performed within the first few days of injury if the injury is severe enough to produce tissue necrosis. Sympathectomy is also useful in alleviating the late sequelae of cold injury, including hyperhidrosis, cold sensitivity, and pain. Treatment directed against intravascular thrombosis has theoretical advantages; however, no consistent benefit has been achieved with the use of agents such as heparin or dextran.

The late sequelae of frostbite were well documented in a study by Ervasti of 812 cases with follow-up over 5 to 18 years. Long-term cold sensitivity was present in 82 percent of the patients, color changes of the skin in 73 percent, hyperhidrosis in 59 percent, pain with use of the extremity in 39 percent, and sensory loss in 23 percent. Sympathectomy was associated with improvement in symptomatic residua of frostbite in over 80 percent of the patients in whom it was performed.

ARTERIOVENOUS FISTULAE

An arteriovenous fistula is a direct communication between the arterial and the venous circulation that bypasses the capillary bed. The congenital variety is present at birth and grows or regresses. Although fistulae rarely cause hemodynamic symptoms, they can produce severe local problems that are often refractory to therapy. Fortunately, most congenital fistulae do not require operative therapy. The acquired arteriovenous fistula is most often the result of penetrating trauma; it can produce serious cardiac dysfunction and usually requires repair.

Congenital Fistulae

Classification

Congenital arteriovenous fistulae may be circumscribed or diffuse and are a result of abnormal development of the primitive vascular system. Szilagyi and associates have proposed an embryologic classification that divides these malformations into the hemangiomas and the fistulae. Circumscribed fistulae, called cavernous or simple hemangiomas, make up 19 percent of these lesions. There are two forms of this entity: a nonneoplastic lesion that appears at birth and grows with the child, and a neoplastic lesion that begins as a small lesion just after birth, grows rapidly, and then usually spontaneously involutes. The distinction is important because of the tendency of the latter form to involute.

The diffuse group makes up 81 percent of these lesions and is characterized by anomalous micro or macro arteriovenous communications. These lesions are complex and have hemangiomatous, fistulous, and aneurysmal elements. Although present at birth, they may not become apparent until the second or third decade of life.

Clinical Manifestations

Many of these lesions are disturbing solely because of their cosmetic appearance, which ranges from innocent-looking varicose veins to an ulcerated, bleeding, pulsatile mass. The cutaneous changes may be only a small part of the problem, however. CT and MRI provide the most complete diagnostic information about the extent of the malformation and the need for resection. Arteriography and venography have many limitations as diagnostic procedures because they do not define the extent of muscular and bony involvement and do not allow visualization of the fistulous connections. The principal role of arteriography is in preoperative localization of the afferent arterial supply, but it also may be used to embolize the localized lesion prior to resection.

High-output cardiac failure is rare with the congenital arteriovenous malformation. The congenital malformations cause symptoms ranging from mass effect and thrombosis to bleeding from varicosities. Ulceration of the overlying skin with pain and infection is not uncommon, because of the accompanying venous hypertension. Progressive growth is the rule, because the low resistance on the venous side produces high flows and enlargement of the arterial inflow vessels. Malformations with extensive hemangiomatous changes can be associated with thrombocytopenia and purpura, the Kasabach-Merritt syndrome. One form of congenital fistula is the Klippel-Trenaunay syndrome, characterized by cutaneous hemangiomas with port-wine staining, varicose veins, and hypertrophy of the involved extremity. Patients with this syndrome may not have a deep venous system, and it is an error to excise the abnormal superficial veins without evidence of an intact deep system.

Treatment

The hemangiomas are more circumscribed and can be completely removed when local symptoms dictate. If changes in consistency or appearance are observed, selected biopsies are indicated because the potential for malignant growth exists. Attempts to control the fistulous malformations are more difficult, and surgery should not be considered unless the lesion is life- or limb-threatening. Simple elastic support may suffice in an extremity lesion with low flow.

Intraarterial embolization and staged operative procedures for symptomatic lesions have been largely unsuccessful except as a means of short-term palliation (Fig. 20-41). When operative intervention is required as a last resort, a multidisciplinary team of vascular, orthopaedic, and plastic surgeons offers a better chance for cure. The major vessels entering and leaving the tumor are initially ligated, and the mass is resected with any involved tissue, including muscle, bone, subcutaneous tissue, and skin. Reconstruction is then accomplished by the use of appropriate tissue transfers. Lesions in the chest, abdomen, and pelvis may require circulatory support and deep hypothermia.

Acquired Fistulae

Etiology and Diagnosis

Acquired fistulae are usually a result of penetrating or iatrogenic trauma but can occur spontaneously, such as an aortocaval fistula. Although it was once thought that the fistula would continue to enlarge, it now seems that spontaneous closure occasionally occurs. On examination, there usually are visible veins surrounding the fistula that may have a palpable thrill and a machinery-type murmur throughout the cardiac cycle. The heart rate may slow with compression of the fistula (Branham's sign). The Duplex scanner allows visualization of the increased flow velocities with the lowered resistance, the dilated outflow vessels, and the often-associated false aneurysm. Arteriography is the essential diagnostic tool for identifying the site of the communication and planning the operative approach.

Pathophysiology

When an arteriovenous connection is suddenly created, there is a decrease in blood flow distal to the lesion and an increase in venous pressure (Fig. 20-42). The peripheral vascular resistance is lowered to that of the venous system, and an increase in cardiac output occurs. As the fistula matures, the collateral circulation increases, the distal perfusion approaches normal, and each component dilates. Over the longer term, venous hypertension may develop in the extremity. Growing children may exhibit limb length discrepancies, and cardiac failure may occur.

Treatment

Acquired arteriovenous fistulae should be repaired, because of the risk of life- or limb-threatening complications. Repair is technically easier when done at an early stage of fistula

development. Successful repair requires that all four limbs of the fistula be controlled. An angiographically placed balloon occluding device in the afferent arterial limb can be a helpful adjunct. Once control is obtained, the fistula itself should be exposed and directly divided and sutured. The connection itself may be quite small and can be repaired with several sutures. Larger defects require a patch angioplasty or interposition graft. The long-term prognosis after direct repair is excellent. Early attempts at control by proximal artery ligation were both unsuccessful and dangerous. Ligation of all four vessels without direct exposure of the communication has resulted in a significant incidence of recurrences and distal arterial ischemia. Inaccessible fistulae can be managed angiographically with a variety of coils or plugs of gelatin foam. Recurrences are high with this technique because it requires great skill on the part of the angiographer.

VASCULAR TRAUMA

Major advances have been made in the treatment of arterial injuries on the basis of lessons learned from combat injuries and from the management of atherosclerotic disease. The amputation rate after extremity vascular trauma was 50 percent during World War II and 13 percent during the Korean War; it is now 2 percent for civilian injuries.

General Considerations

Once the airway has been stabilized and volume replacement begun, patients with major vascular injuries are triaged into one of three groups: (1) those whose injuries are life-threatening and require immediate operation; (2) those whose vascular injuries are obvious but whose vital signs permit an arteriogram; and (3) those whose injuries, while not obvious, require evaluation because of their proximity to vascular structures. The signs of arterial injury include a pulsatile or expanding hematoma, pulsatile bleeding, a bruit or thrill, or end-organ ischemia. Suggestive features include unexplained shock, likely location, a stable hematoma, an injury to an adjacent nerve, or a questionable history of arterial bleeding (Table 20-8). Palpable pulses may be present even when there is a significant arterial injury. Arteriography is the gold-standard test for the diagnosis of an arterial injury and should be performed whenever possible.

A major priority of treatment is the control of hemorrhage. Direct digital pressure is the most effective maneuver at the scene of the trauma or in the emergency department. Tourniquets should be used only as a last resort, because they can occlude collateral flow and increase tissue damage. Attempts at blind clamping with hemostats should be avoided, because this may further damage the injured vessel. Neither an embedded weapon nor a hematoma should be removed until proximal control is achieved, because brisk bleeding may ensue.

Since many of these patients are operated on before complete data are available, maximum flexibility is necessary in planning the operative approach. Both legs should be prepared for surgery if there is any chance that a saphenous vein graft might be necessary. The chest should be prepared if there is any chance that the arch vessels have been injured or that control of intraabdominal bleeding cannot be achieved transabdominally. Preparations should be made to administer blood with a rapid-infusion device.

Proximal and distal arterial control should be obtained before the injury is approached. If an intraluminal thrombus is present, it generally can be removed with a balloon catheter. Systemic heparin is not given to patients with multisystem trauma, and the thrombectomized artery can be flushed with heparinized saline. Once the injury is identified, wide debridement of the vessel is performed. This is especially important in high-velocity missile wounds, in which the damage may extend beyond the obvious injury. The debridement should not be compromised in an attempt to simplify the arterial repair.

The type of arterial repair depends on the nature and extent of the injury. Although some sharp, penetrating wounds can be repaired with lateral arteriorrhaphy or a venous patch angioplasty, most arterial wounds require segmental resection. When an end-to-end anastomosis cannot be performed, an interposition graft using size-matched autogenous vein is used. Contralateral saphenous vein is recommended for extremity repairs to preserve the superficial venous drainage in the affected limb. Use of prosthetic grafts is controversial. Rich and associates reported complications when prosthetic material was used in 20 of 26 patients in the Vietnam Vascular Registry, while a conflicting civilian experience was reported by Mattox and associates from Houston. In their experience 10 percent of vascular repairs became infected when autogenous vein or polyester was used, and no infections were noted using *ePTFE*. Completion arteriography is recommended; distal pulses cannot always be felt even when the repair is satisfactory because of arterial spasm.

Extracranial Vascular Trauma

Carotid Artery

Penetrating trauma as well as blunt trauma can injure the carotid artery. In the former, by far more common and easier to diagnose, the artery is partially or completely transected. Blunt injuries are the result of hyperextension or a direct blow and are usually diagnosed when neurologic signs appear, sometimes hours or days after the insult (Fig. 20-43). This lesion is characterized by intimal or medial tears that lead to dissection, stenosis, or thrombosis of the artery. Carotid arteriography is necessary in any hemodynamically stable patient with a suspected blunt injury.

Patients with penetrating carotid artery injuries may present with shock or respiratory distress from an expanding cervical hematoma. Pressure may affect cranial nerves IX, X, XI, or XII. Penetrating carotid injuries are anatomically divided into three zones. Zone I is the area below the top of the sternal notch; zone II extends from the sternal notch to the angle of the mandible; zone III is the region above the angle of the mandible. Arteriography is most helpful in zone I injuries, for planning the proper incision, and in zone III injuries, where the specific site and extent of the injury may influence operative strategy.

Preparation must be made for either a left anterior thoracotomy or a median sternotomy to gain control of the proximal carotid artery, if necessary. An oblique incision anterior to the sternocleidomastoid muscle provides exposure of the artery and can be extended caudally into a median sternotomy or craniad behind the ear. Most penetrating carotid injuries can be managed by lateral suture or end-to-end anastomosis. If an interposition graft is needed, either the external carotid artery or the saphenous vein can be used. External carotid artery injuries can be treated with proximal and distal ligation.

There are two areas of controversy in the management of carotid artery injuries. The first concerns the proper management of the asymptomatic penetrating neck wound. Recommendations vary from obligatory exploration of any wound penetrating the platysma muscle to arteriography and observation. It is reasonable just to observe these patients if they remain asymptomatic and the arteriogram is of sufficient quality to rule out a carotid artery injury. The second controversy concerns the implications of a neurologic deficit. Although it is accepted that carotid artery injuries should be repaired in a neurologically intact patient, opinions vary on the indications for operation when a deficit exists. The best evidence suggests repair for patients with mild deficits and those with a severe deficit who have prograde flow preoperatively. Ligation is recommended for patients with a severe deficit and absence of prograde flow and patients with a distal thrombus that cannot be extracted. Morbidity and mortality rates are lower for all groups of patients, except those in coma, when repair rather than ligation is performed.

Vertebral Arteries

These injuries are usually from penetrating trauma (95 percent). The increased use of arteriography has increased awareness of the problem. Vertebral artery injury rarely causes neurologic symptoms because of the dual blood supply to the basilar artery, but massive hemorrhage can occur. In most angiographically demonstrated occlusions there is complete transection (52 to 81 percent), which, if left alone, develops into arteriovenous fistula or false aneurysm (5 to 30 percent).

Although there are some proponents of radiologically guided occlusion of the injured vertebral artery with detachable balloons, this technique does not control the distal vessel and carries the risk of a guide wire being pushed through a false lumen within the bony foramina of the transverse cervical processes. Operative therapy, which consists of proximal and distal ligation of the injured vessel, remains the procedure of choice.

If the proximal artery appears occluded on the arteriogram and there is no pulsatile hematoma present, the proximal artery can be approached at its origin from the subclavian artery through an oblique skin incision in the neck, with retraction of the carotid sheath medially and the anterior scalene muscle laterally. The distal ligation may require control of the interosseous portion of the vessel. This can be achieved by extending the skin incision to the mastoid process, exposing the prevertebral fascia, and gently removing the bone that forms the anterior border of the interosseous canal. If there is a pulsatile hematoma caused by an injury to the proximal vertebral artery, control should be obtained in the chest. Occasionally, repair of the injured vessel is indicated when a proximal transection can be easily transposed onto the common carotid artery or when the contralateral vertebral artery is diseased or absent.

Proximal Brachiocephalic Injuries

Injuries to the innominate, subclavian, and common carotid arteries require immediate diagnosis and control of hemorrhage. The initial presentation may range from an asymptomatic patient with an innocuous-appearing wound to a patient in hypovolemic shock with a massive hemothorax. The wounds should not be explored digitally. Tube thoracostomy is necessary for pneumothorax or hemothorax. Nasogastric tubes should not be placed. Patients in shock require an immediate operation. Arteriography can be performed in a stable patient with a penetrating wound, but a negative result does not replace exploration, because it does not rule out a major injury (22 percent), and sudden hemodynamic collapse is possible (5 percent). By contrast, arteriography is mandatory in the evaluation of blunt trauma to this region.

In operative treatment, the patient is positioned supine with both arms abducted, and the neck, chest, and shoulders are prepared for surgery. In hemodynamically unstable patients, lesions on the right side can be controlled with a median sternotomy, and lesions on the left side with an anterolateral thoracotomy. The oblique cervical incision sparing the mediastinum provides exposure for most cervical injuries, but extension to remove the medial portion of the clavicle is sometimes necessary. The chest wall musculoskeletal flap, the so-called trap door incision, can be used for proximal left subclavian or common carotid artery injuries. Standard principles of arterial repair follow control of the bleeding. Wound debridement and lateral closure are often possible in these large vessels. Occasionally an autogenous patch is required. Ligation and remote prosthetic bypass is reserved for injuries with contamination.

Extremity Injuries

Associated injuries to bones, muscles, veins, and nerves complicate the management of extremity arterial trauma. Signs of arterial involvement may be masked by these associated injuries, and therefore arteriography becomes the most important diagnostic study for establishing the diagnosis and defining the extent of the lesion.

Upper-Extremity Injuries

Subclavian artery injuries are the most difficult to manage. They often are associated with injuries to the brachial plexus—either direct injuries, or indirect injuries from the pressure of an accompanying hematoma. Proximal exposure may require a thoracotomy for left-sided injuries and a median sternotomy for injuries on the right. Results of arterial repair are good, but long-term disability usually follows the nerve injury.

Injuries to the axillary and brachial arteries are handled by incisions directly over the penetrating wound. End-to-end anastomoses are usually possible. The most common upper-extremity vascular injuries are iatrogenic injuries in coronary arteriography and transluminal angioplasty and injuries obtained in intravenous drug abuse. Early repair of these injuries is indicated. Although only one-third of the patients with brachial artery occlusion will develop arm symptoms, given the rich collateral network around the elbow, delayed

repair will necessitate a more complicated operation.

Lower-Extremity Injuries

Most penetrating injuries of the lower extremity involve the superficial femoral artery; achieving proximal and distal control is similar to performing an elective femoral-popliteal artery bypass. Arteriography should be performed when the diagnosis is in question. Wounds entering the lateral aspect of the thigh and exiting posteriorly do not need further study unless suggestive signs of arterial injury are present. Medial-entry wounds or those in proximity to a major artery should undergo arteriography.

Limb loss rates from arterial injuries associated with fractures and other soft-tissue trauma range as high as 44 percent in civilian practice and higher in military series. Many of the patients with salvaged limbs have decreased function from muscle loss or nerve injury. Long-bone fractures may cause arterial damage by acute angulation, laceration, or longitudinal stretching and subsequent thrombosis (Fig. 20-44). The arteries most often involved in the lower extremity are the distal superficial femoral and popliteal arteries and the proximal tibial arteries around the knee; in the upper extremity, it is the brachial artery just above and below the antecubital fossa. These lesions may cause immediate or delayed arterial occlusion, and arteriography is indicated for high-risk fractures in these regions when any evidence of arterial compromise is present.

Fracture stabilization in the sequence of repair of concomitant bony and arterial injuries is controversial. Initial bone stabilization may correct perfusion abnormalities and protect any subsequent vascular repair from damage due to movement, but it delays restoration of flow unless a temporary shunt is placed, which may impede access to the injured vessel. On the other hand, initial stabilization of the fracture might result in damage to a vascular repair that was performed first. In general it is preferable to stabilize fractures before the vascular repair, but this sequence must be determined in each case according to the degree of distal ischemia and the presence of hemorrhage. Important technical considerations in repair of combined injuries include: (1) early fasciotomy before orthopaedic repair; (2) preferential use of the contralateral saphenous vein; (3) repair of major venous injuries; (4) completion arteriography; and (5) adequate soft-tissue coverage using appropriate muscle flaps.

The liberal use of fasciotomy in the injured extremity is important in reducing postoperative edema and the compartmental hypertension that may lead to myonecrosis and nerve damage. Reperfusion injury of ischemic skeletal muscle is related to damage from oxygen-derived free radicals. These compounds increase vascular permeability, which permits the loss of intravascular fluid from the capillaries into the interstitial space. Mannitol, a free-radical scavenger, therefore should be given before limb revascularization. Objective measurements of compartment pressures can be made by determining the pressure within a closed system that is required to overcome the pressure in the compartment. Pressures higher than 40 mmHg mandate immediate decompression if clinical signs of an elevated compartmental pressure is present. These include sensory loss in the first metatarsal space, inability to dorsiflex the great toe, or foot drop. Patients in whom the pressure is greater than 50 mmHg should have immediate decompression regardless of the clinical findings. Fasciotomy should be performed through limited skin incisions if possible, because the fascia is the limiting tissue.

Progress in microvascular tissue transfers and improvements in skeletal fixation in recent years have increased the number of salvaged extremities. Occasionally early amputation is required because of the magnitude of the injury. Avulsed nerves and extensive soft-tissue and bony destruction may make the limb useless even after successful arterial repair. The term “crush syndrome” was applied to the ischemia-induced syndrome of myonecrosis, myoglobinuria, and renal failure during the London Blitz in World War II. Complications in the postoperative period, such as metabolic abnormalities from injured skeletal muscle, sepsis, or failure of the arterial repair, also may lead to amputation. The principle of life over limb must always be followed when treating these complicated injuries.

Abdominal Vascular Injuries

Most abdominal vascular injuries follow penetrating trauma. Approximately 20 percent of patients with abdominal missile wounds have a major vascular injury. The aorta, the inferior vena cava, and the iliac vessels are most often involved, and the presenting clinical problem is usually exsanguinating hemorrhage. Diagnostic arteriography, although helpful, is rarely indicated. Blunt abdominal trauma is less common and usually involves avulsion of the hepatic veins and visceral arteries. Most of these injuries are associated with trauma to other organ systems, so the surgeon is faced with two problems: control of bleeding, and management of solid organ or hollow viscus wounds that may be contaminated.

Principles of Management

Hemostasis is an essential part of the resuscitative effort, so operation must not be delayed for a detailed diagnostic evaluation. The beneficial tamponading effect of the intact abdominal cavity should not be forgotten. In cases of severe hypotension aortic control can be achieved with the use of a left anterolateral thoracotomy through the sixth or seventh interspace. An autotransfusion device should be available, but the salvaged blood should not be reinfused until an intestinal injury has been ruled out. The chest, abdomen, and groins are prepared for surgery. A midline abdominal incision should be made for rapid and generous exposure. The aorta is clamped at the diaphragm if arterial bleeding is profuse and cannot be controlled with packing. Blood and clot are rapidly removed, the abdomen packed with large pads, and the blood volume restored. Any intraperitoneal soilage is identified and controlled and then specific injuries are sought out.

Aortic injuries have mortality rates from 50 to 90 percent, depending in part on the location of the injury. They are associated with a midline retroperitoneal hematoma and must be approached initially with aortic control at the diaphragm. The hematoma can then be opened and the clamp moved distally. Inferior vena cava and renal artery lesions require mobilization of the right or left colon. Lesions of the suprarenal aorta can be isolated with the use of the medial visceral rotation technique. Occasionally the incision must be extended into the chest for proximal aortic control in the face of uncontrolled hemorrhage. Most patients with aortic injuries who reach the hospital can have the injury repaired with the lateral suture technique. Occasionally a graft is required, and the decision to use one depends on the degree of contamination present. Prosthetic grafts have been used successfully in situations with minimal contamination. Ligation and extraanatomic bypass is preferred, however, in cases where contamination is severe. Iliac artery injuries usually require graft replacement, and the same rules about contamination apply. A femoral-to-femoral artery bypass usually serves as a satisfactory replacement for the injured iliac artery. The celiac axis and its branches can be ligated in most patients without adverse consequences. Injuries to the proximal superior mesenteric artery or vein should be repaired by lateral suture or by saphenous vein bypass. More distal lesions are ligated, and bowel resection may be indicated. These patients require a second-look operation.

Pediatric Injuries

Pediatric arterial injury is uncommon, and standardized approaches are lacking, especially for the group of iatrogenic injuries of infancy that make up virtually all of the injuries in children under the age of 2 years. These injuries present special challenges because of the small vessel size and the different implications of non-limb-threatening ischemia. The manifestations of acute arterial ischemia in a child are similar to those in an adult. Chronic ischemia causes impaired growth in the affected limb rather than claudication or tissue loss. In general pediatric arterial lesions should be repaired promptly. The only exception may be an iatrogenic femoral artery occlusion in an infant, in which case heparin therapy and observation of an extremity whose viability is not in question may yield better long-term results than early operation.

AMPUTATIONS

The indications for amputation of part or all of an extremity include: (1) loss of viability secondary to nonreconstructible ischemia; (2) trauma that is sufficiently extensive to preclude repair; (3) tumor of the bone, soft tissue, muscles, blood vessels, or nerves; and (4) extensive infection that does not respond to conservative measures or contributes to septicemia. More than two-thirds of the amputations in civilians in Western societies are performed for peripheral vascular disease, usually as the result of extensive lower-extremity atherosclerotic occlusion.

The most commonly used procedures are *above-knee (AK)* and *below-knee (BK)* amputations, though a variety of eponyms have been applied to specific levels of amputations and refinements of technique. Transmetatarsal and digital amputations are less frequently indicated in patients with peripheral vascular disease. The Syme amputation, knee disarticulation, hip disarticulation, and hemipelvectomy are usually reserved for malignant and traumatic lesions. At each level, the operative procedure and postoperative management are directed at achieving primary healing, a painless stump that will withstand the pressures of a prosthesis, and relatively unrestricted ambulation.

Selection of Level

Several factors contribute to the decision concerning level of amputation (Fig. 20-45). When amputation is performed for malignant disease, the principal factor is wide excision of grossly apparent tumor. In the case of amputation subsequent to trauma or for peripheral vascular disease, the major factor determining the optimal level is usually the extent of healthy tissue. Other factors include the length of the stump sufficient for function of a prosthesis and the cosmetic effect of a prosthesis.

The longer the amputation stump, the more functional will be the limb, and the better control the patient will have of the prosthesis. For traumatic amputations with irregular damage to the skin, grafting procedures and mobilization of flaps are indicated to maintain as much of the bony length as possible. AK amputees expend about twice as much energy to ambulate as BK amputees.

There are exceptions to the general rule of “the longer, the better.” For the lower extremity, proximal to the Syme amputation, the best level below the knee as far as prosthetic fitting is concerned is at the distal musculotendinous junction of the gastrocnemius muscle. Every effort should be made to preserve the knee joint. A below-knee prosthesis can be fitted to a stump as short as 5 cm, improving the chances of successful ambulation. In the pediatric population, disarticulation is frequently the procedure of choice in order to save the growing epiphysis and minimize the problem of distal bone overgrowth. In the upper extremity, as much length as possible should always be saved, even if there is only a very short below-elbow or humeral stump.

Other factors that contribute to the selection of amputation site are the patient's general condition and, particularly, the feasibility of rehabilitation. AK amputations are indicated for bedridden patients who require nursing care, since these procedures are associated with a higher incidence of primary healing. A BK amputation in a nonambulatory patient will result in a tight flexure contracture and therefore should not be performed. If there is established contracture, it is always preferable to amputate the leg above the level of contracture.

A correlation has been reported between the healing of digital and transmetatarsal amputations and the presence of peripheral pulses. In patients with a palpable popliteal pulse healing at this level occurred twice as effectively as in those in whom only a femoral pulse was demonstrated. The presence of a popliteal pulse has been associated with a 97 percent success rate for BK amputations, as compared to 82 percent of patients with absent popliteal pulses. In another study, 79 percent of BK amputation failed when the femoral pulse was absent, as compared to 29 percent wound breakdown when the femoral pulse was palpable. When the popliteal pulse was present, the failure rate was 10 percent.

The value of Doppler systolic blood pressure and arterial waveform analysis using the segmental plethysmograph in the selection of BK and forefoot amputation sites has been assessed. A calf systolic pressure higher than 70 mmHg was associated with a 97 percent success rate for BK amputations. However, because of a high false-negative rate, a patient should not be denied a BK amputation on the basis of this determination. The ankle systolic pressure has not proved valuable in predicting the healing of forefoot amputations. An assessment of Doppler ultrasonography and digital plethysmography concluded that they are imprecise indices of the healing of amputations for arterial insufficiency.

At all levels, healing ability is a function of nutritional skin blood flow. The status of the limb circulation may be evaluated by assessing the skin circulation. The clearance of radiolabeled xenon, measured at the anterior incision line, has been found to be well correlated with the healing of amputations. Quantification of skin fluorescein delivery by fiberoptic fluorometry also has predicted the healing of an amputation site. Measurement of transcutaneous P_{O_2} has been used to determine amputation levels; high saturation was indicative of good healing potential, but no level could be defined below which healing would not occur. Thermography tended to overestimate the degree of ischemia and to indicate a higher level of amputation than necessary. Despite the availability of these objective tests, most amputation levels are chosen on the basis of clinical findings alone. When in doubt, one should err conservatively, explaining the possibility of subsequent revision to the patient and family.

The two general classes of amputations performed for lower-limb ischemia are the standard or conventional amputation and the provisional or open (guillotine) amputation. The conventional amputation is performed by constructing curved skin and fascial flaps that have their base at the level of amputation. The muscles, major blood vessels, and bone are divided at the level of amputation, and the major nerves are put on slight stretch and transected so that they retract proximally. The muscles may be tapered so that not too much soft tissue remains over the bone end, and the distal inch of periosteum may be removed to avoid leaving a detached segment that can cause bone spur formation. After the amputation has been completed, the deep and superficial fascia are approximated over the bone, and the skin is loosely closed. The wound may be closed with or without drainage with a catheter connected to a closed suction apparatus. Guillotine amputations should be regarded as provisional procedures performed in the case of sepsis, when the infection precludes primary closure of the wound. They are of historical interest as the most commonly employed procedure before the age of modern surgery. The procedure is performed well below the level planned for a subsequent, definitive amputation. A circumferential incision is made and carried down to the bones. The bones are transected and the wound is packed open. Dressings are changed frequently and a definitive amputation is performed when the septic process has resolved.

Postoperatively, a lightly compressive dressing is applied to the stump, usually in the form of a “stump sock.” The wound is not examined for four or five days, unless signs of infection develop. Contractures are prevented by stump exercises and stretching the proximal joint in the hope of achieving full motion of the remaining joints of the amputated extremity. Rehabilitation requires the services of an experienced physiotherapist, an occupational therapist, and a prosthetist. *Immediate postoperative prostheses (IPOP)* are used increasingly in the management of extremity amputees. In this situation, a rigid dressing is applied at the operating table. The stump is covered with a sterilized sock and a plaster cast, and a prosthetic unit that includes socket attachment flaps may be applied immediately after the rigid dressing has dried. The patient is then encouraged to walk between parallel bars 24 h after the operation, and weight bearing is increased daily. The rigid dressing is kept in place for approximately 2 weeks, at which time the sutures are removed and a new socket is applied without delay. Ten days later the second socket is removed, and at that time stump measurements and a cast are made for a permanent prosthesis. Another temporary socket is provided until the permanent prosthesis is available. With both an IPOP and conventional management, shrinkage of the stump occurs in time and may necessitate a new socket for use with a permanent prosthesis.

Morbidity and Mortality

The operative mortality for amputations performed for isolated tumor, trauma, or infection is less than 3 percent. The mortality rate increases by an order of magnitude in patients who exhibit signs of arteriosclerotic disease. Diabetes does not significantly affect the mortality rate. The mortality rates are generally higher for more proximal amputations in the lower extremity. The most frequent causes of morbidity include cardiac (52 percent) and pulmonary (26 percent) complications. Phantom limb pain is a problem in many patients, with incapacitating pain in the distribution of the amputated extremity.

In a series of 130 patients with unilateral amputation, 70 (54 percent) received prostheses, 84 percent of whom subsequently had successful rehabilitation. Of those who received unilateral amputations, only 30 percent of the *AK* group had successful rehabilitation, whereas 66 percent of the *BK* group were able to walk on prostheses. In the patients with bilateral amputations, 30 percent had successful rehabilitation, which required that at least one of the amputation sites be at the BK level. For the *IPOP* group, the time was shorter from amputation to the fitting of a permanent prosthesis, and rehabilitation rate was higher.

Lower-Extremity Amputations

Toe Amputation

Amputation of one or more toes may be performed for gangrene or osteomyelitis. A transphalangeal level may be used if necrosis is distal to the *proximal interphalangeal (PIP)* joint and if there is absence of cellulitis, necrosis, and edema in the skin to be used for the flaps. If a more proximal amputation is required, transmetatarsal resection with removal of the metatarsal head is preferable to disarticulation (Fig. 20-46). The latter is associated with an increased incidence of breakdown, since the stump is bulky, and in the face of infection or poor vascular supply the exposed cartilage resists infection less effectively than does bone. The incidence of healing at both the transphalangeal and transmetatarsal levels has been relatively poor, with reports ranging from 40 to 60 percent. A direct relationship has been demonstrated between healing at these levels and the presence of popliteal and pedal pulses, suggesting that the poor results were related to poor selection of amputation sites.

Transmetatarsal Amputation

The indications for this procedure include necrosis proximal to the *PIP* joint, but distal to the level of the transmetatarsal incision. Necrosis in the interdigital creases constitutes a relatively frequent indication. In one series, primary healing was achieved in 54 percent of patients with absent pedal pulses. The healing of transmetatarsal amputations in patients with peripheral vascular disease is correlated with the potential for revascularization. Healing occurred in only 24 percent of patients with nonreconstructible disease and over 80 percent of patients with reconstructible disease. Prior lumbar sympathectomy did not alter the rate of healing. If infection or gangrene compromises the standard plantar skin flap, open or guillotine amputation at the midtarsal level can achieve a high level of healing and a durable stump. Wound closure can occur by contracture or partial-thickness skin graft.

Skin incisions and subcutaneous flaps are established so that the plantar flap is long, with a very short dorsal flap cut directly over the line of bone transection (Fig. 20-47). This configuration permits placement of the scar on the dorsal, non-weight-bearing surface of the foot. The metatarsals are divided just proximal to the level of the dorsal incision are removed, and the skin and fascia are closed in layers.

Syme Amputation

Terminal arterial disease involving the distal part of the foot constitutes one of the few peripheral vascular disorders for which the Syme amputation is indicated. In one study the Syme amputation following preliminary ray or transmetatarsal amputation for infection in diabetic patients with peripheral vascular disease was successful in about 85 percent of cases when the posterior tibial pulse was palpable. Functionally, this operation represents an acceptable level for amputation, because it maintains the length of the extremity and preserves the heel skin, which provides an excellent weight-bearing stump (Fig. 20-48). The prosthesis, however, is more difficult to fit.

Below-Knee (BK) Amputation

The below-knee level is probably the most common initial choice for patients with a nonviable foot. Major functional advantages of the BK amputation over the *AK* level are the ability to provide a more functional prosthesis for complete rehabilitation, the ability of the patient to move more easily in and out of bed, and a reduction in energy expenditure in ambulation and in the incidence of phantom limb pain. The stumps of BK amputations can be lengthened by distraction osteogenesis to improve prosthesis fit and increase ambulation. The BK amputation constitutes the level of choice for ischemic lesions of the foot that do not extend above the malleoli, but generally it is not applicable if the gangrene extends in continuity above the ankle. Ischemic rigor of the calf muscles is another contraindication. Contracture of the knee or hip is a relative contraindication, because it negates the functional advantages of this level.

The choice between the BK and AK amputations is determined by the patient's general health and potential for rehabilitation and an evaluation of the vascular status at the proposed level of amputation. No essential difference in healing rate is noted in patients with or without palpable popliteal pulses, but the absence of a femoral pulse is associated with a high failure rate for BK amputation. Healing is not influenced by the presence of diabetes. Healing of BK amputations has been reported in 68 to 90 percent of cases in large series, compared with healing rates of 82 to 98 percent for AK amputations. By contrast, a higher percentage of BK amputees regain ambulatory abilities with a prosthesis. Rehabilitation rates for patients with ischemia who had BK amputations range from 50 to 90 percent.

The amputation should be performed proximal to the lower third of the tibia, since the preponderance of tendinous structures distal to this area predisposes to poor circulation and an unstable, painful stump (Fig. 20-49). A long posterior flap technique is normally used. The skin and soft tissue are cut perpendicular to the axis of the leg over the anterior 180 degrees of the calf circumference. The incision is extended distally to create a long posterior flap of suitable length to close the defect. The posterior incision is beveled to reduce the bulk of muscle in order to provide a better stump. The soleus muscle should be completely excised, because no contribution to the blood supply of the skin passes through it. Equally short anterior and posterior flaps may be made if there is compromised circulation or if recent distal scars preclude the usual incisions. The tibia is usually transected 5 cm above the distal level of the skin flap, and the fibula is transected slightly above the tibial level. The tibia may be beveled approximately 45 degrees to prevent a sharp edge of bone from eroding into the thin anterior skin flap. The nerves are pulled gently down, transected, and then allowed to retract, and the vessels are ligated above the level of the end of the tibia. The fascia and skin are closed with interrupted absorbable sutures. If desired, closed suction drainage may be used, removing the drain on the second postoperative day.

An *IPOP* may be used, in which case the patient is encouraged to stand and bear weight on the first postoperative day, with activity rapidly increased thereafter. A permanent prosthesis usually can be applied at the time of the second cast change, i.e., at 3 weeks, when the stump is sufficiently mature (Fig. 20-50). In a series of cases, the use of IPOPs at this level has shown promise in reducing the incidence of phantom limb symptoms.

Above-Knee (AK) Amputation

Indications for an AK amputation include extension of skin gangrene above a level that would permit *BK* amputation and rigor of the calf muscles. Peripheral gangrene coupled with contracture of the knee or hip constitutes a relative indication, since there is no advantage to a BK amputation.

The AK amputation has the highest healing rates, 90 percent or more, in patients with peripheral vascular disease. The rate of reamputation is lower, which is an important consideration in elderly and debilitated patients. The higher mortality rate of AK amputations is related to the fact that they are more frequently performed in elderly and debilitated patients.

Anterior and posterior skin flaps are fashioned of equal length in a “fish- mouth” configuration (Fig. 20-51). The vascular bundles are divided individually at the level of proposed transection of the femur. Gentle traction is applied to the sciatic nerve before it is transected so that it retracts proximally to the stump of the femur. The periosteum of the femur is scraped with a periosteal elevator, and the femur is transected at a point well proximal to the soft-tissue resection. The edge of the bone is beveled and rasped to provide a smooth radius. Rehabilitation toward ambulation usually requires more time than with BK amputations.

Hip Disarticulation

The indications for hip disarticulation usually include extensive lower- extremity ischemia with proximal gangrenous change, tumors, and extensive traumatic injuries. Flaps are fashioned so that they will come together anteriorly, and the posterior flap is constructed such that it forms a unit of skin, fascia, and muscle to swing forward so that the patient may

sit comfortably in the socket of the prosthesis. The femoral vessels are ligated anteriorly, and, after the posterior dissection is completed, the femur is disarticulated. The sciatic nerve is cut high and tied to prevent bleeding from vasa nervorum. The undersurface of the gluteal muscle is tapered to facilitate approximation of the tissue to the lower abdominal wall. When hip disarticulation is performed in patients with peripheral vascular disease, the results are dismal. These patients typically have had a succession of failed lower amputations, often complicated by peripheral bypass graft sepsis.

Upper-Extremity Amputations

Upper-extremity amputations usually are performed for trauma. Malignant disease is the second most common indication. Peripheral vascular disease, particularly thromboangiitis obliterans, is a rare indication. In all circumstances, the treatment is directed at conserving as much viable tissue as possible and, with more distal amputations, maintaining the function of the hand as a grasping organ (Fig. 20-52). The latter requires preservation of intrinsic muscle-tendon systems of the hand and the longer and more powerful flexors and extensors originating in the arm. The technique used for amputations in the upper extremity is directed at producing minimal scar tissue, and the procedure is planned so that the scar is in an area as far removed as possible from bones, tendons, nerves, and points of external contact. Above the level of the metacarpals, opposing tendons should be fixed to one another in order to preserve muscle length and tone.

Digital Amputation

When part or all of the distal phalanx requires amputation, an attempt should be made to create a longer volar flap and a shorter dorsal flap, so that the scar may be positioned dorsad, away from pressure. Neither bone nor viable soft tissue should be sacrificed to produce an ideal flap. If the amputation removes less than one-half of the nail, it is preferable to retain the nail bed, but if more than one-half of the nail is removed, the entire root should be excised. When it is necessary to disarticulate at the distal interphalangeal joint, the cartilage should be removed from the head of the middle phalanx, and the flexor digitorum profundus tendon should be withdrawn, transected, and allowed to retract.

Amputations of fingers distal to the *PIP* joint should leave pinching and finer movements. Amputations through the middle phalanx result in little alteration of extension when only the more distal portion of the dorsal expansion of the extensor digitorum communis tendon is cut. By contrast, the flexor digitorum superficialis tendon has a long insertion on either side of the phalanx, and if a significant portion of this insertion cannot be left intact, disarticulation may be necessary. Whenever possible, amputation through the proximal phalanx is preferable to disarticulation at the metacarpophalangeal joint, since the stump, however short, has function. In the case of the index finger, a stump less of than one and a half phalanges usually gets in the way, and amputation through the metacarpal should be considered. Stumps that stick out when a fist is made should be avoided. The fifth ray is important for gripping, and even a short but mobile stump is advantageous.

When one or more fingers must be removed in entirety, attention should be directed to the hand as a working unit. When the long and ring fingers are both removed, the adjoining fingers may continue to function perfectly, or they may deviate toward each other, and deviation weakens flexion. If only the long or ring finger is removed, it is best also to remove the head of its metacarpal to allow the adjoining metacarpals to approximate each other. This tends to prevent adjoining fingers from rotating and crossing toward each other. As a secondary procedure, the index (in the case of amputation of the long finger) or the little finger (in the case of amputation of the ring finger) may be moved over by resecting the metacarpal of the missing finger through the proximal third of its shaft and shifting the adjoining finger onto the stump.

When the middle finger and either the index or little finger are removed together, the heads of the metacarpals should be cut in an oblique fashion to produce a smooth contour. When the index finger has been disarticulated, the metacarpal should be partially excised to improve the final appearance of the hand and to prevent its interference with function. In the case of the little finger, it is preferable not to resect the whole metacarpal, because the extensor and flexor carpi ulnaris tendons are attached to the base.

The thumb is the most important of the digits, and every bit possible should be saved, since even the smallest stump is preferable to a prosthesis. When only the thumb and small finger remain, a rotation osteotomy of the metacarpal of the thumb can be performed to permit approximation and grasping.

Forearm Amputation

To avoid excessive scarring and immobility of skin, skin flaps should not be dissected extensively from the fascia. The wound should run across, and not between the ends of the radius and ulna. The long flexor and extensor muscles and tendons are dissected and tapered so that they adhere to underlying tissue. After the bones have been transected, a margin of periosteum should be removed to prevent spur formation, and the fascia should be closed over the end so that muscle length is maintained.

With longer stumps, the actions of pronation and supination should be preserved, and the tendinous muscles and bones are therefore treated atraumatically to prevent fibrosis. If only a short stump can be achieved, this is superior to an above-elbow amputation. The formation of a functioning stump through the upper forearm can be successful if there is a short segment of radius or ulna that might be preserved distal to the tuberosity.

Upper Arm Amputation

Every attempt should be made to preserve as long a stump as possible, because the longer the stump, the greater the applicability of subsequent kineplastic procedures for functional prosthesis. If a high amputation is necessary, it is preferable to preserve the head of the humerus; this maintains the width of the shoulders and serves as a support for a prosthesis.

After anterior and posterior flaps are developed, the muscles are transected so that there is sufficient length of biceps and brachialis tendons and triceps muscle for these tendons to be joined over the end of the bone to avoid retraction.

When the stump is mature it can be fitted with a prosthesis with a terminal device operated by a shoulder cable and with an elbow that can be fixed at varying degrees of flexion. The type of elbow joint used depends on the type of function required by the patient. The most functional terminal devices are hooks that may be voluntarily opened. In the case of the adult unilateral amputee, the prosthesis usually functions only as an aid for the contralateral hand.

Extremity Replantation

The first report on the reattachment of a fully amputated hand, with return of useful function, was in 1963. In the ensuing years, refinements in microvascular techniques resulted in an increasing success rate. A summary of the experiences from China, Switzerland, and the United States in 1981 noted that one-third of the patients had excellent functional recovery, one-third had good results, and one-third had fair-to-poor results.

No shoulder replantation achieved excellent or good functional results. Good function resulted with replantation above the elbow in 0 to 40 percent of cases. Replantation at the proximal forearm was associated with excellent to good functional results in 0 to 70 percent. Much better results have been reported when replantation was carried out at the distal forearm, with success rates ranging from 50 to 83 percent. At the wrist, success rates are around 80 percent; at the palm, 25 to 70 percent; at the thumb, 32 to 90 percent; the fingers proximal to the *PIP* joint, about 70 percent; and distal to the PIP joint, almost 100 percent.

Guillotine injuries are associated with a higher degree of success after replantation than compression or crush injuries, with the worst results occurring in avulsion injuries. Reasonable success rates are achieved if the duration of anoxia is less than 20 h; the success rate drops sharply when anoxic periods are longer than 20 h.

The essential structures of the dismembered extremity must be intact and the severed limb reasonably well preserved if replantation is attempted. Following thorough debridement, the bones are repaired before the circulation is reestablished. A main vein is anastomosed before the arterial repair. The muscles, tendon, and nerves are then repaired and skin coverage is provided. Decompressive fasciotomy is frequently indicated.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 21 - Venous and Lymphatic Disease

Richard M. Green
Kenneth Ouriel

VENOUS DISEASE

Venous Anatomy and Physiology

Lower extremity veins can be divided into three types: superficial, deep, and perforating veins (Fig. 21-1). The systemic veins contain approximately two-thirds of the circulating blood volume under relatively low pressure, and venous flow from the lower extremities must overcome gravity and intraabdominal pressure to return blood to the right ventricle. The initial force produced by the left ventricle is reduced through the capillary bed to a pressure of about 15 mmHg in the venules. The calf muscles provide an additional pump function as they compress deep veins within an unyielding fascial compartment. Proximal flow is assured by the presence of the delicate but strong venous valves, which prevent reflux.

The superficial venous system is composed of the greater and lesser saphenous veins (GSV and LSV) and lies above the investing fascia. The GSV begins in the dorsum of the foot and ascends cephalad anterior to the medial malleolus. It runs along the medial aspect of the leg, crossing the knee joint 8 to 10 cm dorsal to the medial edge of the patella. The saphenous nerve accompanies the vein from the foot to the upper thigh, where it penetrates the fascia and enters the subsartorial canal along the superficial femoral artery. The GSV pierces Scarpa's fascia in the midthigh and enters the fossa ovalis in the groin, 4 cm lateral and inferior to the pubic tubercle. The LSV begins laterally from the dorsal venous arch, courses posterior to the lateral malleolus, and enters the popliteal vein between the medial and lateral heads of the gastrocnemius muscle. The sural nerve lies lateral to the GSV.

The deep veins are primarily responsible for lower extremity venous return. These veins follow the course of the major arteries and share their names. In the lower leg the veins are paired and join at the knee to form the popliteal vein, which continues through the adductor hiatus to become the superficial femoral vein. The latter is joined by the deep femoral vein in the upper thigh to become the common femoral vein, which becomes the external iliac vein as it enters the pelvis beneath the inguinal ligament. Numbers of valves increase with distance from the heart, though the vena cava and common iliac veins are valveless. Each valve is based within a dilated sinus of the vein, which keeps the valve cusps away from the walls and promotes rapid closure when flow ceases. Valves are the focal point of most of the pathology of venous thrombosis because their sinuses are where the initial thrombus forms, and the loss of valvular function after recanalization of a thrombus produces venous insufficiency (Fig. 21-2). Autopsies suggest that it is more common for thrombi to originate in the veins of the soleus and then propagate proximally, but there is evidence that primary thrombosis of the femoral and iliac venous tributaries occurs as well. There also are a number of venous sinuses within the substance of the soleus muscle, which empty into the posterior tibial vein, and in the gastrocnemius muscle, which empty into the popliteal vein. These sinuses are critical to the function of the calf muscle pump.

The perforating or communicating veins connect the superficial venous system with the deep and direct flow internally from the superficial veins in all areas of the lower extremity except the foot, where the opposite occurs. The perforating veins are so named because they penetrate the fascia of the lower leg to connect the superficial and deep systems. The perforators adjacent to the medial malleolus often are responsible for the development of stasis ulcers at that level when they become incompetent. Boyd's perforator connects the GSV to the deep veins 10 cm below the knee. Cockett's perforators connect the posterior arch vein with the posterior tibial vein and often become varicose. The Hunterian perforator connects the GSV to the superficial femoral vein; its incompetence accounts for many thigh varicosities when the saphenofemoral junction is competent.

When a person is in the supine position, lower extremity venous return is primarily dependent on the respiratory cycle. Intraabdominal pressure increases as the diaphragm descends during inspiration and the external pressure on the external iliac veins decreases venous return. Valve closure prevents significant reflux during inspiration. The converse occurs during expiration. Venous return is increased as intraabdominal pressure is decreased with upward movement of the diaphragm. When a person is in the upright position, venous flow is dependent on the contractile force of the heart, static filling pressure, and gravity. The expiratory enhancement of venous flow is insufficient, and the calf muscle pump is necessary to overcome the hydrostatic forces. A single contraction of the calf muscles can empty 60 percent of the blood pooled in the tibial veins and muscular sinuses. The standing venous pressure in the foot veins is 80 to 90 mmHg while standing and falls to 40 mmHg during ambulation (see Fig. 21-22).

Deep Vein Thrombosis

Etiology of Deep Vein Thrombosis

Three factors are primarily responsible for the development of a thrombus within a vein: abnormalities of blood flow, abnormalities of blood, and injury to the vessel wall. The occurrence of this process in a nontraumatized vein was recognized by Rudolf Virchow, who introduced the term thrombosis in 1856.

Stasis

Although stasis alone is not sufficient, it is the most important factor in the development of *deep vein thrombosis (DVT)*. The main event in the formation of a venous thrombus is the generation of thrombin in areas of stasis. This leads to platelet aggregation and fibrin formation. When contrast medium is injected into the veins of the lower extremities of a bedridden patient, it may remain in venous valve sinuses for as long as an hour, confirming the pooling effect in the soleal veins. Primary and secondary vortices are produced beyond the valve cusps, the favored location for the formation of a thrombus, and trap red cells to form the early nidus for thrombus formation. Early thrombi attach to normal endothelium and consist of loosely packed red cells within a fibrin network accompanied by a variable number of leukocytes. The propagation of the thrombus depends on the relative balance between activated coagulation and fibrinolysis. More commonly, in about 60 percent of patients the thrombus propagates without interrupting flow and develops a long floating "tail" that is more susceptible to breaking loose from its tenuous anchor within the valvular sinus. It is the latter sequence of events that is the most dangerous aspect of the disorder, because major pulmonary embolism can and does occur without premonitory signs or symptoms at its point of origin. This process can begin under general anesthesia in the operating room but usually requires other contributing factors such as shock, infection, trauma, or congestive heart failure. Aging, obesity, pregnancy, and malignant disease also are important risk factors.

Endothelial Damage

The role of endothelial injury is questionable. It appears that it is neither a necessary nor a sufficient condition for thrombosis. With the exceptions of hip arthroplasty and central venous catheters, there is little evidence that gross or microscopic venous injury has a role in venous thrombogenesis. Routine histologic examination of veins containing thrombus usually fails to show an inflammatory response consistent with vessel wall injury. It is possible that hypoxic or biochemical injury has a role, but definitive evidence is lacking.

Hypercoagulability

Abnormalities of the blood include aberrations of the clotting and fibrinolytic systems. Stasis and injury alone are not sufficient to cause thrombosis experimentally in the absence of low levels of activated coagulation factors. Patients who present at an early age with spontaneous venous thrombosis, who have a strong family history of *DVT*, or who develop recurrent venous thromboembolisms are usually considered "prothrombotic" or "hypercoagulable." These conditions are listed in Table 21-1.

Activated protein C resistance (APC-R) is a common hereditary condition that results in decreased efficacy of the natural anticoagulant protein C. It is transmitted as an autosomal dominant trait, and 90 percent of the cases are because of a mutation in factor V (factor V Leiden mutation). The syndrome was initially described in young patients with venous thrombosis but may also have a role in hypercoagulability after arterial reconstruction.

Antithrombin III deficiency often is associated with unexplained arterial thrombosis. It is found more often in patients with serum albumin levels less than 3.0 g/dL. Patients with antithrombin III deficiency present as a resistance to heparin. Typically, heparin is given and no increase in the *partial thromboplastin time (PTT)* is noted. In this setting, immediate anticoagulation can be achieved by providing substrate (fresh frozen plasma) in addition to heparin and then conversion to coumarin derivatives.

The antiphospholipid syndrome (APS) is another hypercoagulable state with recurrent thrombotic events and antibodies directed against phospholipids. There are primary (no associated autoimmune disease) and secondary forms. APS occurs in younger patients rather than in atherosclerotic populations, and very few of these patients smoke. Antiphospholipid antibodies, which include anticardiolipin, are being recognized with increasing frequency in association with a variety of thrombotic disorders. Their association with unexplained juvenile *DVT* indicates that screening for antiphospholipid antibodies should be included in the workup of any unexplained thrombosis. One-third of patients with systemic lupus erythematosus have antiphospholipid antibodies. Women with APS often have a history of spontaneous abortions. The diagnosis is suggested in a patient with an appropriate history and a spurious elevation of the *PTT*. Patients on estrogen therapy for postmenopausal replacement, birth control, or chemotherapy are at increased risk for venous thrombosis.

The association between venous thrombosis and cancer was first suggested by Armand Trousseau in 1856 and often has been confirmed in postmortem studies. In a series reported by Aderka and associates, 34 percent of otherwise healthy patients with idiopathic *DVT* were found to have malignant disease diagnosed an average of 24 months later. Increased likelihood of cancer in these patients was associated with age over 65 years, anemia, and eosinophilia. The earliest-onset malignancies were found within 1 year and usually occurred in the pelvic organs and breast.

Patients presenting with a thrombotic episode at a young age or those with previous events should be screened for hypercoagulability. Routine screening should include measurements of prothrombin time, activated partial thromboplastin time, hematocrit level, white blood cell count, sedimentation rate, and platelet count. Measurements of homocysteine levels, antiphospholipid antibodies, protein C and protein S, antithrombin III, activated protein C resistance, platelet aggregation, and mutant factor V should be done in very high-risk patients. Screening is difficult once anticoagulation has begun. For instance, coumarin derivatives interfere with measurements of proteins C and S and the functional assay for activated protein C resistance, heparin reduces circulating levels of antithrombin III, and antiplatelet drugs may produce false negatives when testing for heparin-induced thrombocytopenia.

Sequence of Pathology

The venous lumen is most often recanalized after an episode of *DVT*. This process is a result of spontaneous lysis and involves a complex series of cellular and humoral processes. Organization of the thrombus begins at the attachment zone as endothelial cells activate thrombus-bound plasminogen. This results in enlarging pockets within the thrombus and eventual fragmentation. The clot itself undergoes softening and contraction during this process, with the potential to restore the venous lumen. Serial studies using duplex ultrasonography show that the process of recanalization begins by day 7 in roughly 50 percent of thromboses and is uniformly observed by 90 days.

Recurrent thrombotic events compete with recanalization early in the course of a *DVT*. This encompasses those patients with propagation of clot in previously uninvolved areas, thromboses in another extremity, and rethrombosis of a partially recanalized segment. The incidence of these recurrences is reduced tenfold when patients are given adequate anticoagulation therapy for a 3-month period.

Clinical Manifestations

The site of venous obstruction determines the level at which swelling is observed clinically (Fig. 21-3). Calf vein thrombosis is localized to one or more of the three major named veins below the knee. Calf tenderness is frequently present, but because the thrombi are rarely completely obstructive and the veins are paired, swelling is not a universal finding. Femoral vein thrombosis usually is associated with swelling of the foot and calf. Iliofemoral venous thrombosis represents the most extensive form of DVT and usually is associated with tenderness in the groin and swelling of the entire leg. Major venous thrombosis involving the deep venous system of the thigh and pelvis produces a characteristic presentation of pain and extensive pitting edema. The extremity may have bluish discoloration (phlegmasia cerulea dolens) or blanching (phlegmasia alba dolens, or “milk leg”). The latter usually occurs in association with pregnancy. Other mechanical factors that can affect the left iliac vein include compression from the right iliac artery, an overdistended bladder, and congenital webs within the vein. These factors are responsible for the observed 4:1 preponderance of left versus right iliac vein involvement.

Phlegmasia cerulea dolens occurs when the venous thrombosis progresses and impedes most of the venous return from the extremity. There is danger of limb loss from cessation of arterial flow. Fortunately, this occurrence is rare. Loss of sensory and motor function and venous gangrene are likely unless an aggressive approach is implemented to remove the thrombus and restore blood flow. This condition almost always occurs with advanced malignant disease.

Diagnosis

Only 40 percent of patients with venous thrombosis have any clinical signs of the disorder. Homans's test is performed by dorsiflexing the foot. It is considered positive for DVT if the patient complains of calf pain. False-positive clinical signs occur in more than 30 percent of patients studied. Venous duplex ultrasonography has relegated other noninvasive tests, such as radioactive-labeled fibrinogen scans and all types of plethysmography, to historical interest. In some centers duplex scans have replaced contrast venography as the best diagnostic test for DVT. Accuracy rates above 90 percent have been consistently reported for venous duplex exams. Indications for duplex venous scans include patients with pulmonary emboli, patients with extremity pain or swelling, and patients at increased risk for developing a DVT. The latter group includes those with trauma, joint replacement, other major surgeries, prolonged immobilization, and known hypercoagulability states.

There are three essential phases to the venous duplex scan: (1) thrombus visualization, (2) vein compressibility, and (3) venous flow analysis. Accuracy is dependent on the examiner's skill. Thrombus may be difficult to visualize in its acute form, and the addition of color flow imaging facilitates the identification of nonoccluding clots. Thrombus echogenicity increases with age of clot. Venous compressibility is determined by placing the probe directly over the vein and applying gentle pressure while observing under B-mode imaging (Fig. 21-4). Veins filled with thrombus do not collapse with this maneuver. Venous flow assessment evaluates the respiratory phasicity and response to external extremity compression. Persistent lack of a flow signal indicates total obstruction. A negative scan performed by a well-trained ultrasonographer is sufficient to rule out a DVT of the lower extremity.

The role of venography has been diminished by the advances in ultrasound technology. Nonetheless, the injection of contrast material for direct visualization of the venous system of the extremity remains the most accurate method of confirming the diagnosis of venous thrombosis and the extent of involvement. The main indication for its use in the diagnosis of an acute DVT is a nondefinitive duplex scan. Injection usually is made into the foot while the superficial veins are occluded by tourniquet, and a supplemental injection into the femoral veins may be required to visualize the iliofemoral system (Fig. 21-5). Potential false-positive examinations may result from external compression of a vein or washout of the contrast material from venous flow from collateral veins.

Prophylaxis

Because the first manifestation of a DVT may be a fatal pulmonary embolus, some form of prophylaxis is indicated in high-risk patients. Patients older than 70 years of age, those with previous thromboembolism, malignant disease, paralysis, multiple trauma, or lower extremity joint surgical procedures have a very high risk for DVT. Prophylactic measures are directed toward altering blood coagulability or eliminating or reducing venous stasis.

Efforts to reduce stasis include elastic compression stockings, intermittent external leg compression, leg elevation, and early ambulation. Intermittent pneumatic leg compression is the most effective measure. It reduces stasis and increases fibrinolytic activity with virtually no side effects. The pneumatic boots can be applied in the operating room to minimize the risk of venous thrombosis beginning under general anesthesia and are of proven efficacy in patients undergoing total knee replacement, radical prostatectomy, or operations where adjuvant anticoagulation therapy is contraindicated.

Pharmacologic prophylaxis includes low-dosage unfractionated heparin (UFH), adjusted-dose heparin, low-molecular-weight heparin (LMWH), warfarin (international normalized ratio 2.0–3.0), and dextran 70. Prophylactic low-dosage subcutaneous UFH that does not alter the clotting profile has been extensively tested and is safe and effective in moderate-risk patients. A 5000-unit dose is given subcutaneously 2 h preoperatively and then every 12 h postoperatively for 6 days. This provides protection for most high-risk groups with the exception of those undergoing orthopaedic or urologic procedures. Higher-risk patients require adjusted-dose UFH with the activated partial thromboplastin time (APTT) held in the upper normal range. Both regimens are associated with an increased incidence of wound hematomas. The studies comparing the prophylactic use of LMWH and UFH have concluded that there is little difference between the two drugs. LMWH is ten times more costly than UFH, however.

There are good data to support the use of preoperative oral anticoagulant therapy with coumarin derivatives in high-risk patients. When given the night before operation, warfarin anticoagulation is achieved within 3 to 4 days. The “two-step” or minidose warfarin regimen is designed to circumvent the delay in anticoagulation. Warfarin is started at a dose of 1 mg 14 days before operation to prolong the prothrombin time (PT) by 2 or 3 seconds. This procedure increases the risk of hemorrhage, and because of the added difficulties of laboratory control of prothrombin time, there has not been widespread acceptance of this approach. A national task force on prophylaxis for patients undergoing hip surgery recommends warfarin or adjusted-dose heparin to prolong the APTT to the upper normal range. The administration of dextran, which produces a variety of effects on platelets and clotting factors, has been demonstrated to reduce the incidence of detectable thrombi. It too can produce hemorrhagic problems, allergic reactions and, in older patients, congestive heart failure. Recommendations for prophylaxis are listed in Table 21-2.

Medical Treatment

The approach to management of the patient with DVT is based on three objectives: minimizing the risk of pulmonary embolism, limiting further thrombosis, and facilitating resolution of existing thrombi to avoid the postthrombotic syndrome. The traditional treatment places the patient at bed rest with the foot of the bed elevated 8 to 10 inches. Intravenous UFH is administered, and oral warfarin is started when the patient's APTT is in a satisfactory range. As the pain, swelling, and tenderness resolve over a 5 to 7 day period, ambulation is permitted with elastic stocking support. Standing still and sitting should be prohibited to avoid increased venous pressure and stasis. Patients with large thrombus loads are candidates for fibrinolytic agents in an attempt to preserve venous valvular competence. The practice of mandatory bed rest and hospitalization has been challenged by the encouraging results of outpatient treatment of DVT with LMWH.

Anticoagulation

The foundation of therapy for DVT is adequate anticoagulation, initially with heparin and then with coumarin derivatives for prolonged protection against recurrent thrombosis. Unless there are specific contraindications, heparin should be administered in an initial dose of 100 to 150 units/kg intravenously. Heparin is an acid mucopolysaccharide that neutralizes thrombin, inhibits thromboplastin, and reduces the platelet release reaction. It may be administered by continuous or intermittent intravenous doses regulated by whole blood clotting time or APTT. Recurrent episodes of thromboembolism are 15 times more common in patients with inadequate anticoagulation treatment within the first 24 hours. Bleeding complications can be minimized by doses of heparin that prolong the laboratory clotting determinations by about twice the normal time. Continuous intravenous infusion regulated by an infusion pump minimizes the total dose required for control and is associated with a lower incidence of complications and no loss of effectiveness.

Thrombocytopenia is the most common complication of heparin therapy and is estimated to occur in 1 to 5 percent of patients receiving the drug. Unlike other drug-induced thrombocytopenias, heparin-induced thrombocytopenia often is associated with thromboembolic complications from antibody-mediated platelet activation. The paradox of thrombosis occurring in a patient receiving heparin was first described in 1958 by Weismann and Tobin. Towne and associates described the “white clot” syndrome, a peripheral vascular complication of heparin therapy in 1979. Heparin-induced thrombocytopenia (HIT) represents the prodrome to the thrombotic syndrome (HITTS) that occurs in 1 in 2000 patients who receive more than 20,000 IU of UFH per day for more than 5 days, 1 in 5 patients with HIT, and 1 in 3 patients who have heparin-dependent antiplatelet antibodies. Patients who develop HITTS have a mortality rate ranging from 25 to 37 percent resulting from diffuse uncontrolled clotting with limb ischemia and organ infarction. There are no known factors that predict risk. Development of thrombocytopenia from heparin is independent of sex, age, blood type, amount of heparin given, type of heparin, and route of administration. HIT has been documented after minimal heparin dosages such as those received with I.V. flushes and heparin-coated indwelling catheters.

Two forms of HIT exist. Type I, the most frequent, is mild (platelet counts $>100,000/\text{mm}^3$), reverses despite continuation of heparin, is due to a direct pro-aggregant effect of heparin, and is not associated with thromboses. In contrast, Type II HIT is severe (platelet counts $<100,000/\text{mm}^3$), resulting from antibodies binding to a platelet-heparin complex that leads to platelet activation and aggregation and is often associated with arterial or venous thromboses. The thrombocytopenia typically occurs after 5 days of heparin therapy but can occur earlier in patients who have a prior exposure to heparin. Type II HIT requires the immediate withdrawal of all heparin.

The most important advantage of the LMWHs over UFH is their superior pharmacokinetic properties, allowing their use without laboratory monitoring. LMWH preparations have been compared with UFH for the acute treatment of DVT in 13 well-designed trials. Pooled results from these studies show that LMWHs administered subcutaneously are as effective and safe as UFH but have the advantage of the potential for home treatment and do not require laboratory monitoring. These advantages may offset the increased cost of LMWH.

Oral administration of anticoagulants is begun shortly after initiation of heparin therapy. There is a risk in giving coumarin derivatives to a patient who is not already anticoagulated with heparin. The coumarin derivatives block the synthesis of the vitamin K-dependent clotting factors and inhibit vitamin K carboxylation of proteins C and S. These latter proteins are naturally occurring anticoagulants that function by inhibiting activated factors V and VIII. A vitamin K antagonist potentially can create a hypercoagulable state before achieving its anticoagulant effect because the half-lives of proteins C and S are shorter than the half-lives of the other clotting factors. Heparin should be continued for the 4 to 5 days required to achieve full anticoagulation with coumarin derivatives.

Data from prospective studies indicate that the level of anticoagulation with coumarin derivatives are effective at an international normalized ratio of 2.0 to 3.0. Higher levels are not more effective and are associated with a higher incidence of bleeding complications. Administration of fresh frozen plasma usually can normalize the prothrombin time and control hemorrhagic complications. Coumarin derivatives cross the placenta and should not be used during pregnancy. After an episode of acute DVT, anticoagulation therapy should be maintained for a minimum of 3 months; some investigators favor 6 months for treatment of thrombi in the larger veins. Many drugs interact with coumarin derivatives (e.g., barbiturates), and it is essential that a routine for regular monitoring of prothrombin time be established after the patient leaves the hospital.

Thrombolysis

Anticoagulant therapy is designed to prevent recurrent thromboembolism. Ideally, a treatment would be available with the potential to eliminate the thrombus and maintain valvular function. A number of trials have been performed comparing thrombolysis with standard anticoagulant therapy. Complete clearing of thrombus was noted in 45 percent of patients treated with thrombolytic agents, compared to 4 percent of those treated with heparin. This seems to translate into a long-term improvement in venous function. Popliteal valve incompetence was documented in 77 percent of those patients who did not have clearing, compared to a 9 percent incidence in those with complete lysis. These agents have no advantage over heparin in the treatment of recurrent venous thrombosis or thrombosis that has existed for more than 72 h, and they are contraindicated in postoperative or posttraumatic patients.

There are more bleeding complications with thrombolytic treatment, and this approach is reserved for those patients with clot in the common femoral and iliac venous systems. In a prospective study of 29 patients with thrombosis involving the popliteal veins, with or without involvement of calf veins, Kakkar compared hemodynamic and clinical results in patients receiving 5-day treatment with heparin or streptokinase, followed by a 6-month course of a coumarin derivative. Overall, at 2-year follow-up they found more than half of the limbs to have evidence of the postthrombotic syndrome. Clinically, 14 percent of patients had no symptoms, 20 percent had severe symptoms, and the remainder demonstrated mild to moderate changes. No difference was seen between patients receiving heparin or streptokinase. These studies were done with systemic infusion of lytic agents. Studies evaluating the use of catheter-directed thrombolysis have had encouraging results in selected areas and are discussed in detail in the section Subclavian Vein Thrombosis.

Surgical Approaches

Operative Thrombectomy

There are very few indications for operative thrombectomy because catheter-directed thrombolysis is so effective in treating iliofemoral venous thrombosis. The procedure is reserved for limb salvage in the presence of phlegmasia cerulea dolens and impending venous gangrene and in patients with a contraindication to thrombolysis. The direct surgical approach to remove thrombi from the deep veins of the leg uses the common femoral vein and is facilitated by the use of a Fogarty venous balloon catheter and an elastic wrap for milking the extremity (Fig. 21-6). Results are improved when the extent of thrombus is documented preoperatively, when completion phlebography is performed to assure complete clot removal, when a small arteriovenous fistula is constructed to maintain high blood flows in the iliac vein, and when anticoagulation therapy is given over prolonged periods. Early results in 57 patients treated in this fashion reported by Einarsson and associates showed patency of the iliofemoral segment by venography in 61 percent, and 75 percent had a good clinical result. Measurement by venous function, however, using plethysmography and foot volumetry, showed normal results in only 29 percent.

The use of arteriovenous fistulas after iliofemoral thrombectomy or reconstruction of the venous system is controversial. Most of the experience has been accumulated in Europe, where it is believed to reduce the incidence of early rethrombosis. The two most commonly used sites are the femoral triangle and the ankle. After surgery on the iliofemoral system, an H-shaped fistula can be established easily by anastomosing a branch of the saphenous vein end-to-side to the proximal portion of the superficial femoral artery. At the ankle, the posterior tibial artery may be anastomosed to the posterior tibial vein or the greater saphenous vein. Two problems have led to the reluctance of some surgeons to adopt this procedure: the fear of damaging functioning valves distal to the fistula and the requirement for a second operation to close the fistula. Fistulas usually are closed 3 to 4 months postoperatively, and problems with incompetent valves distal to the fistula have not been reported. Two steps during primary venous reconstruction simplify operative closure of the fistula later. The fistula is made distal to the venous reconstruction, avoiding damage to this area at reoperation, and a ligature is wrapped around the fistula and left in the subcutaneous tissue, where it can be found under local anesthesia. Obliteration of the fistula percutaneously by a detachable balloon has been described.

Vena Caval Interruption

Vena caval interruption also is indicated when there is a contraindication to or failure of anticoagulation therapy (Table 21-3). Early operations in which one or both common femoral veins were ligated were associated with high recurrence rates and a high incidence of sequelae because of stasis in the lower extremity. Control of the inferior vena cava for ligation required a laparotomy and added the adverse effect of a sudden reduction in cardiac output under general anesthesia. This effect, coupled with stasis sequelae and recurrent embolism through dilated collateral veins, led to efforts to compartmentalize the vena cava by means of sutures, staples, and external clips in order to provide filtration without occlusion (Fig. 21-7).

Because these procedures required general anesthesia and laparotomy, the next logical step was to devise a transvenous approach that could be performed under local anesthesia. The Mobin-Uddin "umbrella" unit was inserted from the jugular vein and positioned under fluoroscopic control below the renal veins. The incidence of vena cava occlusion was 70 percent, and fatal embolism sporadically occurred with device migration.

The Greenfield cone-shaped filter was developed to maintain patency after trapping emboli. This is possible because of the geometry of the cone, which collects emboli in its apex and retains perimeter flow. Preservation of flow avoids stasis and facilitates lysis of the embolus (Fig. 21-8). It can be inserted percutaneously from either the jugular vein or the femoral vein. The rate of recurrent embolism with this device has been 4 percent over 12 years of follow-up. Its long-term patency rate in excess of 95 percent allows it to be placed above the renal veins when necessary for embolism control, such as when there is a thrombus within the renal veins or the vena cava. There are a number of proprietary devices available for percutaneous insertion that are equally successful in preventing pulmonary embolism.

Superficial Thrombophlebitis

The term thrombophlebitis should be restricted to a disorder of the superficial veins characterized by a local inflammatory process that usually is aseptic (Fig. 21-9). Patients present with a painful swelling and erythema along the course of a superficial vein. The cause of thrombophlebitis in the upper limb usually is acidic fluid infusion or prolonged cannulation. In the lower extremity it is often associated with varicose veins and may coexist with *DVT*. Its association with the injection of contrast material can be minimized by washout of the contrast material with heparinized saline. The diagnosis usually is obvious. Duplex scans are very accurate in confirming the diagnosis and should be performed especially when swelling is present to rule out a concomitant deep venous problem.

Symptoms usually last for 2 to 3 weeks. Nonsteroidal anti-inflammatory agents provide significant pain relief. Patients should not be kept at bed rest. Activity should be encouraged with the extremity in external elastic support. If the thrombus extends into the saphenofemoral junction, the patient should have the saphenous vein disconnected from the common femoral vein or undergo full anticoagulation therapy.

Thrombophlebitis Migrans

Thrombophlebitis migrans, a condition of recurrent episodes of superficial thrombophlebitis, can be associated with visceral malignancy, Buerger's disease, the hypercoagulable states, systemic collagen vascular disease, and blood dyscrasias. Involvement of the deep veins and the visceral veins also has been described. The presence of this condition should alert the clinician to search for an underlying condition.

Subclavian Vein Thrombosis

Subclavian venous thrombosis (SVT) may be associated with an anatomic abnormality at the thoracic outlet, may be related to the placement of a central venous catheter, or may occur in a hypercoagulable patient. If left untreated, 25 to 74 percent of affected patients will have some limitations in activity, as many as 12 percent will have a pulmonary embolus, and 1 percent will die. Those cases related to a thoracic outlet abnormality make up 0.5 percent to 1.5 percent of all venous thromboses and often are associated with strenuous activities. These cases are referred to as the Paget-Schroetter syndrome or "effort thrombosis." Catheter-induced thrombosis is an increasingly common event because of the more frequent use of central veins for access, nutrition, chemotherapy, and monitoring. Screening venography in patients with central venous catheters demonstrates that 33 to 60 percent have thrombus in the axillosubclavian segments. Clinically evident SVT develops in 3 percent of these patients.

Patients with SVT present with a bluish, swollen arm and a pattern of upper extremity venous hypertension. Collateral veins usually are visible around the shoulder and chest wall. Patients typically describe an aching pain that is exacerbated by exercise. The color duplex scan has virtually replaced contrast venography in the diagnosis of lower extremity venous thrombosis, but the opposite is true in the upper extremity. Venography has a greater diagnostic accuracy when performed with the catheter in the basilic vein. A typical SVT is shown in Fig. 21-10.

Data support a role for conventional anticoagulation in all patients with SVT for prophylaxis against pulmonary embolism and for reduction of residual symptoms. Aggressive surgical therapy focused on clot removal is no longer recommended. Instead, physically active patients with SVT diagnosed in the acute phase should undergo catheter-directed thrombolysis. Current protocols recommend urokinase delivered via a catheter placed through the thrombus. This approach is successful in more than 50 percent of patients and ideally identifies the anatomic cause of the thrombosis (Fig. 21-11). Although the timing of correction is debated, most surgeons agree that extrinsic and extrinsic venous lesions should be aggressively treated. This means eliminating the compression at the costoclavicular space by a transaxillary first rib resection or a medial claviclectomy. The former is used when a direct approach to the vein is not required; the latter is used when the vein requires repair.

Inferior Vena Caval Thrombosis

Thrombosis of the inferior vena cava can result from tumor invasion or propagating thrombus from the iliac veins. Tumors of the vena cava are rare and are usually malignant and have a poor prognosis. They may be primary, such as a leiomyosarcoma, or secondary, such as a hypernephroma and a retroperitoneal sarcoma. Symptoms and signs depend on the segment of vena cava affected by the tumor and the degree of obstruction of adjacent organs. Involvement of the suprahepatic cava may cause the Budd-Chiari syndrome. The diagnosis can be made with a variety of imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), ultrasonography, and contrast cavography.

The most common cause of vena cava obstruction is iatrogenic, resulting from ligation, plication, or insertion of partially occluding caval devices. Any caval filtration device can become totally occluded by a trapped massive thrombus, causing sudden reduction in venous return and cardiac output. In a patient with known prior pulmonary embolism, it is a grave error to ascribe the resulting hypotension to recurrent pulmonary embolism and treat the patient with vasopressor agents. In this situation, the cause of the hypotension may be functional hypovolemia that can readily be confirmed by measurement of central venous pressure. Thrombosis of the renal vein can result from extension of vena caval thrombosis, but it is most likely to occur in association with the nephrotic syndrome. It can be a source of thromboembolism and has been treated successfully by suprarenal placement of the Greenfield filter.

Visceral Venous Thrombosis

Mesenteric venous thrombosis is an uncommon but often fatal cause of intestinal ischemia. Patients present with nonspecific abdominal complaints. Pain out of proportion to abdominal findings is a common occurrence. Unlike arterial ischemia, progression is slow. Delay in diagnosis is frequent. Laboratory findings are not usually helpful. Abdominal CT scanning is the best diagnostic tool for demonstrating thrombus in the superior mesenteric vein. Color duplex scans also are accurate, especially if performed early in the course of the disease. The event may be spontaneous and idiopathic or associated with hypercoagulability, pancreatitis, trauma, cirrhosis, splenomegaly, or infection. Polycythemia is the most common associated condition. The goal of treatment is to prevent intestinal necrosis, or, if bowel necrosis has already occurred, to resect the nonviable areas. Anticoagulation therapy with heparin is the mainstay of medical management. Patients with local or diffuse peritonitis should undergo abdominal exploration.

Portal vein thrombosis can occur in the neonate, usually secondary to propagating septic thrombophlebitis of the umbilical vein. Collateral development leads to the occurrence of esophageal varices. In the adult, thrombosis of the portal, hepatic, splenic, or superior mesenteric vein can occur spontaneously but usually is associated with hepatic cirrhosis or a hypercoagulable state.

Hepatic vein thrombosis (Budd-Chiari syndrome) usually produces massive hepatomegaly, ascites, and liver failure. It can occur in association with a congenital web, endophlebitis, or polycythemia vera. Some success has been reported using a direct approach to the congenital webs, but the usual treatment is a side-to-side portacaval shunt to allow decompression of the liver.

The development of pelvic sepsis after abortion, tubal infection, or puerperal sepsis can lead to septic thrombophlebitis of the pelvic veins and septic thromboembolism. Ligation of the ovarian vein and vena cava has been the traditional treatment, but the emphasis should be on drainage or excision of the abscesses and appropriate antibiotic therapy. It also is appropriate to use the Greenfield filter in this situation because it is inert stainless steel and avoids the development of an intraluminal abscess that can occur after ligation of the vena cava, as demonstrated experimentally by Peyton and associates in 1983.

Pulmonary Thromboembolism

Pulmonary embolism is the third leading cause of death from cardiovascular events, second only to myocardial infarction and stroke. Estimates of the mortality in the United States alone range from 50,000 to 200,000 per annum. It may be the most common form of preventable hospital death. It is estimated that 5 of every 1,000 adults undergoing major surgery will die from massive pulmonary embolism.

Virchow first recognized the association between the venous thrombosis and pulmonary embolus after performing autopsies in 76 patients with antemortem thrombi obstructing their pulmonary arteries. It also became obvious in the early reports by pathologists that pulmonary embolism could be well tolerated by some patients who then died of other causes. The full spectrum of the disorder ranges from asymptomatic minor embolism to sudden death from massive embolism. Autopsy studies consistently demonstrated a 10 to 15 percent incidence of fatal pulmonary embolism until the 1970s. With aggressive prophylaxis, the incidence has been reduced to 6 percent.

Clinical Manifestations

The signs and symptoms of an embolic episode depend primarily on the magnitude of embolus and, to a lesser extent, on the cardiopulmonary status of the patient. Less than 33 percent of patients with documented pulmonary embolism show clinical signs of venous thrombosis. The diagnosis is unsuspected in the majority of patients who die of pulmonary embolism. The vast majority of patients suddenly develop chest pain or dyspnea. Other early symptoms may include tachypnea, diaphoresis, and marked anxiety. Hemoptysis is an uncommon sign, and when present it usually occurs late in the course of the disease and represents pulmonary infarction. Objectively, the patient with major embolism usually shows tachycardia, an increased pulmonary second sound, cyanosis, prominent jugular veins, and varying degrees of collapse. Less commonly, there may be wheezing, a pleural friction rub, splinting of the chest wall, rales, low-grade fever, ventricular gallop, and wide splitting of the pulmonic second sound. The incidence of these findings found in the Urokinase Pulmonary Embolism Trial is shown in Table [21-4](#).

The differential diagnosis includes esophageal perforation, pneumonia, septic shock, and myocardial infarction. Because all these entities are life-threatening, it is mandatory that an orderly approach be formulated to confirm or reject the working diagnosis. Laboratory studies in general are not very helpful in the differential diagnosis, although a white blood cell count of less than 15,000/mm³ might be suggestive when a pulmonary infiltrate is present to help rule out pneumonitis.

Diagnostic Studies

Electrocardiography

The primary value of the electrocardiogram (ECG) is to rule out a myocardial infarction. In the presence of a pulmonary embolus, the ECG may have signs of right ventricular overload such as the S1, Q3, T3 pattern. This only occurs in 16 percent of patients with documented pulmonary embolism. More commonly, the ECG has nonspecific ST and T wave changes that are nondiagnostic.

Chest Radiography

The primary value of the chest radiograph is to exclude other diagnostic possibilities, such as pneumonia, pneumothorax, esophageal perforation, or congestive heart failure. Although central vascular enlargement, asymmetry of the vascular markings with segmental or lobar ischemia (Westermark's sign), or pleural effusion may suggest pulmonary embolism, they are rarely sufficient to establish a diagnosis. The chest radiograph is critical in the interpretation of a lung scan, because any radiographic density or evidence of chronic lung disease makes a perfusion defect less likely to represent a pulmonary embolism. Pulmonary infarction as a consequence of embolism is a rare finding. A wedge-shaped density usually is seen on the chest radiograph (Fig. [21-12](#)).

Arterial Blood Gases

Hypoxemia with PaO₂ of less than 60 mmHg is highly suggestive of pulmonary embolism, especially when the chest radiograph does not show any other pulmonary pathology. The low PaO₂ is believed to be a result of shunting by overperfusion of nonembolized lung and a widened alveolar-arterial oxygen gradient from reduced cardiac output. The reduction in arterial P_{CO2} that follows major embolism is the most discriminating finding because hypoxemia is present in several disorders likely to be misdiagnosed as massive embolism (e.g., septic shock). If hypoxemia and hypocapnia are not present, the diagnosis of major embolism in the severely ill patient is unlikely, and an alternative diagnosis should be sought.

Central Venous Pressure

Low *central venous pressure (CVP)* virtually excludes pulmonary embolism as the primary cause of the hypotension because massive embolism almost always is accompanied by right ventricular overload and elevated right atrial pressures. Elevated right ventricular filling pressures may be transient, however, as hemodynamic accommodation occurs, and in subacute or chronic embolism the central venous pressure may be normal.

Lung Scan

The most commonly used diagnostic test is the perfusion lung scan. A normal scan rules out a major pulmonary embolus, but an abnormal scan does not ensure the diagnosis. In a nonhypotensive patient with a normal chest radiograph, the lung scan is a valuable screening test that has increasing validity as the size of the perfusion defect approaches lobar distribution (Fig. [21-13](#)). Smaller peripheral perfusion defects are more difficult to interpret because pneumonitis, atelectasis, or other ventilation abnormalities alter pulmonary perfusion. When the ventilation/ perfusion lung scan is interpreted as high probability, the diagnostic accuracy as compared to pulmonary angiography is 96 percent.

Pulmonary Arteriography

Selective pulmonary arteriography is the most accurate method of confirming the presence of pulmonary emboli and should be performed for any equivocal ventilation/perfusion scan. The procedure is invasive, requiring passage of a cardiac catheter into the pulmonary artery for injection of a bolus of contrast medium. A series of radiographs that outline areas of decreased perfusion and usually show filling defects or the rounded trailing edge of impacted emboli (Fig. [21-14](#)) is obtained. Straight cutoffs of the smaller pulmonary arteries are more difficult to interpret, particularly if there is associated chronic lung disease that obliterates pulmonary vessels. The procedure can be performed with low risk, although pulmonary hypertensive and cardiac patients are at highest risk for this type of study, which usually carries a 0.3 to 0.5 percent mortality rate. Avoidance of injection of contrast medium into the main pulmonary artery minimizes the complications and mortality rates. Additional useful information is obtained before contrast injection by measurement of pulmonary arterial pressures. A normal pulmonary angiogram excludes the diagnosis of pulmonary embolism in acutely ill patients.

Pathophysiology

It is estimated that 85 to 90 percent of all pulmonary emboli originate from the veins of the lower extremity, and the remainder arise from the right side of the heart or other veins. Once the embolus has lodged and interrupted pulmonary blood flow, the ratio of regional ventilation to perfusion increases, and the lung responds by bronchoconstriction to reduce wasted ventilation. This response is mediated by a local reduction in CO₂ output because it can be prevented by ventilation with increased concentration of CO₂. The bronchoconstriction is exacerbated by the release of serotonin from platelets adherent to the embolus. The ability of heparin to inhibit the release of serotonin adds further justification to the early use of this drug. Other vasoactive agents, such as histamine and prostaglandins, may have a role, but the net effect is a reduction in size of peripheral airways, reduced lung volume, and reduced static pulmonary compliance. The hypoxemia that characterizes major embolism is thought to be due to a ventilation-perfusion imbalance secondary to the ventilation changes described above, although the findings in some patients resemble true arteriovenous shunting. Although there may be some improvement in PaO₂ after supplemental oxygen is administered, the effects usually are minimal. The pulmonary vascular and cardiac effects of embolism are a direct consequence of the degree of occlusion of the pulmonary vascular bed. The loss of more than 30 percent of the vascular tree is required to begin to elevate mean *pulmonary artery (PA)* pressure, and usually more than 50 percent occlusion is required to reduce systemic pressure.

Treatment

Anticoagulation

The hemodynamic variables previously described provide a means of classification of patients that uses four grades of severity and is a useful guide to therapy and prognosis (Table 21-5). The minor degrees of embolism usually can be managed with anticoagulants alone with a satisfactory outcome. Continuous-infusion heparin is the initial treatment, in a dosage designed to prolong the partial thromboplastin time to at least twice normal (approximately 150 units/ kg). Most clinicians also begin oral anticoagulation therapy to allow several days' overlap of the drugs as prothrombin time is extended into the therapeutic range. Adequate anticoagulation stops the progression of thrombosis and is associated with a recurrence rate of less than 5 percent.

Fibrinolytic Therapy

Emboli typically undergo dissolution as a result of the active fibrinolytic mechanism in the pulmonary circulation. Activation of plasminogen to plasmin, which is found in high concentration in the pulmonary circulation, promotes this fibrinolytic effect. Fibrinolytic agents have been administered to increase the rate of lysis after pulmonary embolism. Tissue plasminogen activator (t-PA) works more rapidly than urokinase, but both are costly and are associated with a high incidence of bleeding complications. Indications for fibrinolytic treatment include any critically ill patient with a pulmonary embolus. Bleeding complications can be reduced by taking a careful neurologic history to eliminate patients with any brain pathology, minimizing venous and arterial punctures, stopping heparin administration, and identifying any other potential causes of bleeding.

The advantage of thrombolytic therapy may well be to improve the ultimate resolution of major thromboembolism, as demonstrated by Sharma and associates. Their follow-up studies in patients treated with urokinase or streptokinase showed a better restoration of pulmonary-capillary blood volume and diffusing capacity at 2 weeks than in patients treated with heparin and anticoagulants alone. The reason for the continued improvement at 1 year was not clear but was believed to be related to more complete early resolution of the

embolic condition, allowing more effective natural lytic processes, or to more complete clearance of peripheral venous thrombi, preventing silent recurrent embolism. Therefore, the patient who is not in shock and who has no clear contraindication should be treated with a fibrinolytic agent.

Vena Caval Interruption

In some patients, anticoagulants cannot be used because of associated problems (e.g., peptic ulcer disease), and management must be directed toward a mechanical means of protection against recurrent embolism as outlined earlier (see Table 21-3). Other patients, in whom anticoagulation appears to be adequate, sustain recurrent embolism and become candidates for surgical intervention. The third indication is when there has been a complication of anticoagulant therapy, forcing it to be discontinued and leaving the patient with untreated *DVT*. Another indication for a vena caval filter is protection against recurrent embolism in a patient who has sustained massive pulmonary embolism requiring open or catheter embolectomy. In these patients, in spite of a satisfactory embolectomy of the pulmonary circulation, the original focus of venous thrombosis remains untreated and recurrent embolism is likely.

There are two additional relative indications for a vena caval filter in a patient with active or recent *DVT*. One is the high- risk patient over 40 years of age who is obese and has a serious associated medical illness (e.g., heart disease), malignant disease, or a history of previous embolism and who undergoes a major abdominal or vascular procedure. The final relative indication is the patient in whom 40 to 50 percent of the vascular bed has been occluded (major) and who would most likely not be able to tolerate additional emboli, particularly if there is associated cardiac or pulmonary disease.

Pulmonary Embolectomy

The direct surgical approach to pulmonary embolism can be traced back to Trendelenburg (1908), who demonstrated the feasibility of pulmonary embolectomy experimentally but had no successes clinically. It remained for his pupil Kirschner (1924) to confirm the possibility of embolectomy by a successful clinical outcome. Because this procedure was attempted without circulatory support using a direct approach to the pulmonary artery at thoracotomy, the number of survivors was very small, and the first successful case in the United States was not reported until 1958 by Steenburg. The first successful open embolectomy during cardiopulmonary bypass was reported by Sharp in 1962. Since then, partial bypass support has also been used for the patient in shock. Local anesthesia is used, and the femoral artery and vein are cannulated for venoarterial bypass. The equipment is fully portable (Fig. 21-15), and patients can be supported during pulmonary arteriography and then transported to the operating room, where they can tolerate general anesthesia and sternotomy much better while being maintained on partial cardiopulmonary bypass.

Emergency pulmonary embolectomy rarely is indicated but should be considered in any patient with an acute pulmonary embolism who appears preterminal. These patients rarely survive the trip to the angiography suite and operating room, but when they do, the operation is probably not necessary if fibrinolytic therapy is available. Documentation of the diagnosis of massive pulmonary embolism by pulmonary arteriography is mandatory because the clinical diagnosis often is incorrect. The initial approach to patients who have transient collapse or persistent systemic hypotension should include full heparinization and administration of inotropic drugs, if necessary, to support the circulation while the diagnosis is confirmed. Isoproterenol (4 mg in 1000 mL of 5% dextrose in water) is useful initially because of its bronchodilating and vasodilating effects and its positive inotropic cardiac effect. It may provoke arrhythmias, however, necessitating the use of dopamine. In the patient who responds to heparin and does not require vasopressors for systemic pressure or urine output, careful monitoring is essential to determine whether anticoagulation and fibrinolysis will control the disorder. Open pulmonary embolectomy carries a mortality rate in the range of 50 percent, however, and uncontrollable pulmonary hemorrhage may follow open restoration of pulmonary perfusion.

An alternative approach using local anesthesia has been suggested by Greenfield and associates for transvenous removal of pulmonary emboli. A cup device attached to a steerable catheter is inserted in the jugular or the femoral vein, and the cup is positioned under fluoroscopy adjacent to the embolus seen on arteriography (Fig. 21-16). The position is verified by injection of contrast medium through the catheter. Then syringe suction is applied to aspirate the embolus into the cup, where it is held by suction vacuum as the catheter and captured embolus are withdrawn. Clinical experience with the technique in 32 patients demonstrated that emboli could be extracted in 29 of them (91 percent) with an overall survival rate of 76 percent. Emboli could not be removed when they had been impacted for more than 72 h or if the patient suffered cardiac arrest at the time of angiography, in which case open embolectomy was required. Placement of a Greenfield vena caval filter after removal of sufficient emboli to produce near- normal hemodynamics protected the patients from recurrent embolism.

Pulmonary Hypertension and Thromboembolism

Pulmonary emboli may accumulate gradually over a prolonged period if they fail to undergo lysis and obliterate the pulmonary vascular bed. The clinical picture in this case is one of chronic cor pulmonale because significant pulmonary hypertension results from changes in the pulmonary vascular bed. The presentation may be subtle with only dyspnea or syncope on exertion, but there is a loud P₂ and right-sided strain on the electrocardiogram. The sequence also may occur unaccompanied by significant respiratory symptoms, and this may explain the cause in some of the patients considered to have primary pulmonary hypertension. When the diagnosis is made, there is limited life expectancy, but the patient may benefit from a vena caval filter to prevent additional embolism even if the disorder is primary pulmonary hypertension, as reported by Greenfield and associates. The rationale for this is that they ultimately develop right heart failure, predisposing to pulmonary embolism that is lethal even if small. When acute cardiopulmonary decompensation occurs in these patients after embolism, they are not good candidates for embolectomy because of fixation of the older thrombi to the pulmonary arterial wall. They should be classified separately (chronic) and managed by long-term anticoagulation therapy, or in some cases should be considered for open pulmonary thromboendarterectomy or heart-lung transplantation.

Recurrent thromboembolic pulmonary hypertension produces exertional dyspnea and signs of right heart strain with cor pulmonale. With further progression of right heart overload, tricuspid insufficiency may develop. This disorder may be difficult to distinguish from primary pulmonary hypertension, although the latter is more likely to be found in women under 20 years of age without a history of *DVT*. Severe pulmonary hypertension is a serious problem and usually limits the life expectancy to less than 2 years from diagnosis.

Open thrombectomy for chronic occlusion was first performed by Allison and associates in 1958 and remains a possibility for improving pulmonary blood flow. For a patient to be eligible for this procedure, the occlusion must involve the proximal portion of the pulmonary arterial tree, and the distal bed must be patent. The physiologic basis for continued distal patency after proximal occlusion is bronchial arterial collateral flow. The procedure also has a significant mortality, but this has been decreasing with greater experience and identification of risk factors. Daily and associates performed pulmonary thromboendarterectomy on 127 patients under deep hypothermic circulatory arrest with a mortality rate of 12.6 percent. For the majority of patients with severe pulmonary hypertension, however, the outlook is poor unless they receive maximum protection from recurrent embolism, which in the authors' experience has required anticoagulation therapy and vena caval filter placement.

Fig. 21-17 presents an algorithm for the management of pulmonary embolism.

Varicose Veins

Varicose veins are the most common vascular disorder affecting human beings. A definition of a varicose vein that lays the groundwork for a unified theory of causation was given by Carl Arnoldi, who described them as “any dilated, elongated, or tortuous vein, irrespective of size.”

Etiology

There are four factors that affect the development and progression of varicose veins: heredity, female sex hormones, gravitational hydrostatic force, and hydrodynamic muscular compartment forces. A familial tendency toward the development of varicosities may be the most important predisposing factor. Female sex hormones also have a profound effect on the lower extremity superficial veins. Varicose veins are common occurrences in pregnancy, usually appearing in the first trimester (70 to 80 percent) when the corpus luteum is secreting progesterone. Progesterone is known to inhibit smooth muscle contractility and increase venous distensibility. These effects are maximal on the first day of the menstrual cycle, when the effects of progesterone are amplified by estrogen. Hydrostatic forces produce venous dilatation from the weight of the blood column transmitted through incompetent valves. The other force is exerted by the contracting muscles on adjacent veins via the perforating system. These forces regularly exceed 150 mmHg.

Clinical Manifestations

The usual distribution of varices is below the knee in branches of the greater saphenous system (Fig. 21-18). The symptoms associated with varicose veins are nonspecific aching and heaviness of the legs that can be attributed to the congestion and pooling of blood in the enlarged superficial venous system. These symptoms worsen as the day progresses, requiring the patient to rest with leg elevation to obtain relief. Calf-length elastic stocking support in the range of 20 to 30 mmHg may provide symptomatic relief for those whose vocations require long periods of standing or sitting. Although mild edema may occur from varicosities alone, it usually reflects additional incompetence of the deep or perforating venous system and other medical conditions, such as cardiac or renal failure. Associated night cramping of the legs may be helped by the administration of quinine sulfate, which reduces muscular irritability.

Diagnosis

The Trendelenburg test is useful in distinguishing between primary varicose veins and the more serious condition of varicosities secondary to underlying deep venous disease. In the Trendelenburg test the limb is elevated to evacuate the veins, then pressure by hand or tourniquet is applied to the saphenofemoral junction (Fig. 21-19). With the patient standing, the lower leg is observed for the rate of filling of the varicosities. Gradual filling occurs in normal patients when the perforating veins are competent. Rapid filling occurs if the perforators are incompetent. The second phase of the test consists of release of the pressure to see if the upper thigh varices fill rapidly, indicating incompetence of the saphenofemoral valve. There are four possible results of this test. A *negative-negative* result occurs when there is only gradual filling in the distal one-third of the leg with compression in place and only continued slow filling when the compression is released. This indicates that the valves of the perforating veins (phase I) and superficial veins (phase II) are competent. In a *negative-positive* result, the release of compression is followed by a rapid filling of the saphenous vein, indicating that its valves are incompetent. A *positive-negative* result indicates that the perforating veins are incompetent, but the superficial veins are competent. A *positive-positive* result indicates that both systems have valvular incompetence. These principles have been refined by the use of color-flow duplex scanning. Changes in the direction of flow are detected by changes in color, and venous valves may be seen on the grey- scale image.

In the Perthes test, a tourniquet is placed around the upper leg and the patient is instructed to walk. If the varicose veins disappear, the deep venous system is patent and the perforating veins are competent. If pain occurs with walking, the deep system is obstructed and the superficial system represents the major source of venous outflow. It would be a serious error to excise superficial veins under these circumstances. Sequential tourniquets also may be used to define and isolate areas of incompetent perforating veins (Ochsner-Mahorner test).

Treatment

The majority of patients can be managed by conservative methods, but if these fail to control symptoms or if additional complications of venous stasis develop, such as dermatitis, bleeding, thrombosis, or superficial ulceration, the patient may become a candidate for more aggressive management.

The two methods of treatment currently used are injection sclerotherapy and ablative surgery. Sclerotherapy should be reserved for those patients who do not have evidence of axial saphenous reflux. Large varices of the thigh also should be removed surgically because they are subject to superficial thrombophlebitis after sclerotherapy and are often associated with large perforating veins. Injection sclerotherapy destroys the endothelium of the vein and promotes its obliteration by scarring. Pressure must be applied to the vein after injection of the sclerosant to prevent thrombus formation and later recanalization. The technique for injection involves placement of the needle and syringe with the patient standing followed by elevation of the leg, injection of the agent, and bandage compression of the area for 2 to 3 weeks. Sodium chloride 23.4% is the agent most preferred, but a wide selection is available. Current indications for sclerotherapy include superficial venules (<1 mm), varicosities 1 to 3 mm in diameter, postoperative residual veins, small congenital vascular malformations of venous predominance, bleeding varices, and large varices around an ulcer. The most common complications include hyperpigmentation, skin necrosis, pain, anaphylaxis, and matting.

The goals of operative treatment are the elimination of the hydrostatic forces of saphenous reflux, the removal of the hydrodynamic forces of perforating vein reflux, and the eradication of the varicosities in as cosmetic a manner as possible. Each case must be planned thoroughly because routine stripping of the greater saphenous vein from groin to ankle usually is not required. Patients with saphenous reflux should have groin-to-knee stripping. It is unnecessary to strip the below-the-knee portion of the vein unless it is varicose. Stab avulsion of vein clusters, which are marked preoperatively, supplements the stripping.

Removal of the greater saphenous vein requires its detachment from the common femoral vein and ligation of its tributaries at the saphenofemoral junction (Fig. 21-20). If an ankle incision is made, care must be taken to avoid injury to the saphenous nerve. Bleeding can be reduced by the use of a tourniquet and leg elevation during the stripping. The incisions made for stab avulsion of varices are 1 to 2 mm in length and are oriented in the skin lines. Ecchymosis is the most common complication after operations for varicose veins. The incidence can be reduced by carefully placed elastic support. Recurrences are usually due to incompetence at the groin or in the mid thigh from perforating veins.

Chronic Venous Insufficiency

Chronic venous insufficiency or the postthrombotic syndrome develops in approximately 50 percent of the patients with deep venous thrombosis. It is estimated that there are 500,000 patients in the United States with venous ulcers. Homans noted in 1917 that "overstretching of the vein walls and destruction of the valves upon which the mechanism principally depends bring about a degree of surface stasis which obviously interferes with the nutrition of the skin and subcutaneous tissues." It is now known that recanalization of the deep veins results in valvular incompetence, which in turn results in a long column of blood that transmits pressures of over 100 mmHg to the venules, causing the development of abnormal capillaries. These new vessels have an increased permeability to fibrinogen and red blood cells. Lymphocyte and macrophage recruitment occurs in response to extravasated protein. Pericapillary cuffing, an attempt by the endothelial cells to limit extravasation, occurs and results in widened endothelial gap junctions. The result is thickening and lipodermatosclerosis of the subcutaneous tissues that produce a characteristic "brawny" edema. The loss of red cells results in hemosiderin deposits, producing the characteristic pigmentation. A deterioration of mononuclear cell function is associated with chronic venous insufficiency. A decreased capacity for lymphocyte and monocyte proliferation in response to various challenges translates into poor or prolonged wound healing. When the distal perforating veins become incompetent, there is additional pressure, with skin atrophy leading ultimately to necrosis and chronic stasis ulceration (Fig. 21-21).

Diagnosis

The diagnosis of chronic venous insufficiency usually is made by inspecting the leg. Physical findings do not provide information about the presence, extent, or location of valvular incompetence or obstruction. Duplex scanning is the most reliable method of identifying valvular incompetence and venous obstruction.

Venous valvular incompetence is identified easily with the duplex scanner. The vein to be studied is identified with B-mode imaging. While the velocity spectrum is displayed, various maneuvers are performed to reverse the normal peripheral-to-central gradient. Retrograde flow is indicated by an inverted spectrum and a change in color from blue to red. Reflux is evaluated in the groin by having the patient perform a Valsalva maneuver. A period of reversed flow exceeding 1.5 s is considered abnormal. Manual compression is used above and below the vein in question. Some prefer to examine the veins for reflux while the patient is standing. A pneumatic cuff is placed at various levels beginning at 5 cm below the vein in question. The velocity spectrum is recorded continuously as the cuff is inflated and deflated. Normal valves close rapidly in response to temporary flow reversal. Perforating veins are studied with the patient in the reversed Trendelenburg position. Perforators are identified as veins arising from the superficial veins and penetrating the deep fascia into the muscular compartment. Outward flow with calf compression indicates valvular incompetence. Venous valves are identified with B-mode imaging. Normal valves are thin and mobile; diseased valves are shortened and thick, often with attached echogenic material.

Venous obstruction is identified with the same techniques used to diagnose acute venous thrombosis. After the vein is identified with B-mode imaging, its patency is assessed by its compressibility and Doppler spectrum. With partial occlusion or incomplete recanalization, there will be an encroachment on the flow image that no longer fills the entire vein. Collateral veins will be seen, and that finding is particularly useful in distinguishing between acute versus chronic occlusions. Another distinguishing feature of chronic occlusion is the shrunken size of the vein when compared to the distention seen in acute DVT.

The physiologic response to venous reflux can be measured. A needle is placed in a dorsal foot vein and secured in place. The patient's venous pressures can be determined in the resting and active state. The pressure in the standing position is slightly higher than the hydrostatic force of a column extending from the atrium to the foot. Resting pressures in patients with and without venous insufficiency are similar. In contrast to normal patients, who reduce their distal venous pressure with walking, patients with the postthrombotic syndrome gain no benefit from their muscle pump and their pressure increases (Fig. 21-22). If there has been failure of recanalization with persistent obstruction, the increase in blood flow with exercise may increase venous hypertension to produce ischemic pain referred to as venous claudication.

Nonoperative Management

The goals of treatment are to alleviate symptoms, heal ulcerations, and prevent ulcer recurrences. The vast majority of patients can be managed nonoperatively. Although the mechanism of benefit is unknown, compression therapy is the most important aspect of patient management. Conflicting hypotheses include a reduction in ambulatory venous pressure, improvements of the microcirculation of the skin and subcutaneous tissue, and increase in the pressure of the subcutaneous tissue that reduces the leakage of fluid from the capillaries. The latter is the most plausible because cutaneous metabolism may improve after fluid resorption, allowing an enhanced diffusion of oxygen and other nutrients.

The initial treatment of patients with venous ulceration should include a period of strict bed rest to reduce edema. Systemic antibiotics are given for the surrounding cellulitis. Elastic stockings are fitted when the edema has subsided. Surrounding areas of dermatitis are treated with topical steroids. Patients are then instructed to wear the elastic stockings for life. Two pairs are prescribed to allow for daily laundering of alternate pairs. Ulcer recurrence is 16 percent in compliant patients, but long-term compliance is difficult to achieve with patients who are reluctant to wear the stockings after their ulcer is healed. Some physicians prefer the paste gauze boot (the Unna boot) during the ulcer healing phase. This dressing contains calamine, zinc oxide, glycerin, sorbitol, gelatin, and magnesium aluminum silicate. Patients whose ulcers fail to heal after prolonged outpatient care require hospitalization.

Operative Management

Patients selected for operation have severe, disabling symptoms and a history of recurrent ulceration despite aggressive medical therapy. Candidates for operation should undergo ascending and descending venography in addition to duplex scans and ambulatory venous pressures. These tests provide data allowing an individualized treatment plan that addresses specific areas of obstruction or reflux.

Perforator Vein Ligation

Healing of chronic stasis ulcers is not likely unless the perforating veins responsible for the ulcer are identified and ligated. The typical location of these veins is posterior and superior to the medial malleolus. Ligation of the perforator vessels should be the initial procedure for recurrent ulceration. Treatment failure occurs in 10 percent of patients despite vigorous medical therapy, including support stockings, leg elevation, wound care, and patient education. These patients should be considered for venous reconstruction.

Venous Reconstruction

The present attitude of most surgeons toward venous reconstruction is critical and pessimistic. The venous system, unlike the arterial system, tends to recanalize, thus making it more difficult to quantitate the obstruction and identify the patient who may benefit from venous reconstruction. Dale estimated that the percentage of patients with chronic venous insufficiency who could benefit from reconstruction was 1 to 2 percent.

Primary valvular dysfunction can be treated by valvuloplasty. The valve most suitable to direct valve repair usually is the most proximal valve in the superficial femoral vein. After DVT, most patients have scarred and thickened valves that do not lend themselves to this type of reconstruction. Kistner, after studying 200 limbs with ascending and descending venography, found 28 that could be treated by valve repair, and 72 percent had an excellent result. In this procedure, floppy, incompetent valves are tethered against the vein wall or shortened using interrupted 8-0 monofilament suture (Fig. 21-23). Recent technical advances have allowed this procedure to be done under direct visualization using an angioscope. Results of this procedure are difficult to interpret because the operations are often combined with saphenous vein stripping and perforator ligation. Most investigators have reported improvement in symptoms for prolonged periods in approximately 60 to 80 percent of the patients.

Direct repair is not possible for postthrombotic valvular dysfunction. The two recognized options for surgical candidates are transposition of a deep femoral or saphenous vein valve or transplantation of a valve-bearing segment of the axillary vein to the superficial femoral or popliteal vein (Fig. 21-24). Results are not as good as those achieved by valvuloplasty for primary incompetence. Taheri and coworkers described 66 patients with good results in 78 percent. In this series, 31 patients had postoperative venograms, 28 of these were found to have valvular competence. A number of other investigators report symptomatic relief in 50 to 92 percent and ulcer recurrence in 6 to 54 percent of patients. Most of these patients had good results initially; however, at 1 year, a high proportion of the affected limbs had reverted to their preoperative condition. Bergan and colleagues have pointed out that for venous valve surgery to be successful, it usually must be accompanied by saphenous vein stripping and perforator ligation. The difficulty in identifying patients who could benefit from these procedures was put into perspective by Dale, who, after 2 years of investigating, failed to identify a group of patients who would benefit from venous valve transplantation or valvuloplasty. Husni found that venous reconstruction fails in three situations: when the bypass graft is too small in caliber, when venous hypertension is mild to moderate (less than 80 percent of the standing venous pressure), and when a thrombectomy or endophlebectomy has to be performed before anastomosis. In these patients who are at high risk for failure, he has recommended a distal arteriovenous fistula.

Approximately one-third of patients with chronic venous insufficiency have a predominant obstructive component because of inadequate recanalization after a DVT. Patients typically complain of swelling and pain on ambulation. The pain often is described as bursting, but patients with valvular incompetence refer to their pain as aching. Because the pain of chronic venous obstruction requires the patient to be off his or her feet to obtain relief, it is referred to as venous claudication. Ambulatory venous pressure measurements document the diagnosis, and venography identifies the site of obstruction.

Venous obstruction of the iliofemoral venous system can be bypassed by a saphenous vein cross-over graft, first described by Palma and Esperon in 1958. The procedure consists of isolating the normal contralateral saphenous vein and dividing it distally. The vein is tunneled suprapubically and anastomosed to the contralateral femoral vein, distal to its

obstruction. In 1982 Dale described 59 patients who had the Palma bypass with excellent results in 63 percent, good results in 17 percent, and a failure rate of 20 percent. Husni, in 1983, and Smith and Trimble, in 1977, reported similar results. The saphenous vein cross- over graft generally has been accepted as useful; however, the natural history of iliac vein occlusion is recanalization, and very few patients with iliofemoral thrombosis become candidates for surgery.

Use of the saphenous vein for popliteal-to-femoral vein bypass was described in 1954 by Warren and Thayer with good results in 10 of 14 patients. This operation provides the muscle pump system a means of emptying the calf by bypassing the occluded superficial femoral vein. With rich collateral veins in the thigh, identifying the patient with an obstructed superficial femoral vein who might benefit from the saphenous-to-popliteal vein bypass is difficult. The saphenous vein is dissected free below the knee and anastomosed to the popliteal vein, which is obstructed proximally. Husni has popularized this procedure and has reported the outcome in 27 patients with a good result in 63 percent. Dale reported good results in 10 patients (60 percent), and Smith and Trimble, in a collected series of 59 patients, reported good results in 76 percent.

The majority of iliofemoral thromboses occur on the left side. This is attributed to the right iliac artery compressing the left iliac vein as it crosses the fifth lumbar vertebra. Various autopsy series and operative studies have documented the presence of left iliac vein webs and scarring in patients who have had iliofemoral thrombosis. There was interest in this problem in the 1960s by Calnan and associates and by Cockett and Thomas, who advocated surgical correction of these lesions. Dale reviewed eight such patients identified by venography and subsequently operated on four, trimming out anterior webs or scar tissue and using a venous patch for closure. Two of the patients had excellent results; edema developed later in one patient, and a fourth patient had a complicated postoperative course, complaining of excruciating pain and postoperative swelling. Dale recommended operation only for the patient whose symptoms are severe and who accepts the operation with the understanding that the results are not predictable. Smith and Trimble have followed 30 patients with this problem and have operated on 14, with an 85 percent postoperative improvement rate. Cockett and Thomas, conversely, found the results unsatisfactory, and after operating on 30 patients using several different methods they recommended abandoning the procedure.

Venous Trauma

Venous injuries of the extremities usually are associated with arterial injuries because of their anatomic proximity. Direct ligation of injured superficial veins is appropriate treatment except when they are the sole remaining venous drainage of the extremity, which mandates their repair. Treatment of injuries of the deep veins changed dramatically as a result of the military experience in Korea and Vietnam, as reported by Rich and associates. It was well demonstrated that ligation of major extremity veins resulted in higher rates of disability and limb loss than repair or replacement by autogenous vein segments. The concept of primary repair of venous injuries by suture vein patch or vein graft interposition has been extended to civilian injuries by Agarwal and associates with favorable results. The rationale for primary repair is based in part on the adverse hemodynamic effects in the first 72 h after major venous ligation. Current recommendations are that repair of the common femoral and popliteal veins should be done whenever possible. Repair of the superficial femoral vein is controversial.

LYMPHATICS AND LYMPHEDEMA

Anatomy and Physiology

The exact origin of lymphatic vessels is a matter of disagreement among embryologists. The original theory of Sabin traced the origin from the venous system, while Huntington and McClure suggested that lymphatics form by fusion of mesenchymal spaces or clefts, which has been labeled the centripetal theory. By the sixth week of gestation, there are paired lymph sacs in the neck and lumbar areas, and at the eighth week, there is a retroperitoneal lymph sac with a developing cisterna chyli. These systems develop communicating channels that ultimately form the thoracic duct by merger of the right lymphatic duct with the left across the fourth to sixth thoracic vertebrae, which then drains into the left subclavian vein. Smaller lymphatic ducts persist that drain into the right subclavian vein.

Developmental arrest or abnormalities can result in primary hypoplasia or absence of ducts and lymph nodes. Abnormal growth of jugular lymph sacs can produce unilocular or multilocular lymph cysts termed cystic hygromas. These cysts also may be found in the axilla, mediastinum, retroperitoneum, or intestinal mesentery. Hyperplastic changes may occur to produce lymphangiomas with or without other vascular malformations.

The function of the lymphatic system begins with lymphatic capillaries, which collect fluid and protein from the extravascular spaces. In addition to the protein that cannot be reabsorbed by the venules, red blood cells, bacteria, and other larger particles can be evacuated only through the lymphatics. This permeability is facilitated by the absence of a basement membrane beneath the lymphatic endothelial cells. The lymphatic capillaries are found beneath the epidermis in the superficial dermis. These vessels drain into valved channels in the deep dermis and subdermal tissues, forming larger channels that follow the vascular pathways superficial to the deep fascia. Although lymphatics can be found in the intermuscular fascia, they are absent in muscles, tendon, cartilage, brain, and cornea.

Lymph is transported by afferent vessels to regional lymph nodes that vary in size according to their function and activity. Within the medullary sinuses of the node, circulating lymphocytes are replaced and initial contact of foreign material with the immune system is made. Efferent lymph leaves the node via hilar channels, which are less numerous than the afferent channels that enter the convex side of the node. In addition to direct thoracic duct drainage into the subclavian vein, there are other lymphovenous communications within nodes and in peripheral vessels. Central lymphatic flow is promoted by the lymphatic valves, muscular contractions in larger ducts, respiration, arterial pulsation, and external massage. The main function of the lymphatic system is to clear the interstitial spaces of excess water and particulate matter.

Lymphedema

Classification of Lymphedema

The original classification of Allen was into two types, one in which there was no known cause, and the other secondary to a known disease or disorder. The primary lymphedemas were called *congenital* when present at birth and *praecox* when there was onset in childhood. Kinmonth added the term *tarda* for when the onset was not until later life. With the advent of lymphography it became possible to classify the primary lymphedemas structurally into hyperplasias and hypoplasias. The original classification as proposed by Kinmonth has largely been abandoned for the more simplified version presented in Table [21-6](#).

The congenital lymphedemas are hypoplastic in 92 percent of cases. Their subgroups are defined by lymphography and behave differently. Those with distal hypoplasia have a mild, nonprogressive form of the disorder provided that their proximal pathways are normal. Most of these patients are women and notice the onset after puberty. In proximal hypoplasia, the lymphedema is more extensive, involving the entire extremity, and it occurs equally among males and females. The combination of proximal and distal hypoplasia shows features of both groups and tends to be progressive.

The primary hyperplastic lymphedemas are uncommon (8 percent), and those with bilateral hyperplasia usually can be recognized by diffuse capillary angiomas on the lateral sides of the feet. Lymphography shows dilated lymphatics with normal valves, in contrast to the findings in the megalymphatic group, in which no valves can be seen. In this latter group, chylous reflux may produce chylometorrhea, skin vesicles, or chyluria.

The most common cause of secondary lymphedema in this country is malignant disease metastatic to lymph nodes. Surgical removal of nodes, especially when combined with radiation therapy that produces lymphatic fibrosis, is another common cause. In tropical and subtropical countries, filariasis is the most common cause of secondary lymphedema, producing the typical appearance of elephantiasis. Other infective and chemical agents, such as silica, can enter the lymphatic system via barefoot walking and cause fibrosis of lymphatics and lymph nodes.

Clinical Manifestations

Lymphedema is a clinical diagnosis and should be restricted to situations where other causes of edema have been excluded or a specific lymphatic abnormality has been demonstrated. The presence of bilateral dependent “pitting” edema usually indicates a renal or cardiac etiology. Other generalized hypoproteinemias may be seen in malnutrition, cirrhosis, and protein-losing enteropathy, or they may be idiopathic. Allergies or hereditary causes are unusual. In unilateral edema, venous disease is the most likely cause and can be recognized by the examinations described in the previous section.

The patient with lymphedema complains of swelling and fatigue. Limb size increases during the day and decreases at night but is never normal. It is important to determine whether there is a family history of primary lymphedema and whether the patient has visited any countries where filariasis is endemic. The presence of weight loss and diarrhea suggests small bowel lymphangiectasia. On examination, lymphedema is characteristically firm and rubbery but nonpitting. Lymph vesicles may be present containing fluid of high protein concentration. Complications of lymphedema such as infection, cellulitis, erythema, and hyperkeratosis may be present. It is important to document limb size to identify isolated limb gigantism and the Klippel-Trénaunay syndrome that may have hypoplastic lymphatics in addition to venous abnormalities, capillary nevus, and limb elongation. The patient should be examined for upper extremity and genital lymphedema, hydroceles, and amelogenesis imperfecta.

Diagnostic Studies

Dye Injection

Lymphatics can be visualized by dye injection in the extremities and mesentery, and also by ingestion of cream or milk to visualize intestinal lacteals and major ducts.

A highly diffusible dye such as patent blue, introduced by Hudack and McMaster, or sky blue dye, recommended by Butcher and Hoover, can be injected in 0.2 mL amounts subcutaneously into each interdigital web. Massage of the skin and movement of the joints usually defines a network of fine intradermal lymphatics (Fig. [21-25](#)). If the collecting vessels are obstructed or inadequate, the dye diffuses through the dermal lymphatics to produce a marbled appearance called “dermal backflow.”

Lymphography

The technique of lymphography was developed by Kinmonth, who demonstrated that it was possible to cannulate the lymphatics visualized by dye injection and then inject contrast medium (Lipiodol). This is a meticulous and tedious procedure that may require general anesthesia, as proposed by Kinmonth. If the lymphatics in the foot are not usable, it is possible to cannulate lymphatics adjacent to groin nodes or to inject the node directly. With adequate visualization, the lymphatics in the extremity will be identified, often as parallel tracks that are of uniform size and bifurcate as they proceed proximally in contrast to the venous system. Normally, there is some dilatation at the level of the valves.

Radionuclide Lymphatic Clearance

Radionuclide scanning using human serum albumin labeled with radioactive iodine or technetium 99m colloid has been used to monitor lymphatic clearance by serial scanning. Although the technique is simpler than standard lymphography, it has major disadvantages because of the haziness of the scan, radiation dosage, and distribution of the radionuclide into the extracellular fluid, making calculations of clearance dependent on leg volume.

Analysis of Tissue Fluid

Tissue fluid or lymph can be aspirated or collected from a tube in the subcutaneous tissues but contributes little to the diagnosis of lymphedema. Characteristically, lymphedema fluid has a protein content of more than 1.5 g/dL, in contrast to that of edema fluid from venous hypertension, which usually is less. The ratio of albumin to globulin also is higher in lymphedema fluid than in plasma, which is helpful in the presence of an inflammatory exudate in which the protein content is high but the albumin-to-globulin ratio is normal.

Management

Supportive Treatment

There are significant anatomic and physiologic limitations to the treatment of lymphedema. From the standpoint of physiology, the removal of fluid is not as effective as in edema of other causes because of the residual protein in lymphedema. In addition, from an anatomic standpoint, the development of fibrosis produces irreversible changes in the subcutaneous tissues. Therefore, the options are limited and the primary objectives remain for control of edema, maintenance of healthy skin, and avoidance of the complications of cellulitis and lymphangitis.

The initial objective of control of edema can be approached by elevation and the use of sequential pneumatic compression boots to massage the leg. These treatments can be done at home with equipment rented for this purpose. Once the leg has reached optimal size, the patient should be fitted with firm elastic stockings as described earlier for venous insufficiency. The stockings should be removed at night and the foot of the bed elevated to maintain the pressure gradient from leg to right atrium. Massage therapy was used in the treatment of lymphedema as early as 1882 and has current advocates.

The onset of redness, pain, and swelling usually signifies early cellulitis or lymphangitis, which can be recognized by red streaking up the leg. The usual causative organism is staphylococcus or beta-hemolytic streptococcus, which must be treated vigorously, usually with intravenous antibiotics. In the absence of treatment, the infection may obliterate more lymphatics and produce constitutional signs of fever, malaise, nausea, and vomiting. Another frequent complication is eczema, which usually will respond to hydrocortisone cream. Antifungal agents may be necessary, topically and systemically, for chronic infections, particularly between the toes. In contrast to the stasis edema of venous insufficiency, ulceration is unusual, although fissures and lymph fistulas can develop and require surgical excision.

The secondary lymphedemas may lend themselves to treatment of the underlying disorder, such as using diethylcarbamazine for filariasis or appropriate antibiotics for tuberculosis or lymphogranuloma venereum. In rare cases of long-standing secondary lymphedema, such as in the arm after radical mastectomy, a lymphangiosarcoma may develop, appearing as a raised blue or reddish nodule. Satellite tumors and early metastases may develop if it is not recognized and widely excised.

Operative Treatment

Only 15 percent of patients with primary lymphedema become candidates for operative treatment, which usually is directed to reducing leg size. The indications for operation are related to functional rather than cosmetic improvement, because the appearance of the extremity even after a successful procedure will still be abnormal and show extensive scarring. The best results are obtained when the bulk of the extremity has severely impaired movement or when there have been recurrent attacks of cellulitis. Although some efforts have been made to develop techniques to improve lymphatic drainage, most of the established procedures consist of excisional operations.

Three of the excisional procedures were based on the incorrect assumption that the deep fascia acted as a barrier to lymphatic drainage, and the efforts of Kondoleon and associates to excise fascia or insert a dermal flap into muscle proved ineffective in improving lymphatic drainage. The original procedure devised by Charles consisting of wide excision of lymphedematous tissue followed by skin grafting still is useful when the overlying skin is in poor condition, as in elephantiasis. The procedure used most often, however, is Kinmonth's modification of Homans's procedure, in which skin flaps are raised to allow excision of the underlying subcutaneous tissues.

The most logical, albeit technically demanding, approach has been directed to establishing lymphaticovenous anastomoses. Initial efforts in this area were made by Nielubowicz and Olszewski, who divided a lymph node, removed the pulp under magnification, and sutured the node capsule with its afferent lymphatics into a vein. This procedure is more suitable for secondary lymphedema than primary, in which the disorder lies in the lymphatic channels themselves. Another promising technique of direct lymphovenous connection was developed by Cordeiro and modified by Degni, who used a special needle for insertion of lymphatic vessels directly into veins and fixed them there by a single suture. Using this technique, Fox and associates treated eight secondary and 12 primary lymphedema patients, with follow-up as long as 4 years. Good results were obtained in two of four postmastectomy lymphedemas, with poor results in the other two, who had postoperative lymphangitis. Nine of 11 patients with primary lymphedemas had good functional results, allowing them to resume normal activity. The authors recommend long-term preoperative anti-inflammatory and antimicrobial therapy to avoid postoperative lymphangitis.

It is difficult to evaluate the results of such procedures when combined with resectional operations and in the absence of postoperative lymphography to demonstrate patency of the anastomoses. However, the deleterious effects of lymphangiographic contrast on lymphatics were well demonstrated by O'Brien and associates, who measured limb volume after lymphangiography in 100 patients and found that 32 percent had a significant increase in leg volume and 19 percent developed lymphangitis. Therefore, it seems advisable to use lymphangiography only for diagnostic studies and not for pre- or postoperative evaluation until safer contrast material becomes available. Additional efforts to combine resectional operations with microlymphovenous anastomoses as reported by O'Brien and Shafiroff may offer some brighter prospects for improvement of these debilitating disorders.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 22 - Manifestations of Gastrointestinal Disease

Josef E. Fischer
Michael S. Nussbaum
William T. Chance
Frederick Luchette

Pain

Pain (from the Latin *poena*, meaning penalty, punishment, torment) is the most common symptom of gastrointestinal disease and leads to surgical intervention more than any other gastrointestinal manifestation. Most disturbances in gastrointestinal function and diseases of the abdominal viscera are associated with pain at some time during their course.

General Considerations

Although there is no unequivocal proof that each peripheral nerve fiber is devoted to one type of sensory modality (pain, touch, cold, or warmth), most physiologists subscribe to the *specificity theory*, which contends that pain is a separate sensory modality with its own specific neural apparatus. Melzack and Wall, in 1965, proposed the *gate-control theory*, which suggested that the amount and quality of perceived pain are determined by many physiologic and psychologic variables. Modulation of nociceptive impulses occurs at the dorsal horn and various levels of the ascending afferent systems. The gate-control theory has been the basis for methods of pain-inhibiting electrical stimulation, e.g., with transcutaneous nerve stimulation (TNS).

Appropriate stimuli initiate impulses in skin, muscle, or viscera. These sensory impulses are transmitted to the posterior horn of the spinal cord in the primary sensory neuron that has its cell body in the dorsal root ganglion. The secondary sensory neuron in the posterior horn transmits the impulses in the contralateral spinothalamic tract to the posterolateral nucleus of the thalamus. The tertiary sensory neuron transmits the impulses from the thalamus to the postcentral gyrus of the cerebral cortex (Fig. 22-1). Three kinds of pain have been described: superficial, or cutaneous pain; deep pain from muscles, tendons, joints, and fascia; and visceral pain. The first two can be combined in the category of somatic pain.

Pain Pathways

Until well into the twentieth century, the viscera were thought to be completely insensitive. William Harvey noted when observing a boy who had sustained a chest wound leaving the heart exposed that touching the heart caused no sensation. In 1901 Lennander stated that the viscera are insensitive to cutting, crushing, and burning and that only traction or irritation of the parietal peritoneum could cause abdominal pain. In 1911 Hurst demonstrated that distention of any hollow viscus is painful. Ryle amplified this by emphasizing that contraction of smooth muscle of hollow viscera is adequate to cause pain.

The current and preferable terminology is *visceral afferents* to denote all afferent fibers from the viscera, including those that give rise to visceral reflexes as well as those that subserve pain. The terms *autonomic*, *sympathetic*, and *parasympathetic* are reserved for the visceral efferent fibers.

Pain impulses from the abdominal cavity reach the central nervous system by three routes: from the viscera via visceral afferents that travel with the (1) sympathetic and (2) parasympathetic nerves; and from the parietal peritoneum, body wall, diaphragm, and root of mesenteries via somatic afferents that travel in the (3) segmental spinal nerves or phrenic nerves. Primary sensory neurons for pain, both visceral and somatic, are mostly small (1 to 2 mm) unmyelinated fibers, but with some small (3 to 4 mm) myelinated fibers (Gasser classes C and A delta, or groups IV and III in Lloyd's terminology).

The route of a typical afferent from an abdominal viscus is as follows: The axons of nerve endings in the wall of the viscus follow the artery to the aorta and then through the collateral sympathetic ganglion without synapsing. They then enter the splanchnic nerve, traverse the paravertebral sympathetic ganglion, again without synapsing, and join the spinal nerve via the white ramus communicans. The cell body of this primary visceral afferent neuron is located in the spinal ganglion from which central processes are sent to the dorsal horn of the spinal cord via the dorsal root.

The central processes of the primary sensory neuron synapse with at least three distinct spinal tracts: (1) secondary pain neurons whose axons ascent for two or three segments and then cross in the anterior commissure to the anterolateral spinothalamic tract; (2) secondary sensory neurons whose processes ascend the posterior columns; and (3) many small neurons in the substantia gelatinosa that contribute to the tract of Lissauer. As the secondary sensory neurons ascend, collaterals are given off to the reticular substance, forming the core of the brainstem, and to the hypothalamus. The role of these extraspinothalamic pathways is uncertain; they may be alternative pain pathways, may inhibit central pain responses, or may be involved in the affective aspects of pain. Sensory tracts synapse in the thalamus with tertiary neurons that project to the cortex. The cortical areas involved with pain are not completely known. Stimulation of various areas of the postcentral gyrus in the conscious human being produces contralateral paresthesia of small areas of the body, but even total hemispherectomy does not consistently abolish pain, though localization is defective. The functions of the thalamus in human beings, in contrast to animals, are largely expressed through the cortex, but one function that may have been retained in the evolutionary process is the expression of the affective aspects of sensation. Affectivity, pleasantness and unpleasantness, is considered a primitive function that has remained at the thalamic level despite development of the cerebral cortex.

Though the thalamus and frontal cortex are the principal areas of the brain involved with pain, they cannot be considered *the* brain centers, since the hypothalamus, limbic system, brainstem reticular formation, and parietal cortex also are involved in pain reception. Opiate receptors exist in the brainstem and spinal cord, and naturally occurring opiates are found in the brain and pituitary gland; these are collectively called *endorphins*. Pain relief from low-frequency electrical stimulation exerts its effect by increasing these endorphin levels, an effect that can be blocked by the opiate antagonist naloxone.

Abdominal Pain

Visceral Pain

The visceral peritoneum has no somatic pain fibers and is innervated only (inadequately, at best) by autonomic C-type fibers. C fibers are slow transmitters; the sensation reported is dull, poorly localized pain, gradual in onset and longer in duration. These fibers are normally insensitive to stimuli that produce pain when applied to the skin, but respond to a limited number of noxious stimuli. These include changes in diameter of a hollow viscus, including distention and tension, e.g., rhythmic tension that is perceived as cramping. Spastic contraction, stretching of the capsules of solid viscera, ischemia, and certain chemicals also produce visceral pain. This perception of pain often is so poorly localized that it is of only limited value in determining the nature of the disturbance.

The role of chemical substances in visceral pain is not clear. Experimentally, pain can be produced by the intraarterial injection of acid, alkaline, or hypertonic solutions (lactate, potassium ions, or bradykinin). Potassium released from cells by injury or ischemia has long been known to be pain-producing, and some suggest that the release of intracellular potassium ions may be the actual physiologic stimulus for pain. Some pain receptors have been classified as chemoreceptors, and the pain of ischemia has been attributed to increasing concentrations of hydrogen ions. The pain of inflammation is thought to be caused by the accrual of algescic bradykinin peptides that activate pain receptors more or less selectively. The bradykinin effect is facilitated in the presence of prostaglandins, and it is this mechanism that is inhibited by aspirin.

Visceral pain is diffuse and poorly localized, has a high threshold, and exhibits an exceedingly slow rate of adaptation. The high threshold and poor localization are partially attributable to the relatively sparse distribution of sensory endings in the viscera. The pain is felt by the patient to be “deep” in those cutaneous areas or zones that roughly correspond to the segmental distribution of somatic sensory fibers that originate from the same segments of the spinal cord as the visceral afferent fibers from the viscus in question.

With severe visceral or deep somatic pain, concomitant responses, presumably due to autonomic reflexes, may be prominent. These include sweating, nausea (sometimes with vomiting), tachycardia or bradycardia, fall in blood pressure, cutaneous hyperalgesia, hyperesthesia or tenderness, and involuntary spastic contractions of abdominal wall musculature. The muscular rigidity accompanying severe pain is most marked when the body wall is involved by the pain-inciting lesion, e.g., the boardlike rigidity associated with perforated peptic ulcer. The distribution is regional rather than segmental, and it involves sustained reflexes in several segmental nerves. Maintained muscular rigidity may become painful, so that occasionally the deep muscular tenderness outlasts and outweighs the original visceral pain.

The parietal peritoneum and the abdominal wall are innervated by somatic pain fibers. These A-delta fibers are distributed not only to the parietal peritoneum but also to muscle and skin, and they transmit sharp, well-localized pain sensations of limited duration to the spinal nerves. Pain from these locations yields information that localizes well and that may vary in character depending on stimulus. Zachary Cope first emphasized that the pelvic peritoneum has no significant pain fibers, and significant disease can occur in the true pelvis without pain detection.

Unreferred Visceral Pain

The location of the noxious stimuli from various portions of the gastrointestinal tract influences the type of pain reported. Abdominal viscera are innervated by afferent autonomic innervation bilaterally, and visceral pain usually is perceived as midline. Exceptions include the kidneys, ureters, cecum, and ascending and descending sigmoid colon, which are unilaterally innervated. The pain is poorly localized because most viscera contain multisegmental innervation and few nerve endings compared to the skin. The characteristics of the pain are gnawing, cramping, and associated with nausea, sweating, vomiting, perspiration, and pallor. When the nociceptive stimulus is intense, the pain also is referred to the skin and deeper somatic tissue. The midline location is the result of the embryologic development of the gut. Midline (epigastric, periumbilical, and hypogastric) pain is caused by a wide variety of diseases, each of which is localized to its corresponding dermatome on the basis of embryologic development. Since the gastrointestinal tract comprises a foregut, midgut, hindgut, and cloaca, each develops its own separate embryologic blood supply and innervation.

Epigastric pain is a characteristic of a foregut disturbance. The foregut is that part of the digestive tube extending from the oropharynx to the duodenum, generally terminating at the entrance of the common bile duct at the junction of the second and third portions of the duodenum, and it includes the liver, biliary tree, pancreas, and spleen. Midepigastric pain also can result from pylorospasm, a nonspecific reflex response to any disturbance in the abdominal viscera, from appendicitis to the prodrome of epigastric pain that may be a harbinger of carcinoma of the colon. Midgut pain is manifested as periumbilical pain. The midgut comprises the distal duodenum, jejunum, ileum, appendix, and ascending colon as well as the proximal transverse colon vascularized by the superior mesenteric artery. Lower abdominal or hypogastric midline pain, especially cramping pain, reported as transverse across the lower abdomen, is a manifestation of hindgut pain. The hindgut comprises the remainder of the colon and rectum down to the cloaca, and endoderm-lined cavity giving rise to the anorectal canal, and is in direct contact with the surface ectoderm at the level of the dentate line.

Suprainguinal pain is an important and highly reliable differential sign. It often is seen in the female and rarely in the male, and it almost always indicates gynecologic inflammation or disease. It sometimes is associated with hyperesthesia immediately above the inguinal ligament. Pain in this location probably is a result of stimulation of the parietal peritoneum

just above the true pelvis, because the parietal peritoneum within the true pelvis transmits no pain sensation. With a purulent inflammatory process involving the fallopian tubes or the uterus, pain and tenderness perceived in the suprainguinal area usually are present.

Referred Visceral Pain

This is more localized than true visceral pain. It is localized to the dermatomes and myotomes that are supplied by the same spinal cord segments as the affected viscus. For example, distention of the intestine with a balloon causes a vague, aching, poorly localized discomfort at first, but with greater distention the pain is referred to the abdominal wall and the back.

Parietal Pain

Unreferred Parietal Pain

Unreferred or local parietal pain occurs when the inflammation of the viscus involves the adjacent nociceptive fibers of the parietal peritoneum. The pain is localized in the body wall directly over the site of the inflammation, such as in the localized pain of acute appendicitis at McBurney's point.

Referred Parietal Pain

This is characterized by pain felt in an area remote from the site of the nociceptive stimulation. A classic example is Kehr's sign (left shoulder pain) with traumatic rupture of the spleen.

Perception of Stimulation

Stimulation of the visceral peritoneum usually does not result in pain in a normal sense; only distention, contraction, torsion (probably as a result of ischemia), and occasionally certain chemicals or other inflammatory agents result in pain. The neural receptors for the transmission of these perceived noxious stimuli are supplied by autonomic nerves (sympathetic and parasympathetic) and are located within the walls of the hollow organs or on the serosal structures, such as the capsules of solid organs. Visceral pain may or may not signal the need for surgical therapy. The pain of intestinal ischemia, for example, perhaps perceived by sustained contraction of the intestinal wall musculature and intramesenteric receptors associated with large mesenteric vessels, signals the need for surgical intervention, whereas gastroenteritis with severe cramping does not.

The response of the visceral receptors also reflects the state of the adjacent serosa and mucosa. A good example is the gastric mucosa; acid or food touching a noninflamed gastric mucosa does not result in pain. If the mucosa becomes inflamed, however, as in alkaline reflux gastritis, food ingestion results in significant pain.

Pain arising from the parietal peritoneum is mediated principally by somatic nerve fibers that transmit through the posterior tracts of the spinal cord. Even if fibers cross, pain can be localized and usually is limited at least to one of the four quadrants, and often is considerably more focused. These nerve endings are stimulated by chemical irritants, changes in pH, inflammatory mediators, and probably by local transudation of bacteria and products of polymorphonuclear leukocyte destruction. When a somatic component is added to visceral pain it is likely that surgical intervention will be necessary.

One type of pain associated with gastrointestinal disease is a direct result of neural infiltration. The back pain associated with pancreatic cancer, or infiltration of the celiac axis and left gastric nodes by gastric cancer are examples of severe, unrelenting pain resulting from direct invasion of pain-transmitting nerves. This type of pain makes operative intervention likely, if only for pain relief. In pancreatic cancer severe, persistent back pain usually indicates unresectable disease, because of the size of the tumor, invasion of adjacent structures, or lymph node metastases.

The various causes of gastrointestinal, intraperitoneal, and abdominal pain are outlined in Table 22-1. Pain of intraabdominal origin may emanate from the peritoneum, hollow intestinal viscera, solid viscera, mesentery, or pelvic organs and can be caused by inflammation, mechanical processes such as obstruction and acute distention, and vascular disturbances. The extraperitoneal causes of abdominal pain are listed in Table 22-2. The intrathoracic causes of upper abdominal pain are confusing, because their segmental distribution is similar to that of the referred pain from abdominal pathologic conditions, but they usually are not accompanied by tenderness or muscular rigidity. Inflammation of peripheral nerves, as in herpes zoster, may be painful before the lesion becomes apparent.

Referred Pain

Referred pain is the result of afferent neurons that innervate two entirely separate, anatomically distinct structures that have a common embryologic origin. Three common examples of referred pain in gastrointestinal disease are the fourth cervical nerve route, the sixth to eighth thoracic nerves, and the tenth thoracic nerve route. The parietal peritoneum of the diaphragm, the area around the shoulder, and the supraclavicular hollow are innervated by C4. Pain under the left diaphragm is manifested as left-shoulder pain (Kehr's sign) and occurs secondary to a ruptured spleen or pancreatic inflammation. Right supraclavicular pain can be caused by blood or bile under the right hemidiaphragm. The thoracic afferents of T6–T8 that innervate the right subscapular area also innervate the biliary tree, liver, and peripancreatic area. The pain at the tenth thoracic nerve route usually is the result of irritation to the kidney or ureter and is usually manifested as flank and genital pain, and classically as testicular pain in the male. Pain referred to the labia in the female is far less common but may occur with renal colic. Possible sources of referred pain are listed in Table 22-3.

Acute Abdominal Pain Clinical Manifestations

The nature of the onset of pain is critical to the diagnosis (Table 22-4). Sudden onset of abdominal pain is more likely to result in operative intervention, because it usually is the result of perforation, volvulus, passage of a stone or other obstructing material in a hollow viscus, or embolization in the arterial supply of a hollow viscus. Gradual onset of pain usually indicates an inflammatory process or a slower progressive obstructive process. The relationship of other symptoms to the onset of pain also is important to the diagnosis. Acute onset of pain followed shortly by nausea and vomiting suggests high intestinal obstruction following a volvulus. Nausea and vomiting 6 to 12 h after the onset of pain suggests a site of obstruction lower in the gastrointestinal tract, and pain of even longer duration suggests colonic obstruction secondary to neoplasm, diverticulitis, or volvulus. Inflammation in the foregut or midgut often is followed by anorexia, nausea, or vomiting because of the sensory afferents that are carried by vagal fibers, and hindgut inflammation usually is not associated with these symptoms unless complicated by luminal obstruction. Conversely, anorexia, nausea, and vomiting that precede the onset of pain generally indicate a disease process that does not require operative intervention.

The character of the pain also is important. The patient may describe pain as pressure, sharp pain, dull steady pain, or crampy pain. Each suggests a different type of disease process. Epigastric pressure, particularly in the upper epigastrium, raises the suspicion of myocardial ischemia, reflux esophagitis, gastric distention, or inflammation. Colicky pain in the right upper quadrant radiating to the subscapular area suggests a stone in the cystic or common duct; back and flank colicky pain radiating to the labia or to the testicle in the male suggests a stone in the ureters (Fig. 22-2).

It is important to note what aggravates the pain (Table 22-5)—posture, bodily functions such as eating, moving the bowels, passing gas, vomiting, etc. Patients with significant peritoneal irritation do not move; they lie still, legs flexed, to avoid aggravating the pain by sudden movement. Patients with ureteral colic often writhe, moving around more with the exacerbation of pain.

The progression of the pain with time is important; is the pain becoming more severe, or is it subsiding? Referral of the pain might give a clue to the anatomic area and the nature of the process. Changes in location and severity with time also are important. The relationship between the onset of abdominal pain, especially lower abdominal pain, and the menstrual cycle should be noted. Previous episodes or whether any family members have had similar episodes should be determined. The nature of the medications taken by the patient, including any medications taken in an attempt to relieve the pain, should be elicited. Patients often do not think of acetaminophen or aspirin, for example, as medications that should be reported.

Different inflammatory processes occur in different age groups. Acute colicky pain is suggestive of intussusception in young children. Cholecystitis is most commonly seen in the 20-to-40-year age group, and diverticulitis rarely occurs before age 35, although it is increasingly seen in younger patients.

Physical Examination

Probably the most important aspect of the physical examination is a general visual assessment. Is the patient sick? What is the breathing pattern? Are the nares flaring? Is the patient experiencing tachypnea? Is the facies drawn, tense, with sunken eyeballs, implying dehydration or acute volume loss? The patient's chest and abdomen should be undraped for observation. Slanting light from across the bed or examination table may reveal rib retraction or the splinting of one side of the chest. Right upper quadrant movement may be inhibited in acute cholecystitis. The patient's response to pain is an important clue. Is the patient still or writhing? In breathing, do both sides expand, or is there limitation on one side, suggesting peritoneal irritation?

Inspection

A scaphoid abdomen is unlikely to harbor significant acute disease. A diseased abdomen is most likely to be distended. The abdomen should be inspected in indirect light for bulges or other irregularities in the abdominal wall pattern. In thin individuals it is possible to see a ladder pattern if there is intestinal obstruction or masses indicating neoplastic disease, volvulus, etc. The ventral, inguinal, and femoral regions should be evaluated for the presence of hernias.

Examination of the chest anteriorly and posteriorly in a good oblique light enables the physician to evaluate movement of the ribs. The chest should be inspected for movement of equal excursions on a deep breath (Fig. 22-3). There should be an attempt to elicit tenderness in the supraclavicular area (C4) or subscapular area (T6–8), in the area where the pain is referred. If the pain is referred from a thoracic structure, there will be no tenderness, but occasionally there is hyperesthesia. Additional examination should be performed for pleural fluid, tactile fremitus, or egophony. Percussion of diaphragmatic movement is performed. A normal diaphragmatic excursion, even in a smoker, is 4 to 6 cm. In pneumonitis, referred abdominal pain may mimic cholecystitis or pancreatitis (see Table 22-3).

Auscultation

Auscultation of the lungs should be done to identify any crepitus in the lung fields or mediastinum, rales, or bronchial breathing. Auscultation of the abdomen is performed before percussion or palpation to avoid any alteration in peristalsis by palpation. Stethoscopes are not all equally suitable for abdominal auscultation. Stethoscopes designed for cardiac auscultation, for example, are not ideal for evaluation of bowel sounds. Stethoscopes should be warmed before use. The stethoscope can be used to evaluate tenderness because the patient is distracted.

Examination should determine whether bowel sounds are present. If bowel sounds are not heard within 1 min over various parts of the abdomen, the conclusion is that the patient has "no bowel sounds," which usually means profound ileus, hypokalemia, hypomagnesemia, narcotic overdose, peritonitis, or mesenteric thrombosis. If bowel sounds are heard, are they hyperactive or hypoactive? Hypoactive bowel sounds may result from hypokalemia, inflammation, and ischemic bowel. In the early stages of diverticulitis, when obstruction

is prominent and peritonitis is not, bowel sounds tend to be hyperactive. After arterial thrombosis or embolus bowel sounds are very hyperactive, and they become hypoactive as bowel viability is impaired. Bowel sounds also may be correlated with cramps, nausea, or vomiting.

Percussion

Percussion is less important in abdominal evaluation, but it may be useful in identifying areas of localized or referred rebound pain. Percussion distinguishes between gas and fluid causes of distention. Ascites is determined by percussing areas of dullness, and shifting dullness, with changes in patient position. A “fluid wave” is elicited by pressing vertically on the midabdomen and tapping the flanks. If there is suspicion of ascites, the stigmata of liver disease or neoplastic disease, including a prominent abdominal venous pattern, umbilical hernia, caput medusa, palmar erythema, spider angiomas, Dupuytren’s contracture, testicular atrophy, rhinophyma, etc., are confirmatory.

Palpation

Palpation of the abdomen should be directed by the history as well as the physical findings noted on inspection, auscultation, and percussion. Examination of the abdomen should proceed in a systematic fashion. A light touch is essential. Palpation should include all sites of possible hernia—inguinal, femoral, or umbilical. If an acute inflammatory or other process is suspected, examination should begin away from the pain and the painful quadrant examined last. Once a painful or tender area is palpated, the abdominal wall may spasm, making additional information difficult to obtain. If the significance of the tenderness is uncertain, the patient should be distracted while palpation continues.

Rebound tenderness indicates inflammation of the parietes of the abdominal wall by an adjacent viscus. Rebound tenderness is subjective. Rebound tenderness also may be elicited by having the patient breathe deeply, cough, or by striking the bed suddenly. Indirect rebound tenderness—palpating another area of the abdomen and letting go, with increased pain in another quadrant—usually indicates significant intraabdominal surgical disease. In advanced stages of peritonitis, “boardlike rigidity” can occur. This regularly occurs in patients with perforated peptic ulcer. The rigidity of perforated diverticulitis with generalized peritonitis is unlike, and less severe than, the boardlike rigidity of a perforated ulcer.

Hyperesthesia over the involved area usually corresponds to the involved intraabdominal viscus. Hyperesthesia in the right lower quadrant may be present in patients with appendicitis. Hip flexion (psoas sign) indicates a retroperitoneal abscess or invasive neoplasm. With deep but gentle palpation, a mass may be appreciated. Bimanual pelvic and rectal examinations, with stool samples examined for occult blood, are routine components of examination of the abdomen.

Laboratory Evaluation

Laboratory evaluation should include white blood cell count and differential, especially the number of polymorphonuclear leukocytes and immature polymorphs (bands), the number of eosinophils, which may indicate an allergic reaction, and mononuclear cells and lymphocytes, which may suggest a viral or tubercular process. Urine is examined for red blood cells (calculi), white blood cells (infection), diabetes mellitus, and porphyria. In the female of child-bearing age, b-HCG or other accurate pregnancy test is essential. Blood urea nitrogen (BUN) and creatinine elevation indicate dehydration, hypovolemia, or upper gastrointestinal bleeding (elevated BUN). In patients with epigastric pain, amylase, lipase, and liver chemistries are essential to rule out pancreatic and biliary disease. In suspected pancreatitis, a lactic dehydrogenase (LDH) study should be added for prognostic purposes (Ranson’s criteria) as well as tests for calcium and phosphorus. In African-American patients, the status of sickling should be determined. An electrocardiogram may reveal a myocardial infarction that might mimic reflux or cholecystitis.

Radiologic Examination

An upright anteroposterior and a lateral chest x-ray and an upright and a supine abdominal film constitute the usual standard. In addition to reviewing all the bony structures, the diaphragmatic sulci should be examined for evidence of fluid and the lung parenchyma for evidence of a pneumonic process or neoplastic disease. The bowel should be evaluated for abnormal amounts of intraluminal air. Proceeding from the upper abdomen, the gastric bubble (the amount of fluid in the stomach) as well as the thickness of the gastric wall should be evaluated. If the gastric wall between the diaphragm and the gastric air bubble is thick, gastric neoplasia should be suspected. Biliary air suggests gallstone ileus or a previous biliary tract procedure. The thickness of the small bowel wall, the presence of fluid within the bowel, a stepladder pattern, biliary calculi, pancreatic calcifications, ureter calculi, vascular calcification, and fecaliths should all be ruled out. Properitoneal fat lines should be present, and their obliteration provides evidence of inflammation. Obliteration of pelvic lucent fat lines also suggests fluid or inflammation.

Ultrasonographic examination of the acute or posttraumatic patient is rapidly becoming the province of the surgeon, because it is noninvasive, easy to perform, readily available at the bedside, and reproducible. Ultrasonography is useful in evaluating suspected pancreatic or hepatobiliary disease. Gallstones, dilated ducts, liver tumors, ascites, edematous pancreatitis, and complications such as phlegmon, abscess, and pseudocyst are apparent in ultrasonographic examination. In patients with blunt abdominal trauma a focused abdominal sonogram for trauma has been reported to have a sensitivity of 93.4 percent, a specificity of 98.7 percent, and an accuracy of 97.5 percent in detecting hemoperitoneum and organ injury. In the patient with ambiguous clinical findings, ultrasonography can diagnose appendicitis or diverticulitis. Ultrasonography is particularly helpful in identifying gynecologic abnormalities that can mimic gastrointestinal inflammatory processes. Bedside ultrasonographic evaluation of the abdomen in patients with suspected bowel obstruction is as sensitive and more specific than plain roentgenograms in identifying a specific point of obstruction.

Computerized tomography (CT) scanning of the abdomen often is helpful, but it requires oral or rectal contrast agents, which the examiner may be reluctant to administer, particularly when there is any question of perforation. CT is very accurate, however, in demonstrating small-bowel obstruction and may be the technique of choice, particularly when extraluminal abnormalities are suspected or when prompt intervention is required. Magnetic resonance imaging (MRI) has proved useful in some situations.

The administration of a water-soluble contrast enema usually can be performed safely in the acute phase of diverticulitis, e.g., CT of the abdomen and pelvis with oral, intravenous, and water-soluble rectal contrast. Routine administration of rectal contrast in the patient with suspected diverticulitis is discouraged. CT scanning can identify diverticulitis and serve as a guide to percutaneous drainage of diverticulitis-associated abscess (Fig. 22-4). Arteriography or digital subtraction venography may be required in cases of suspected intestinal ischemia. Paracentesis for fluid or blood, laparoscopy, culdoscopy, and culdocentesis are useful adjuncts, and celiotomy is, in some cases the definitive diagnostic procedure.

Surgical Decision Making

The diagnosis is made primarily by the history. Physical examination is confirmatory. Laboratory examinations are confirmatory of the presumptive diagnosis based on the history and physical examination. Laboratory tests are focused, depending on the suspected diagnosis. The nature of surgical decision making, especially in the acute abdomen, is that it does not require a specific diagnosis, but a plan of action. Surgical decision making consists of (1) presumptive diagnosis, (2) indication for operation, (3) timing, and (4) approach.

In some cases of acute abdomen, symptoms can be masked, particularly in the pediatric population (Table 22-6), the elderly, those on steroids, the immunocompromised (Table 22-7), and patients who do not mount a normal inflammatory response. In the immunocompromised patient and in elderly patients who are unaware of their surroundings and do not have family, ethical issues arise. Patients in a prolonged vegetative state who have no interested family pose ethical problems. Patients with functional complaints, with psychologic abnormalities, those habituated to drugs, or patients who want an operation (Munchausen syndrome) should not be operated on.

Intermittent and Recurrent Pain

Various hematologic disorders can produce abdominal pain as part of their symptom complex. The majority of those disorders include pain as a prominent feature during a “crisis” only. Throughout their lives, patients with sickle cell disease are plagued by recurrent painful crises.

Autoimmune hemolytic anemia, thrombotic thrombocytopenic purpura, and other hemolytic disorders can begin abruptly and cause moderate to severe abdominal pain. The accompanying muscular spasm and rigidity may simulate an acute surgical emergency.

Chronic hemolytic anemias, hereditary spherocytosis, thalassemia, and glucose 6-phosphate dehydrogenase (G6PD) deficiency can cause abdominal pain. Clinical manifestations relate to the occurrence of crises, anemia, jaundice, splenomegaly, and cholelithiasis. This group of patients commonly has mild to moderate enlargement of the spleen. In some patients, symptoms of cholelithiasis or cholecystitis are the initial manifestation of a hemolytic crisis.

Acute intermittent porphyria is a dominantly transmitted inherited disorder that can exist in latent form indefinitely or present as an acute attack of neurologic dysfunction precipitated by various environmental and endogenous factors. The basic enzyme defect in this disease is a 50 percent decrease of uroporphyrinogen synthetase. Consequently, d-aminolevulinic acid, porphobilinogen, and uroporphyrinogen are produced in excess by the liver and excreted in the urine. An attack is precipitated by one of four factors: drugs, starvation, sex hormones, and infection. Drugs implicated include barbiturates, sulfonamides, phenytoin, griseofulvin, meprobamate, among others. Starvation and crash dieting precipitate attacks of porphyria. This is because of the ability of glucose and certain other carbohydrates to normally block the induction of hepatic d-aminolevulinic synthetase.

Acute attacks result in mental or abdominal neurologic symptoms. Symptoms of an acute attack result from nervous system damage. Any part of the nervous system can be involved; the specific clinical findings depend on the areas affected. Symptoms rarely occur before puberty. Abdominal pain frequently is the initial and most prominent symptom of the porphyric attack, occurring in about 95 percent of the patients. The pain can be moderate but often is severe and cramping or colicky, and it is typically localized in one of the lower quadrants or in the periumbilical region, with radiation to the back or flanks, but may be felt throughout the abdomen. Vomiting, constipation, and mild abdominal tenderness are common. The pain might result from autonomic neuropathy that causes disturbed gastrointestinal motility or alternating areas of spasm and dilatation. Other autonomic manifestations include labile hypertension, sinus tachycardia, postural hypotension, and sweating.

Chronic Pain

The pain pattern in patients with chronic abdominal pain can provide important diagnostic clues. Bouts of pain with entirely normal intervals usually are explained by a discrete intermittent disorder of physiology. Examples include acute intermittent porphyria, internal hernias, endometriosis, and, occasionally, choledocholithiasis. Chronic abdominal pain that is persistent usually is due to a clear pathophysiologic abnormality, such as chronic pancreatitis or pancreatic or colonic malignancy. Other cases of chronic abdominal pain may have no specific pathophysiologic abnormality. Nonulcer dyspepsia and irritable bowel syndrome are examples.

Pain arising from the abdominal wall frequently is misdiagnosed. Specific diagnoses include iatrogenic peripheral nerve injuries, hernias, myofascial pain syndromes, the rib tip syndrome, abdominal pain of spinal origin, and spontaneous rectus sheath hematoma. Carnett’s test is performed by palpating the abdomen of the supine patient in the usual way. With the palpating fingers located over the tender spot, the patient is asked to contract the abdominal muscles by raising his or her head from the bed. Once the muscles are tensed, pressure is reapplied and the patient is asked if the pain is changed. If the cause of the symptoms is intraabdominal, the tensed muscles should shield the viscera and result in diminished tenderness. On the other hand, if the source is in the abdominal wall, the pain will be worse or no better.

Many psychiatric disorders are associated with chronic abdominal pain. Diagnoses include primary affective disorders, somatization disorders, psychogenic (conversion) pain,

hypochondriasis, anxiety states, substance abuse disorders, schizophrenia, chronic factitious disorder with physical symptoms (Munchausen syndrome), and malingering.

Intractable Pain

The control of pain associated with diseases that cannot be satisfactorily treated is one of the most challenging and often frustrating problems. Examples are unresectable carcinoma of the pancreas and chronic pancreatitis. Opiate analgesics, if given in sufficient dosage, usually can control abdominal pain, but at the risk of addiction and undermining the patient's ability to function effectively. This probably is the best therapy for a patient with incurable disease and a life expectancy of a few weeks or months, but it should not be used in patients with severe pain from nonmalignant causes with an unpredictable life expectancy. Narcotic addiction before surgery usually compromises the outcome, because an addicted patient continues to be addicted even though pain is relieved by the procedure.

Neurosurgical interruption of the pain pathways, though disappointing at times, is the treatment of choice for intractable pain in properly selected patients with malignancy. Splanchnicectomy and celiac ganglionectomy can control abdominal pain effectively if somatically innervated structures are not involved in the pain-producing process. When both visceral and somatic pain fibers are involved, posterior rhizotomy or tract interruption is indicated. The spinothalamic tracts are usually interrupted, by surgical incision or by injection of sclerosing chemicals, into the spinal cord, a few segments above the segments where the noxious impulses are entering. The interruption also has been done at medullary and mesencephalic levels. After anterolateral cordotomy pain and temperature sensation are lost and proprioception and touch are virtually unimpaired. Paresthesia often replaces pain in the anesthetic areas, and in a significant number of patients, after about 1 year the paresthesia becomes disagreeable and painful. For this reason, root or tract interruption is most useful in patients with intractable pain and a short life expectancy.

For patients with pain arising from areas too great to be controlled by peripheral interventions, prefrontal lobotomy may be considered. Such lesions do not abolish pain but diminish the reaction to pain. The price paid for such relief, however, includes inability to experience pleasure, a flattening of all affect, and development of a more or less apathetic state.

A dorsal column stimulator is commercially available. Impulses arising from electrostimulation of descending dorsal column fibers are used to inhibit, in accordance with the gate-control theory, prolonged small-fiber afterdischarge that is uniquely related to pain. Electrodes are implanted four to eight segments above the pain input. The external transmitter is controlled by the patient. There is a buzzing or tingling sensation below the site of the stimulator, but pain usually is controlled without significant alteration of normal sensory function. Transcutaneous electrical stimulation has been used effectively to control postoperative pain. The results of these forms of therapy are variable.

Fever

Fever is not dangerous unless it is unusually high, e.g., in malignant hyperthermia. Fever may, by raising the temperature, improve host defense activity such as ingestion, killing, diapedesis, and other functions of neutrophils. More commonly fever is an indication of illness such as infection, inflammation, autoimmune disease, and, less commonly, neoplasia. Fever is a normal postoperative event, often a consequence of thermostat reset, heat generation as a response to intraoperative body cooling, and the result of normal cytokine activation and other inflammatory mediators from operative trauma. Persistent fever usually indicates an infectious complication.

Pathophysiology

The inflammatory response activates cytokines that are released locally within the brain or peripherally into the bloodstream and act on the hypothalamus in endocrine fashion. Activated macrophages also liberate pyrogens (interleukin-1, tumor necrosis factor, and interferon). These pyrogens cause an upward resetting of a thermoregulatory apparatus, which in turn triggers two physiologic mechanisms for increasing body temperature: vasoconstriction, which limits heat loss, and increased heat production, manifested by shivering. These two processes, in combination, increase body temperature until the temperature reaches the level required by the new hypothalamic set point. The response to alterations in body temperature is closely linked with behavioral changes. During a febrile episode, a patient might curl up in bed under layers of blankets, as if cold. This is because the patient's temperature, while elevated, is below the thermoregulatory set point. Removing the bed covers to reduce body temperature brings no relief and heightens stress; the preferred approach is to administer an antipyretic at an adequate dosage. The patient will start to sweat and remove the blankets as soon as the set point has been readjusted downward or as soon as the fever breaks.

Behavioral adaptation allows patients to establish what environmental temperature is most comfortable for them. This temperature is called thermoneutrality. Certain groups of patients tend to select an ambient temperature that is much warmer than that preferred by the hospital staff. For example, when burned patients without known infection were placed in an environmentally controlled chamber and allowed to regulate the ambient temperature by means of a bedside remote control, the mean ambient temperature they selected as most comfortable was significantly higher than the temperature selected by normal individuals. When comfortable, the burn patients had significantly higher core and mean skin temperatures than did the normal subjects. These results indicate that in burn patients the thermoregulatory set point is reset upward, probably because of the pyrogens that are produced during the inflammatory process that is associated with the injury. Other hospitalized patients also feel cold at temperatures that normal persons consider comfortable, e.g., persons with large wounds, peritonitis, pancreatitis, sepsis, and multiorgan failure; all have an elevated central set point.

Clinical Considerations

Fevers of gastrointestinal origin usually result from infection within the abdominal cavity, which are either monomicrobial or polymicrobial. Classic monomicrobial infections include biliary tract infections and spontaneous or primary peritonitis. Most cases of intraabdominal sepsis are polymicrobial, including both aerobic and anaerobic bacteria derived from the normal intestinal flora. The infections arise from the leakage or perforation of a hollow viscus within the abdominal cavity. Intraabdominal abscesses require drainage, except for most amebic liver abscesses.

Fever also is a presenting complaint in 50 percent of patients with Crohn's disease. Such fevers are low grade and usually associated with reactivation of the disease and its characteristic symptoms of crampy pain, diarrhea, and abdominal masses. Postoperative fevers are common, occurring in 15 to 30 percent of patients after celiotomy. Only 10 to 20 percent of postoperative fevers are caused by established infections. Common causes include pulmonary complications, urinary tract infections, wound sepsis, and thrombophlebitis. Acalculous cholecystitis should not be overlooked. Fever is the presenting symptom in 75 percent of patients with acute hepatitis from halothane or infectious causes. Occult tumors also present initially with fever. Twenty percent of fevers of unknown origin are secondary to cancers, either primary or metastatically involved in the abdomen. About 5 percent of fevers in patients with neoplasms are related to the tumor and not a complication. The most commonly implicated tumors are hypernephromas, liver tumors, lymphomas, and carcinomas of the stomach, colon, and pancreas.

Anorexia

Anorexia complicates many illnesses. In inflammatory disease, it is not necessarily prognostic; anorexia can occur even in mild illnesses. Anorexia usually indicates a significant degree of inflammation, but it is not necessarily the direct result of inflammation. One common clinical dictum is that if a patient is hungry, the diagnosis of appendicitis is probably incorrect.

Anorexia also complicates various endocrinopathies, including hyperparathyroidism, especially with extreme elevations of calcium, and adrenocortical insufficiency. Anorexia is common in liver disease, presumably through a central nervous system mechanism, probably mediated, at least in part, by a combination of ammonia and *neuropeptide Y (NPY)*.

A common cause of anorexia is carcinoma. In this circumstance, the presence of anorexia carries with it a grim prognosis, because it usually accompanies a significant tumor burden. Tumors that produce anorexia without significant tumor burden include carcinoma of the pancreas and small-cell carcinoma of the lung.

Postoperative anorexia is common and may be associated with a loss of taste, especially in patients with diarrhea whose zinc stores are chronically low. Zinc replacement (220 mg daily) will return taste in 10 days to 3 weeks, and, with it, perhaps appetite. Other than zinc, there is no mechanism for stimulating appetite. This is especially true of patients who have been on total parenteral nutrition or enteral feeding for a prolonged period, and who may have difficulty resuming their oral intake. Some patients find alcohol a stimulant; bringing food from home may be useful in others.

Pathophysiology

The physiologic mechanisms that control feeding and satiety are complex and multifaceted, and involve receptors in both the brain and the periphery that must interact to maintain normal nutrient intake and body weight. Several independent variables have been observed to alter food intake consistently in normal animals, including: (1) changes in glucose utilization rate, (2) changes in the rate of lipid metabolism, (3) alterations in brain and peripheral peptides, (4) imbalance in plasma and brain amino acid profiles, (5) increases and decreases in neurotransmitter activity, and (6) alteration in cytokine levels.

Central Nervous System Control

The hypothalamus is the "feeding center" of the brain. Lesions of the lateral area produce anorexia, and lesions of the ventromedial region cause hyperphagia and obesity. Conversely, electrical stimulation of the lateral and medial areas elicits and inhibits feeding, respectively. Other brain areas are important for the modulation of feeding as well as some degree of neurochemical specificity subserving hunger and satiety. Neurons ascending from brainstem indoleamine and catecholamine nuclei and passing through the hypothalamus appear to be as important as local hypothalamic circuits for the modulation of feeding. The nucleus accumbens and central amygdala contribute to the reinforcement value of food and appreciation of food quality, respectively. Vagal afferents are important for communicating peripheral nutrition-related information from the gastrointestinal tract or from glucose-sensitive cells in the liver. This information begins to be processed in the dorsal nucleus of the vagus, which sends inputs to the nucleus tractus solitarius. After bifurcation in the parabrachial nucleus of the pons, branches of these neurons synapse in the paraventricular hypothalamus or continue in the limbic system to terminate in the central amygdala. This pathway, which includes glucose-sensitive neurons, not only provides for communication between the periphery and the central nervous system but also allows the interaction of several neurotransmitters, neuropeptides, and neuromodulators.

Influences of Peripheral Metabolism

Although pathologic alteration in circulating glucose levels affects nutrient intake, smaller decreases in blood glucose level have been related to the onset of hunger. According to the glucostatic hypothesis of feeding, increased use of glucose should signal satiety, and decreased glucose metabolism is associated with hunger. The degree of control, however, appears to be for short-term regulation only.

Increased lipolysis and oxidation of lipids decreases feeding and produces satiety. This "lipostatic" theory of feeding and satiety is thought to be major controlling factor in the diurnal feeding pattern observed in rats. Fat mobilization and oxidation produces the satiety pattern observed in the light phase of the light/dark cycle, while fat synthesis elicits the dark phase feeding pattern. The discovery of the obese gene protein (leptin) provides a proposed mechanism for modulation of food intake and body weight that is directly related to adipose tissue mass. Since leptin is synthesized in adipocytes, levels of circulating leptin increase as the amount of adipose tissue increases. Elevated circulating leptin concentrations have negative effects on the synthesis and release of hypothalamic NPY, resulting in reduced food intake and increased metabolic rate. Therefore, leptin level may serve as a negative feedback mechanism for long-term body weight regulation. Obese rats and mice appear to produce faulty leptin or to have a deficit in leptin receptors.

Central Nervous System Neurotransmitters

The preponderance of evidence for neurotransmitter involvement in the control of feeding suggests that alpha-adrenergic neuronal activity, particularly in the medial hypothalamic area, stimulates feeding, and increased beta-adrenergic activity causes satiety. Dopamine systems also are involved in feeding, with an *optimal* dopamine input necessary to maintain food-associated reinforcement (perhaps localized in the nucleus accumbens), or the motor activity (perhaps in the corpus striatum) necessary for feeding to occur. Increasing dopamine activity beyond this normal level, however, reduces feeding. Stimulation of dopamine neuronal activity is the primary mechanism of amphetamine- and cocaine-induced anorexia. Elevated activity in serotonin neurons also inhibits feeding; reducing serotonin levels usually increases the amount of feeding elicited by orexigenic agents. Serotonergic systems appear to exert a type of tonic inhibition over feeding circuits, with food intake increasing in magnitude when this inhibition is released by blockade of central serotonergic postsynaptic receptors.

Gamma-aminobutyric acid (GABA) also has been implicated in the control of feeding, with stimulation of GABA-a receptors in the medial hypothalamus or central amygdala eliciting feeding, and GABA-a stimulation in the hypothalamus inhibits feeding. It has been suggested that a GABA- benzodiazepine receptor interaction mediates hedonic aspects of taste and palatability. Alterations in circulating levels of other amino acids elicit rapid and severe hypophagia. Changes in meal pattern and meal frequency have been described in normal rats with exposure to dietary amino acid imbalance, deficiency, and excess. As with GABA, the amygdala appears to be a focal area for the effects of amino acid imbalance on food intake.

Peptides participate in the control of feeding and satiety. Evidence exists for involvement of NPY, endorphins, galanin, *cholecystokinin (CCK)*, *corticotrophin-releasing factor (CRF)*, calcitonin, glucagon, insulin, enterostatin, and bombesin in the control of feeding. NPY, endorphins, and galanin stimulate feeding, and the remaining peptides produce satiety.

NPY is abundant in the central and the peripheral nervous systems, and it is colocalized with norepinephrine. It is the most potent orexigenic agent investigated in rats to date. In particular, carbohydrate intake appears to be under NPY control. Although NPY is localized in several hypothalamic nuclei, the arcuate and paraventricular nuclei have particularly high concentrations of this peptide. Some NPY neurons in the paraventricular nucleus arise from brainstem catecholaminergic sources, but NPY also is synthesized in the arcuate nucleus with projections terminating in the paraventricular nucleus. Hypothalamic NPY levels and messenger RNA expression are altered in normal and pathologic feeding. Genetically obese Zucker rats and experimental diabetic rats, both hyperphagic, exhibit increased hypothalamic NPY concentrations in the paraventricular and arcuate nuclei. NPY message also is elevated in the arcuate nucleus of these rats. Food-deprived rats also exhibit elevated NPY levels in hypothalamic tissue, with tissue concentrations declining with feeding.

Galanin is located in several brain regions, including the hypothalamus, and it elicits feeding when injected into the paraventricular nucleus. This peptide appears to selectively stimulate the intake of fat. Enterostatin, a pentapeptide released during fat digestion, reduces the intake of fat after peripheral or central administration. Endorphins appear to be involved in the control of fat intake, which increases after the injection of beta- endorphin into the paraventricular nucleus, and is reduced after treatment with the opioid antagonist naloxone.

The release of CCK during feeding may serve as a satiety signal. CCK a and b receptors are found in the brain and periphery, but a central location for CCK-induced satiety has not been identified. Localization of CCK a receptors within the nucleus tractus solitarius and the area postrema suggests that CCK-induced satiety may share conditioned taste aversion effects. This phenomenon occurs when a taste is associated temporally with a noxious stimulus.

Apart from their regulation of glucose synthesis, uptake, and metabolism, insulin and glucagon both appear to reduce food intake when administered into the hypothalamus. Although large doses of insulin have been associated with hyperphagia, the increased feeding is because of the extreme hypoglycemia.

CRF has been reported to reduce food intake after its injection into the paraventricular nucleus. This peptide controls the release of adrenocorticotrophic hormone from the pituitary gland and corticosterone or cortisol from the adrenal glands. Adrenalectomy reduces feeding elicited by intrahypothalamic NPY or norepinephrine primarily through removal of feedback inhibitory influences on CRF. A balance of many of these peptides must be maintained for normal control of hunger and satiety.

Cytokines (tumor necrosis factor and interleukins) reduce food intake after peripheral infusion or direct injection into the cerebral ventricles or the hypothalamus. Tolerance to the anorectic effect of tumor necrosis factor develops after continuous intravenous infusion of tumor necrosis factor.

Despite our increased knowledge of feeding and its control, anorexia is a nonspecific symptom, but its presence requires investigation. Inflammatory, neoplastic, and viral diseases of the gastrointestinal tract result in anorexia, as does advanced renal disease, congestive heart failure, and endocrine disorders, such as pan-hypopituitarism, adrenocortical insufficiency, and hyperparathyroidism. Many drugs cause anorexia, but there is no medication that restores appetite.

Cancer Anorexia

It is likely that cancer anorexia results from tumor-induced aberrations of neurochemical mechanisms that normally control hunger and satiety. Anorectic tumor-bearing rats do exhibit significant changes in many of these nutrition-related variables. Although tumor-bearing rats exhibit altered glucose metabolism, including decreased insulin and increased glucagon concentrations in the blood, the glucostatic hypothesis of feeding proposes only short-term regulation of food intake by changes in glucose dynamics. Alterations in insulin and glucagon levels in tumor-bearing rats may be secondary to the anorexia.

Significant aberrations in plasma and brain amino acid profiles have been reported in tumor-bearing organisms. Correcting plasma amino acid alterations by providing diets deficient in histidine, lysine, threonine, and excess in tryptophan did not improve anorexia. Glutamine, citrulline, serine, arginine, and the branched-chain amino acids (leucine, isoleucine, valine) are reduced in the plasma of tumor-bearing rats early in the course of anorexia, while concentrations of alanine and phenylalanine are elevated. The brain amino acid profile of tumor-bearing rats, however, reveals a different pattern, with increases in the concentrations of the large neutral amino acids tyrosine, tryptophan, methionine, phenylalanine, histidine, threonine, and glutamine.

Since tyrosine and tryptophan are precursors for catecholamine and indoleamine neurotransmitters, respectively, the alterations in these amino acids observed in anorectic tumor-bearing rats also may affect neurotransmitter synthesis and metabolism. Increased brain serotonin neuronal activity usually has been associated with satiety. A variety of studies have implicated alterations in tryptophan and serotonin in cancer anorexia, but it appears that these alterations might be the result of anorexia/cachexia, rather than the cause.

Although dopamine metabolism is elevated in the hypothalamus, nucleus accumbens, corpus striatum, and amygdala of anorectic tumor-bearing rats and is normalized when the tumor is resected and normal feeding returns, a variety of dopamine receptor blockers were ineffective against cancer anorexia.

CCK-induced anorexia does not synergize with cancer anorexia, and the absence of increased CCK in plasma and brain of anorectic tumor-bearing rats suggests that this peptide is not a primary mediator of cancer anorexia. CCK levels were reduced in the hypothalamus of anorectic tumor- bearing rats, suggesting possible down-regulation of putative CCK satiety systems in response to cancer anorexia. Endogenous opioid systems have been investigated and no evidence for involvement of endorphins in cancer anorexia has been elicited.

NPY, a very potent orexigenic agent, appears to be dysfunctional in anorectic tumor-bearing rats. NPY is a less potent feeding stimulus in tumor-bearing rats, even before the onset of overt anorexia. Hypothalamic concentrations of NPY are reduced in tumor-bearing rats before the onset of anorexia, and continue to decrease as anorexia develops, while pair-fed control rats exhibit the opposite response. Release of hypothalamic NPY and binding affinity of NPY to hypothalamic membranes taken from tumor- bearing rats is decreased. These and other results strongly point to a dysfunction of NPY feeding mechanisms as one neurochemical mediator of experimental cancer anorexia.

Factors produced by tumor tissue may contribute to anorexia directly or induce alterations in host metabolism that cause anorexia. Glucose and glutamine are primary fuels for tumor. Since anaerobic glycolysis is the principal route of tumor metabolism of glucose, considerable amounts of lactic acid are produced. Metabolism of glutamine by tumor tissue releases ammonia into the circulation. Both of these metabolites contribute to anorexia, and their detoxification also presents a continuous energy drain on the host. Hyperammonemia, in particular, appears characteristic of larger tumor burdens and experimentally may produce anorexia. Patients with large tumors have hyperammonemia.

Cytokines may be involved in tumor-related anorexia/cachexia. Mice transplanted with a xenograft that produced human *tumor necrosis factor (TNF)* exhibited cachexia, while those transplanted with xenografts that did not produce TNF were not cachectic. Treatment of tumor-bearing mice with TNF antibodies delayed but did not prevent the appearance of anorexia. The blood concentrations of TNF are not high enough to be measured in anorectic tumor-bearing organisms, and large doses (relative to endogenous levels) of TNF are required to reduce food intake. Interleukins have been suggested as anorexia-producing candidates.

Weight Loss and Cachexia

Weight loss and cachexia have significant negative prognostic implications. Weight loss of 10 to 15 percent during 3 to 4 months is one of the two prime indicators that a patient may be nutritionally or immunologically impaired.

The history taking should discover whether intake has diminished and, if so, for what reason and to what degree. Anorexia also may have been present. Confirmatory laboratory values include serum albumin levels of less than 3.0 g/dL in the hydrated state and decreased short-turnover proteins (retinol-binding protein, thyroxine-binding prealbumin, and transferrin). A functional history emphasizing the inability to carry out normal functions, such as walking, mowing the lawn, taking out the garbage, etc., should be obtained. Inability to perform these functions has serious negative prognostic complications.

Decreased intake and weight loss may result from obstruction, as in neoplastic disease of the esophagus or stomach. Oral intake may be voluntarily restricted because of postprandial pain generated by eating, as in chronic mesenteric ischemia, or because of postprandial cramps and diarrhea in Crohn's disease. In chronic pancreatitis, intake may be normal, but fat malabsorption can result in weight loss.

Chronic infection also causes weight loss. Examples include tuberculosis, chronic low-grade infection such as a blind-loop syndrome, and an increasingly large group of immune-deficiency diseases such as HIV infection. Patients who have had a transplant also may contract diseases such as cytomegalovirus enteritis, with late complications resulting in malabsorption, diarrhea, and significant weight loss.

Thyrotoxicosis and Addison's disease are associated with rapid weight loss over a short period. Diabetes mellitus may result in weight loss if unrecognized; before the diagnosis is

made, polydipsia, polyphagia, polyuria, fatigue, and weight loss are prominent presenting symptoms in new-onset diabetes.

The most common cause of weight loss is malignancy. Cachexia is commonly associated with advanced malignancy. Up to 70 percent of cancer patients die from the effects of starvation, with infection the final common pathway. Cancer cachexia usually is seen in the late stages of pancreatic cancer and small-cell lung cancer, but cancer anorexia can be an early symptom of localized disease with small tumor burden; the mere presence of cachexia should not be a contraindication to surgery. Cancer cachexia is not always associated with diminished intake, but in many cases intake is decreased. More obvious causes of cachexia in patients with neoplastic disease include obstruction of the esophagus, gastric cardia, or colon. There may be aversions to certain foods.

Weight loss and malnutrition in the elderly population is particularly common, requiring special attention, especially when it results from treatable causes, such as poorly fitting dentures. The elderly may gradually lose weight over time. This is associated with loss of height secondary to osteoporosis, and loss of muscle mass because of decreased exercise and decreased exercise ability. Of patients who present with weight loss, the weight loss is related to illness in about 75 percent, and is of psychologic origin in about 10 percent; 15 percent remain undiagnosed.

Hiccups (Singultus)

Hiccups, or singultus, usually is a transient and benign annoyance but can be quite debilitating if persistent. Hiccups can be a manifestation of an underlying severe pathologic process. They remain a medical enigma, with no known useful function to the organism. The origin of the word “singultus” is from the Latin *singult*, which means “the act of attempting to catch the breath while sobbing.”

The relationship between phrenic nerve irritation and hiccups has been recognized for many years. There is believed to be a hiccup reflex arc. The afferent portion of this arc comprises the phrenic and vagus nerves and the sympathetic chain arising from thoracic segments T6 to T12. The central connection between afferent and efferent limbs of the reflex arc is not ascribable to a specific anatomic location. It appears that the center controlling hiccups is a nonspecific anatomic location between spinal cord segments C3 and C5.

Hiccups have been categorized into those that are benign and self-limited, and those causing persistent or intractable episodes; the latter are more likely to result from a serious pathophysiologic process. Benign self-limited hiccups are caused by gastric distention, sudden changes in ambient temperature, alcohol ingestion, and tobacco use. Intractable hiccups are classified as organic, psychogenic, or idiopathic. The organic causes are divided into central, peripheral, and toxic. The surgeon's interest mostly is in the peripheral causes from phrenic or vagal nerve stimulation or direct diaphragmatic stimulation.

Irritation of the vagus nerve along its course may cause hiccups. In the chest, stimulation may result from trauma, neoplasm, myocardial infarction, pulmonary edema, inflammatory or infectious processes, or intraoperative manipulation of thoracic viscera. Phrenic nerve irritation within the chest caused by intrathoracic inflammation, tumor, infection, or trauma also can lead to hiccups.

A diverse group of abdominal conditions can stimulate the afferent vagal branches, e.g., gastric distention, peptic ulcer disease, pancreatic and biliary disease, bowel obstruction, appendicitis, inflammatory bowel disease, hepatitis, and trauma or intraoperative surgical manipulation. Intraabdominal phrenic nerve stimulation can arise from hiatal hernia, diaphragmatic eventration, subphrenic abscess, neoplastic disease, perihepatitis, and intraoperative manipulation. Several toxic causes are relevant to surgical practice, including anesthetics and drugs such as intravenous steroids, barbiturates, benzodiazepines, and methyldopa. Metabolic causes include acute alcohol toxicity, uremia, hypocalcemia, and hyponatremia.

The treatments of hiccups are diverse. Nonpharmacologic hiccup treatments usually rely on some method of nasopharyngeal stimulation; examples include forcible traction on the tongue, gargling or sipping iced water, swallowing a tablespoon of granulated sugar, grape jelly under the tongue, and inhalation of noxious fumes (ammonia). Direct pharyngeal stimulation with a rubber catheter is reportedly successful in 90 percent of cases. Pharmacologic treatments include administering continuous positive- pressure ventilation at 25 to 35 cmH₂O, chlorpromazine, haloperidol, phenytoin, phenobarbital, carbamazepine, and sodium valproate. Other agents sporadically reported in the literature include metoclopramide, amitriptyline, chloral hydrate, and ketamine.

Symptoms Related to Specific Components of the Gastrointestinal Tract

Esophagus, Stomach, and Duodenum

Heartburn and Dyspepsia

Heartburn and dyspepsia are common, and many normal individuals have experienced heartburn while lying down or bending over after overeating. Gastric capacity is exceeded and reflux occurs, resulting in pain and substernal pressure, at times radiating to the interscapular region, blades, the neck, or the left arm. With more regular occurrence, the patients has gastroesophageal reflux disease.

The esophagus normally is protected from damage by four defense mechanisms: a competent lower *esophageal sphincter (LES)*, rapid esophageal clearing of refluxed material, neutralization of refluxed acid by bicarbonate-rich saliva, and an intact mucosal diffusion barrier. The most common clinical abnormality of esophageal motility is incompetence of the LES. As incompetence becomes more frequent and reflux becomes more voluminous, salivary bicarbonate can no longer keep up and esophageal pH falls. This leads to compromise of the mucosal barrier with damage to the subepithelial structures, especially nerves, with resulting spasm and dysmotility further delaying clearing of the distal esophagus, and relaxation of the LES. This cycle set the stage for additional reflux and the development of esophagitis.

Heartburn usually is described by the patient as a substernal burning, often associated with a bitter taste, and, when severe, frank regurgitation of gastric contents. Approximately 40 percent of Americans (100 million people) experience an episode of heartburn monthly, nearly 40 million experience weekly episodes, and 18 million have chronic daily heartburn. Eighteen percent of Americans (45 million) take some form of nonprescription or prescription medication for this problem, and 1.75 million Americans do not respond to these therapies. Most gastroesophageal reflux dysmotility does not result in any serious long- term sequelae, but esophagitis occurs in approximately 1.1 percent of the world's population and is present in approximately 19 percent of patients undergoing upper endoscopy for upper abdominal symptoms.

Dyspepsia is a nonspecific term given to a collection of symptoms involving the esophagus, stomach, duodenum, biliary tree, and pancreas. Dyspepsia generally is a postprandial complaint involving substernal pressure, epigastric distress, nausea, and bloating. In the case of dyspepsia and heartburn, the symptom is described as substernal pressure. Classic esophageal pain radiates directly through the back between the scapulae. It also can radiate to the left shoulder, to the neck and teeth, and down the left arm, mimicking angina. In 15 percent of patients heartburn and substernal pressure with reflux can be so difficult to distinguish from angina pectoris that coronary angiograms, stress tests, echocardiograms, and other forms of cardiac evaluation, such as multiple gated acquisition (MUGA) scans, may be required. Only 3 percent of patients with dyspepsia have true angina pectoris. Another symptom related to reflux is deep pain immediately under the left costal margin, pointing toward the left shoulder.

If there is serious concern that a patient might be suffering from dyspepsia, the work-up should include cinefluoroscopy, contrast radiography, or an upper gastrointestinal series, endoscopy, and biopsy. If esophageal spasm is suspected, manometry will be helpful in distinguishing among achalasia and other motor disorders, scleroderma, and dysmotility complicating reflux esophagitis.

The relationship of symptoms to eating and other activities may give clues in the diagnosis of dyspepsia. *Duodenal* ulcer pain is relieved by food, particularly protein-containing food, which is a good acid neutralizer. *Gastric* ulcer pain may be provoked by food, and may be slightly more central than duodenal ulcer, which is typically slightly to the right of the midline. In alkaline reflux gastritis with severe inflammation, any food results in pain. Constant nausea and periodic clear bile vomitus without food may be present.

In patients with gallbladder and pancreatic disease, not all foods provoke pain. Classically in patients with gallbladder disease, fried, fatty, or greasy foods, chocolate, and, in some patients, cabbage and turnips, provoke right upper quadrant pain with right subscapular radiation. In patients with chronic pancreatitis dietary indiscretions, such as “fast and surfeit”—that is, gluttonous intake after a period for relative fasting—or alcohol with large meals, may provoke panceratitis, which characteristically is associated with central epigastric or left upper quadrant pain radiating to the center of the back.

Dysphagia

Disturbances in swallowing can be categorized according to the etiologies of degenerative, functional, inflammatory, mechanical (sometimes as a result of inflammatory), autoimmune (or collagen-type), and neoplastic diseases. Additional esophagus-related symptoms include painful swallowing (odynophagia) or the inability to swallow. Either symptom requires investigation. The history taking should focus on whether the pain is entirely new or is associated with previous episodes. An example is dysphagia that follows previous chronic odynophagia secondary to heartburn or to spasm associated with reflux esophagitis. The focus should center on chronic medications (e.g., nonsteroidal anti-inflammatory drugs), dietary habits, and substance abuse, particularly smoking or taking alcohol undiluted by water, because this history is significant in patients with squamous-cell carcinoma of the esophagus. Regurgitation of undigested foul-smelling food or a feeling of a bulge in the neck both suggest Zenker's diverticulum. For causes of dysphagia, see Table [22-8](#).

Pain is almost always due to inflammation or spasm. Spasm may be due to a primary dysmotility disorder (diffuse esophageal spasm), or secondary to chronic inflammation associated with gastroesophageal reflux. Diffuse esophageal spasm may simulate angina pectoris, because it is often precipitated or worsened by stress. When gastroesophageal reflux dysmotility is associated with severe pain and spasm it usually indicates associated ulceration, with at least grade II and perhaps grade III esophagitis. Spasm or substernal pressure also may accompany the dysphagia and nocturnal aspiration from achalasia.

Radiologic examination should be followed by upper gastrointestinal endoscopy. The location of the Z line, its relationship to the diaphragm, the presence or absence of a hiatal hernia or esophagitis and ulceration should be noted. The esophagus should be biopsied and, if Barrett's esophagus is suspected, multiple biopsies at various levels should be performed to evaluate the extent of the metaplasia as well as the presence and degree of dysplasia. The stomach and duodenum should be carefully inspected to make certain that there is no duodenal obstruction. If esophagitis, Schatzki's ring, neoplasm, or reflux are not diagnosed, further studies, including manometry and 24-h pH monitoring are required. Zenker's diverticulum must be ruled out. Even if the esophagitis is severe and reflux symptoms are well established, manometry and 24-h pH monitoring should be undertaken to discern what type of repair is necessary. Other diagnoses that can be clarified include achalasia, in which a large, distended, tortuous esophagus and a spastic sphincter are causes of this symptom. Achalasia initially may be amenable to hydrostatic dilation or to esophago- cardiomyotomy.

The esophagus is the most common part of the gastrointestinal tract to be affected buy progressive systemic sclerosis (scleroderma). This chronic connective tissue disorder is characterized by collagen deposition and atrophy of the smooth muscle of the esophagus, resulting in the characteristic decreased amplitude of contractions of the lower

esophagus and decreased LES pressure, which may lead to heartburn and dysphagia. Lower esophageal dysmotility can occur in association with other connective tissue disorders such as rheumatoid arthritis, Sjögren's syndrome, and mixed connective tissue disease. A distal esophageal web (Schatzki's ring) may be visualized above a hiatal hernia and, when associated with dysphagia, may be amenable to graduated mercury bougie (Maloney) dilation. If nothing is evident on endoscopy, and cinefluoroscopy or manometry suggests diffuse spasm or dysmotility, beta blockers or nitroglycerine can be used if these medications are tolerated without significant side effects. Dysphagia lusoria is caused by an aberrant origin of the right subclavian or pulmonary artery. This rare condition is suggested by the observation of a posterior indentation of the upper esophagus and can be confirmed with angiography.

If neoplasm is diagnosed, further studies depend on its location. CT scan of the chest, mediastinum, and upper abdomen may be useful, and, in carcinoma of the midesophagus, bronchoscopy may be useful. Depending on the extent of the disease seen on CT scan, the presence or absence of celiac or other adenopathy, preoperative chemotherapy and radiotherapy may be useful.

Dysphagia is a relatively common manifestation of psychologic disturbances. Inability to swallow may accompany conversion hysteria, anxiety, and anorexia nervosa.

Nausea and Vomiting

Nausea and vomiting may be related or unrelated to diseases of the gastrointestinal tract. Nausea and vomiting can occur independently of each other, but generally are so closely related that they can be considered together. Whenever any part of the upper gastrointestinal tract becomes excessively irritated, overdistended, or excitable, vomiting may result. In the early stages of such excessive gastrointestinal irritation or distention, antiperistalsis begins to occur. This antiperistaltic wave may begin in the ileum, and a large portion of the intestinal contents may be pushed back to the duodenum and stomach. Duodenal distention is a particularly potent stimulus for vomiting. Impulses from the gastrointestinal tract are transmitted by both vagal and sympathetic afferents to the bilateral vomiting center of the medulla. This center lies near the tractus solaritus at the level of the dorsal motor nucleus of the vagus and controls and integrates the act of emesis. Once the vomiting center has been sufficiently stimulated, motor impulses are then transmitted to the upper intestinal tract via cranial nerves V, VII, IX, X, and XII, and to the diaphragm and abdominal muscles through the phrenic and spinal nerves. The vomiting act results from a squeezing action of the muscles of the abdomen associated with gastric and duodenal contractions along with sudden relaxation of the esophageal sphincters, allowing the forceful oral expulsion of gastric contents. The act of vomiting depends on the coordinated closure of the glottis, contraction and fixation of the diaphragm in the inspiratory position, and closure of the pylorus.

In addition to receiving stimuli from the intestinal tract, the vomiting center receives input from other parts of the body, from higher cortical centers, and from the chemoreceptor trigger zone (CTZ). The CTZ is a small area located near the area postrema bilaterally on the floor of the fourth ventricle. Activation of this zone results in efferent signals that stimulate the vomiting center. Many stimuli activate the CTZ via the dopamine receptors present there. In particular, many drugs (e.g., apomorphine, morphine, digitalis, levodopa, quinidine, acetylcholine, ergot alkaloids, atropine, tartar emetic, ipecac) may directly stimulate the CTZ receptors and initiate activation of the vomiting center. Rapidly changing directions of motion may stimulate receptors in the labyrinthine apparatus, from which impulses are carried via the vestibular nuclei into the cerebellum, then to the CTZ, and finally to the vomiting center, leading to emesis associated with "motion sickness." Higher cortical centers also can stimulate the vomiting center directly after various psychic stimuli, such as distressing visual input, foul odors, and other psychologic disturbances.

Vomiting usually is preceded by the sensation of nausea. Nausea represents the conscious recognition of excitation of an area of the medulla that is closely associated with or part of the vomiting center. The same irritative impulses (intestinal, lower brain, cortical) that lead to vomiting may produce nausea. Nausea often is accompanied by anorexia, skin pallor, perspiration, salivation, and altered autonomic activity, which, when severe, may lead to bradycardia and hypotension (vasovagal response). It appears that only certain areas of the vomiting center are associated with nausea, because vomiting may occur in certain circumstances without any prior sensation of nausea.

Nausea and vomiting are commonly associated with central nervous system (CNS) abnormalities. Reduction of the oxygen supply to the vomiting center has been implicated in the vomiting associated with increased intracranial pressure, anemia, vascular occlusion, shock, and severe pain. Any CNS disorder that leads to an increased intracranial pressure may be accompanied by vomiting. There may be no antecedent nausea. Vomiting usually is sudden in onset and often projectile. Migraine may increase intracranial pressure, followed by an interference of blood flow to the cerebromedullary center and the induction of vomiting. Vertigo associated with acute labyrinthitis or Meniere's disease often is associated with nausea and vomiting. Vomiting commonly accompanies migraine headache and acute meningitis.

In many systemic disorders nausea and vomiting are associated symptoms. These include acute systemic infections, particularly in pediatric patients, acute myocardial infarction (posterior wall), congestive heart failure, and endocrinologic or metabolic illnesses. The vomiting associated with diabetic acidosis, uremia, and Addisonian crisis has been related to changes in electrolyte levels and, more specifically, acidosis and hyperkalemia affecting the emesis center. The vomiting associated with thyrotoxicosis appears to be related to an augmented irritability of the emesis center. The hormonal changes of early pregnancy lead to the nausea and vomiting of "morning sickness" in many women. When extreme and unrelenting, this may lead to hyperemesis gravidarum, and the patient may require hospitalization for intravenous hydration, enteral nutrition, and, occasionally, parenteral nutrition. Pain of any origin may be associated with nausea and vomiting. Inflammatory illness of any sort is associated with a generalized malaise and can trigger nausea and vomiting. Acute inflammatory diseases of the intestine and pelvic organs are associated with nausea and vomiting through visceral reflexes. When the inflammatory illness is associated with gastrointestinal tract disorders, such as appendicitis, cholecystitis, hepatitis, and pancreatitis, nausea and anorexia are prominent.

A wide variety of medications can provoke nausea and moving by central mechanisms. Examples include morphine via the chemoreceptor trigger zone, glycosides (overdosage), or chemotherapeutic agents. A variety of antibiotics, including erythromycin, bactrim, and neomycin, may cause nausea and vomiting by direct effects on the gastrointestinal tract.

Any disturbance within the gastrointestinal tract may cause nausea and vomiting. Inflammation or infectious agents affecting the gastrointestinal tract, neoplastic disease, or mechanical obstruction at any level may lead to acute nausea and vomiting. Malignant tumors of the stomach and obstructing gastric or duodenal ulcers are common causes of gastric outlet obstruction. Obstructing prepyloric or gastric ulcer typically interferes with the propagation of the peristaltic wave and normal rhythmic emptying associated with the opening of the pylorus rather than creating a tight mechanical obstruction. Disturbances in the stomach and inflammation such as in gastritis, gastric and duodenal ulcer, and duodenitis as well as cholecystitis, pancreatitis, and various inflammations of the small bowel, such as viral gastroenteritis, can give rise to nausea and vomiting. Obstruction of the colon and diverticulitis also can result in nausea and vomiting.

The patient with an eating disorder is becoming more common in our appearance-conscious society. If bulimia is suspected, a study indicating a normal gastrointestinal tract should be followed by a referral for psychiatric consultation. The surgeon may be needed for nutritional support in these patients. Symptoms may be related to gastric dysfunction or to small- bowel dysfunction.

Consequences of Vomiting

Nausea and vomiting can produce dehydration and volume contraction, resulting in hemodynamic instability. Elevations of blood urea nitrogen (BUN) and creatinine are late features of such dehydration. Electrolyte changes include subtraction hypokalemia by the vomiting of material relatively rich in potassium and hydrogen ions, resulting in hypokalemic alkalosis.

When hydrogen ion is excreted into the gastric juice, there is a shift of bicarbonate ion into the plasma, and to maintain electric neutrality of the blood, chloride is excreted into the gastric juice. Concentration of bicarbonate in the plasma is augmented by the loss of chloride and excessive sodium in the vomitus. In general, an inverse relation exists between the concentration of hydrogen ion and sodium ion in the vomitus. Even in the achlorhydric state, when vomiting results in a depletion of sodium chloride and potassium chloride, chloride is excreted in excess of the physiologic proportions of plasma, i.e., 145 mEq of sodium and 100 mEq of chloride. As the plasma bicarbonate rises, the body compensates by increasing renal excretion of bicarbonate and reducing the rate and depth of respiration to decrease the respiratory loss of HHCO_3 to maintain the acid/base ratio. These compensatory features may maintain the normal blood pH while the urine is alkaline because of excretion of bicarbonate. Determination of plasma electrolyte levels is only partly informative, because it is dependent upon the relative loss of electrolytes and water and on the solutions ingested.

With continued vomiting, the plasma and extracellular potassium concentration become reduced as a result of the increased quantities of potassium excreted in the urine in exchange for sodium, which in turn is related to the lack of availability of hydrogen ions depleted by loss in the vomitus. Adrenocortical stimulation intensifies the potassium loss and potentiates the absorption of bicarbonate by the renal tubular cell. The extracellular and subsequently intracellular potassium concentrations become reduced, sodium cations shift into the cell, and, in order to conserve sodium an acid urine is formed in the face of generalized alkalosis. This paradoxical aciduria intensifies the alkalosis and sets up a vicious cycle that includes the shifting of more potassium out of the cell in exchange for sodium.

The fluid loss results in reduction in the circulating blood volume, and in time the consequences of starvation such as cellular breakdown of protein and increased renal load of nitrogenous waste products cause a rise in the BUN level. Fat stores are used. Ketone bodies are formed, and since they require sodium for excretion, there is additional depletion of the body stores of sodium.

Acute treatment of hypokalemic alkalosis includes administration of large amounts of sodium chloride-containing solutions, and, once diuresis is established, the addition of potassium until the serum potassium level is increased, allowing the kidney to start excreting hydrogen ion. A somewhat uncommon manifestation of profound depletion and extensive vomiting is paradoxical aciduria; this occurs when the extent of potassium depletion is so great that there is no potassium to exchange for hydrogen, and the kidney will then continue to excrete hydrogen rather than potassium ion, compounding the alkalosis. Therapy with potassium alone may not reverse paradoxical aciduria and must include a disposable cation such as ammonium chloride or arginine hydrochloride in addition to potassium. If this treatment fails, administration of 0.1 N hydrochloric acid slowly through a central vein will trigger urinary hydrogen ion excretion and correction of the alkalosis.

Gaseous Distention, Eructation, and Flatulence

A common presenting complaint from patients with gastrointestinal disorders is intestinal gas. Gaseousness cannot be underestimated as a source of social difficulty to patients. Most of these patients are aerophagics; they swallow too much air when they eat, eat too quickly, or talk too much when they eat. There is little question that some patients have demonstrated physiologic changes in which there is greater sensitivity of the stretch receptors to a normal amount of gastrointestinal gas. This may be part of a more generalized motor dysfunction.

Symptoms and signs include gas, bloating, left shoulder pain (splenic flexure syndrome), borborygmi, bowel sounds that the patient believes are audible in public (and they may be), some nausea, dyspepsia, eructation and flatulence. Hyperactive bowel sounds are the rule. Similar symptoms also are present in the early postoperative period but generally resolve spontaneously in 3 to 4 months. Patients may be reassured and treated symptomatically with activated charcoal or simethicone.

Eructation, or belching, usually is due to aerophagia and is relatively normal after a meal to rid the patient of air associated with swallowing. A variety of foods may increase eructation; although these foods vary among patients, usually they include onions, tomatoes, and peppers. Chronic eructation may be an indication of chronic pathology, but it

remains a voluntary phenomenon. Cholelithiasis, peptic ulcer disease, and esophageal reflux disease may be associated with eructation, which partially relieves the symptoms. Anxiety is part of the syndrome, and belching may become habitual. The importance from a surgical standpoint is to make certain that these patients are not operated on needlessly. Patients who do not have decreased LES sphincter pressure or short LES intraabdominal length should not undergo fundoplication, because the combination of fundoplication and aerophagy will result in the patient's being unable to belch (gas bloat syndrome), causing distress.

Small intestine gas usually is present in amounts of less than 1 L, perhaps as little as 200 mL at any given time. Production of gas ranges between 600 to 800 mL per day, which usually is passed anywhere from 10 to 25 times per day. Components of the expelled gas include nitrogen, oxygen (a major source is swallowed air), carbon dioxide (resulting from bacterial fermentation), and methane (a product of bacterial fermentation). Minor components of intestinal gas, probably less than 1 percent, include the volatile short-chain fatty acids, ammonia, hydrogen sulfide, and various volatile amino acids from which the odor is derived. Distention of the colon may reflect chronic partial obstruction secondary to diverticulitis. The splenic flexure syndrome can be manifested as pain in the left shoulder. Hepatic flexure distention is less common.

Lactase deficiency can result in both eructation and flatulence. Excessive flatulence usually is a result of intestinal dysmotility, such as that preceding diverticular disease, or aerophagia from increased amounts of swallowed air. Diet plays a major role; a trial of dietary therapy, eliminating legumes, beans, cabbages, and similar foods may result in improvement. With suspicion of lactase deficiency, lactose-free milk products can be prescribed, or milk can be eliminated from the diet. In patients with diverticular disease who are taking psyllium seed preparations or other soluble fiber, the contained pectin may result in increased production of methane.

If bacterial overgrowth is suspected, a trial of metronidazole or tetracycline is a reasonable therapeutic test. Activated charcoal, simethicone, or both are given as symptomatic therapy. Newer treatments, such as enzymes to aid digestion of legumes and vegetable products, may be helpful in some patients, but the overall results have been disappointing.

Small and Large Intestine

The fate of ingested food is determined by hormonally and neuromuscularly controlled events. Initial digestion is in the stomach. Hyperosmolar gastric contents will not pass the pylorus, and they must be rendered isoosmotic by secretion, at which point the pacemaker begins to fire. The peristaltic waves move down the greater and lesser curve, with rhythmic opening of the pylorus every 30 seconds, transferring 2 to 4 mL of digested chyme. Total gastric emptying is achieved within 3 to 4 hours, and sometimes sooner. Once in the duodenum, 3 to 4 L of biliary and pancreatic secretions per 24 h are added to the duodenal chyme. Additional secretion of 2 L of succus entericus per 24 h occurs in the small bowel. The digested chyme traverses the small intestine, normally at about 1 inch per minute, and its passage through the entire small bowel is complete in 3 to 4 h. Under normal circumstances, protein absorption is complete within the first 120 cm, and carbohydrate absorption within the first 150 to 180 cm. Water absorption takes place mostly in the right side of the transverse colon, with a net water gain in the colon of up to 1,500 mL per 24 h. The remaining stool is approximately 70 percent water. Elimination occurs 24 to 72 h after food ingestion, depending on composition.

Constipation

Constipation is defined as an abnormal retention of fecal matter, a delay in bowel evacuation as compared to the patient's previously normal bowel habits, or stools that are abnormally hard, diminished in quantity, or relatively infrequent. It is estimated that at least \$400 million is spent on laxatives each year, not including the amount spent on evaluation and investigations of constipation. The public perception appears to be that one regular bowel movement daily is normal. In reality, normal bowel movement frequency ranges from three per day to one bowel movement every third day.

One reason for the lack of regularity in the United States is the low fiber consumption. Because processing carbohydrate removes most of the fiber in our diet, the average daily fiber consumption is less than 20 g, stool and average daily weight is about 200 g. In relatively underdeveloped countries, in which diverticulitis is a medical rarity but volvulus is very common, average daily fiber consumption is 75 g, and stool weight approximately 500 g. Investigation is required for a change in bowel habits, but the findings rarely reveal neoplastic disease. Because diverticulosis is endemic in the United States, with close to 60 to 80 percent of the adult population over 60 years of age probably having significant muscular disorders of diverticulosis, these patients appear in physicians' offices more often than with most other disease states.

Physiology of Defecation

The fecal bolus normally is stored in the sigmoid. When defecation is about to occur, longitudinal peristaltic waves move the fecal bolus into the rectum. At the same time, the circular muscle at the rectosigmoid junction relaxes. The normal reservoir function of the rectum also triggers afferent signals via the hypogastric and pelvic nerves to the cauda equina area and where efferent spinal cord impulses are generated. Abdominal and diaphragmatic muscle contraction and voluntary relaxation of the external sphincter usually accompany normal defecation. The external sphincter is a voluntary muscle that receives input from the conus terminalis.

The central nervous system can cause diarrhea or constipation. Parasympathetic nerves augment intestinal motility, while sympathetic stimulation inhibits bowel movements. Anticholinergics and opiates inhibit motility and defecation, whereas cholinergic agents, caffeine, nicotine, potassium, and vasopressin stimulate intestinal motility.

Etiology

(Table 22-9).

Acute constipation is most often mechanical in origin because of intestinal obstruction of the small bowel by volvulus and incarcerated hernia, or diverticulitis and neoplastic disease of the large bowel. Ulcerative colitis or Crohn's disease may be responsible for stricture. In ulcerative proctitis acute constipation might be the principal symptom. Reflex acute constipation may complicate trauma with a retroperitoneal hematoma. In patients with anal fissures, acutely thrombosed hemorrhoids, or after a rectal procedure, acute constipation is common. In neoplastic disease (such as carcinoma of the sigmoid) or diverticulitis, patients may present with acute obstruction.

Constipation can be a result of psychologic factors, dietary constituents, laxatives and drugs, neurogenic causes, decreased skeletomuscular power, and mechanical factors that are intrinsic or extrinsic to the gastrointestinal tract. Psychogenic constipation may be related to improper training, and the symptom frequently dates from early childhood. The eventual result may be a functional megacolon, which is more common than Hirschsprung's disease. Dietary factors include a lack of bulky foods and the use of laxatives, which leads to overstimulation of the bowel, with eventual fatigue. Decreased muscular power in the skeletal muscles of the diaphragm, abdominal wall, and pelvic floor can cause constipation. Weakness of the diaphragm may be associated with a variety of chronic pulmonary diseases, and weakness of the abdominal wall can occur in pregnancy, in the presence of large, rapidly expanding intraabdominal masses, and in patients with marked ascites. Weakness of the pelvic floor usually is a consequence of pregnancy.

The role of atony of the intestinal muscle is difficult to evaluate and may be of minimal importance. Hypokalemia causes ileus. Collagen and endocrine disorders are thought to be associated with intestinal atony. Neurogenic causes include tabes dorsalis, multiple sclerosis, spinal cord tumors, and trauma. These lesions result in deficient reflex activity or directly destroy or depress the autonomic innervation of the intestine. In Hirschsprung's disease, the neurologic deficit in the myenteric and submucosal plexuses interrupts the peristaltic action to that segment. Factors intrinsic to the gastrointestinal tract that contribute to the symptom of constipation include tumors, fecal impaction, intussusception, and volvulus. The mechanical factor is also implicated at the anal sphincter, when spasm and the voluntary avoidance of defecation because of pain occur in patients with hemorrhoids, fissures, or proctitis. Extrinsic causes include large intraabdominal masses such as ovarian cysts, fibroids, pregnancy, and obstructing adhesions.

Clinical Evaluation

The history should focus on previous bowel habits and subsequent changes. Dietary history is essential. Nausea and vomiting may be present. The nature of the changes in stool caliber, color, consistency, blood, mucus, or undigested fat should be determined. Abdominal distention may be present. Physical and rectal examination may reveal a mass. A stool sample to be tested for occult blood should always be obtained. Rigid or flexible proctosigmoidoscopy may define the presence of inflammation, tumors, or melanosis coli (cathartic dependence). If these studies are unrewarding, radiographic examination with barium enema or colonoscopy is indicated. In all cases of acute intestinal obstruction, barium examinations of the colon are essential to determine whether the obstruction is colonic in origin. The exception is toxic megacolon, in which acute obstipation may be present, and acute diverticulitis, in which it is not necessary to perform further diagnostic maneuvers if the diagnosis is clinically apparent.

Chronic Constipation

Chronic constipation can be caused by congenital disorders, motility disorders, functional problems of the defecatory mechanism in the pelvic floor, specific problems of the elderly, medications, diverticulosis with chronic scarring, and neoplastic disease. An early age of onset suggests a congenital cause such as Hirschsprung's disease, requiring a rectal biopsy. Colonic dysmotility, or slow-transit constipation, is a poorly understood colonic motility disorder that affects mostly young men and women but is increasing throughout the population. Stool markers and transit times may reveal the diagnosis. The constipation of irritable bowel syndrome also contributes to this large catch-all area of diagnoses; alternating diarrhea also may be present in this disorder. In the elderly, prolapse can lead to constipation, and, in a vicious cycle, constipation leads to further prolapse. A variety of pelvic floor abnormalities in the elderly result in significant difficulties with constipation. These include failure of the striated muscles of the pelvic floor to relax on straining, failure of the internal sphincter to relax on rectal distention, pelvic floor laxity, rectal intussusception (occult rectal prolapse), acute rectal herniation (rectocele), or deficient or ignored rectal sensation. Medication plays a major role in chronic constipation, such as in laxative abuse or use of sucralfate (for peptic disorders), anticholinergics, and phenothiazines. Chronic diverticulitis, especially repeated episodes, results in narrowing of the sigmoid and consequent alternating diarrhea and constipation, change in stool caliber, and, in many of these patients, a feeling of tenesmus. On physical examination, there usually is tenderness along the sigmoid colon, which in most patients is easily palpated and feels rigid. The most difficult and most important motility disorder to diagnose is neoplasia; the patient may give a history of chronically increasing constipation, stool guaiac usually is positive, and a mass sometimes is palpable.

Directed laboratory examination should include anoscopy, proctosigmoidoscopy, and colonoscopy or barium enema. If indicated, evaluation of pelvic floor abnormalities should be performed, including electromyography, scintigraphic balloon topography, endorectal ultrasonography, rectal defecography, balloon expulsion, and rectal manometry. Pudendal nerve latency testing may rule out some functional abnormalities. After the diagnosis is made, appropriate therapy can be undertaken, including the repair of pelvic floor abnormalities, discontinuation of certain medications, appropriate operation for diverticulitis or neoplastic disease, rectal biopsy if Hirschsprung's disease is suspected, and changes in diet and medications.

Diarrhea

Because the volume of material secreted by the stomach, hepatobiliary tree, and small bowel is large (5 to 8 L per 24 h), diarrhea can result in hemodynamic instability. Five to 10 million individuals die worldwide of acute infectious diarrhea annually, and most are children.

Pathophysiology

There are at least four major pathophysiological mechanisms that can result in excessive loss of fecal water: (1) luminal secretion of solute or water; (2) exudation, or the loss of protein, blood, and mucus; (3) osmotic retention of water; and (4) abnormal or disordered contact between chyme and the absorptive surface of the bowel.

Secretory diarrhea is the result of stimulation of small bowel or colonic mucosa to secrete rather than absorb fluid. The fluid secreted has the composition of extracellular fluid. After ileal resection, bile acids are not absorbed and enter the colon as dihydroxy bile acids, which irritate the colon and cause secretion. Cholestyramine in patients who have had less than 100 cm of ileum resected is effective in preventing the loss of bile salts and the accompanying decrease in bile salt pool, and reducing the diarrhea. Steatorrhea, as in the pancreatic insufficiency of chronic pancreatitis, or improper mixing of bile and pancreatic juice, such as after Billroth II gastrectomy, also allows fatty acids to reach the colon. Binding these fats with calcium decreases the diarrhea.

Pancreatic islet cell tumors may cause diarrhea by two different mechanisms. Approximately 20 percent of gastrinoma patients have diarrhea-based acid hypersecretion. Diarrhea stops with effective nasogastric suction. The Werner-Morrison syndrome includes watery diarrhea, hypokalemia, and achlorhydria or hyperchlorhydria, and alkalosis. Approximately one-third of the patients have hypercalcemia. Vasoactive intestinal peptide (VIP), prostaglandins, pancreatic polypeptide (a marker for islet cell tumors), and peptide histidine leucine have been implicated as etiologic agents. The syndrome also is called pancreatic cholera. The diagnosis is made by elevations in serum VIP and pancreatic polypeptide levels. VIP secretion may be periodic, and several fasting serum VIP values should be obtained. Resuscitation of these patients before operation is difficult because up to 10 L of potassium-rich fluid may be secreted. Octreotide, a long-acting somatostatin analog, decreases serum levels of VIP and ends the diarrhea. More than half of these tumors are malignant, and most are found in the body and tail of the pancreas. In less than 20 percent of the patients, diffuse islet cell hyperplasia is the cause of the syndrome. If the tumor has spread and is malignant, debulking is indicated. A rare somatostatin-producing tumor may present with steatorrhea as a predominant symptom.

Carcinoid tumors are the most common gut endocrine tumors, constituting 55 percent of gut endocrine tumors and 13 to 34 percent of all tumors of the small bowel. They are rare, with an incidence of 1.5 per 100,000 of the general population. Symptomatic carcinoid tumors generally occur in the ileum, appendix, bronchus, or stomach. Carcinoids of the rectum usually are not functional. Products of carcinoids include largely 5-hydroxytryptamine (serotonin) but also histamine and dopamine, tachykinins including kallikrein, substance P, and neuropeptide K, various peptides such as pancreatic polypeptides, chromogranins, neurotensin, human chorionic gonadotropin, and motilin, and prostaglandins.

Almost 30 percent of patients with medullary carcinoma of the thyroid have a secretory diarrhea that is attributed to increased jejunal water and electrolyte secretory stimulation by high plasma calcitonin levels. Plasma calcitonin levels usually are in excess of 1,000 pg/mL. In extensive disease, basal plasma levels may exceed 2,000 to 3,000 pg/mL. If medullary thyroid carcinoma is suspected and calcitonin levels are normal, calcium and pentagastrin stimulation may elicit high calcitonin levels. Treatment is by total thyroidectomy. In the case of multiple endocrine neoplasia (MEN) syndromes secretion of other hormones may result in diarrhea.

Secretory diarrhea also can be caused by various bacterial toxins such in cholera or shigella infection. The toxins of *Vibrio cholerae* activate adenylyl cyclase of the basolateral membrane of mucosal cells, causing a rise in intracellular cyclic adenosine monophosphate concentration, which results in chloride and bicarbonate secretion into the lumen, carrying water passively with it. Other toxins, such as those of *Escherichia coli* and *Cryptosporidium*, have different mechanisms. Laxative abuse, such as ingestion of anthraquinone cathartics, provokes secretory diarrhea.

Exudative diarrhea results from the copious luminal release of serum proteins, blood, or mucus from sites of inflammation, ulceration, or infiltration. Inflammatory bowel disease, parasitic and bacterial infestation (amoeba, salmonella), and infiltrative disease (lymphoma, Whipple's disease) will produce diarrhea.

Disordered contact between the intestinal contents and the absorptive surface has many causes. Bowel resection or bypass, Billroth II gastrectomy, or motility disorders result in altered exposure time that can impair mixing and shorten contact times, not allowing complete digestion and absorption. Lactase deficiency can result in rapid transit and prevent absorption of nutritional moieties. Abnormal motility, as an intrinsic mechanism or as a result of superinfection, prevents sufficient time or contact for reabsorption of the fluid that is secreted into the bowel.

The causes of diarrhea are summarized in Table [22-10](#).

Clinical Evaluation

The history is important in ascertaining whether this is a new, chronic, or recurrent problem. The frequency of the diarrhea and any associated pain or other symptoms may give important clues. The stool should be evaluated: is it foul smelling, or are there fat globules that float on the water? The relationship to food intake is important. Associated symptoms such as fever, pain, nausea, and vomiting suggest an infectious or inflammatory process. If other individuals in the family or community also have diarrhea, this suggests an infectious cause.

Alternating diarrhea and constipation, with thin, ribbonlike stools and tenesmus, suggests diverticular disease or chronic laxative abuse; cancer and partial obstruction are less likely. Variants of irritable bowel syndrome also may be associated with this history.

Mucus secretion occurs with neoplastic disease, cancer, ulcerative colitis, and Crohn's disease, particularly in conjunction with classic postprandial pain. Tenesmus occurs in ulcerative colitis, rectal cancer, and diverticular disease.

A physical examination can reveal abdominal distention or a mass lesion, suggesting neoplasm or inflammation. The area around the anus should be carefully inspected for perianal fistulas. Examination also might reveal evidence of systemic disease, such as collagen diseases, endocrinopathies including Addison's disease, hypoparathyroidism, or hyperparathyroidism, or medullary carcinoma of the thyroid.

Fever suggests an inflammatory process, and rarely neoplastic disease, except with lymphoma, in which fever is common. Various skin disorders, such as pyoderma gangrenosum or erythema nodosum, and arthritis suggest inflammatory bowel disease. Acanthosis nigricans strongly suggests neoplasm. Digital examination of the rectum may reveal carcinoma, granulomas, lymphogranuloma venereum, diffuse polyposis, stricture, or fecal impaction. Stool should always be tested for occult blood.

General and specific laboratory evaluations should be performed. Since protracted diarrhea results in fluid and electrolyte loss, it can result in hypovolemia, acidosis, renal failure, and death if the patient is not replenished. Hematocrit concentration, white blood cell count, liver chemistry studies, and determinations of levels of electrolytes, calcium, phosphorus, and especially magnesium should be obtained. Zinc is almost certainly depleted in chronic diarrhea, which may result in loss of taste as well as inability to carry out certain essential enzymatic functions; it is difficult to measure, however. Malnutrition also may be present. Replenishment of the patient with chronic diarrhea while the work-up is proceeding may involve not only hydration but also the placement of a central venous catheter for total parenteral nutrition.

Once these examinations have been completed, anoscopy and rigid or flexible proctosigmoidoscopy should be performed to rule out Crohn's disease, lymphogranuloma venereum, amebiasis, and chronic ulcerative colitis. Rectal mucosal friability suggests ulcerative colitis or amebiasis; rectal sparing suggests Crohn's disease. Levels of fecal leukocytes and of *Clostridium difficile* toxin should be obtained. If steatorrhea is suspected, a 72-h quantitative fat excretion should be obtained. Microscopic examination for ova and parasites is particularly pertinent to the diagnosis of infestations, but if *Giardia* is suspected because of recent travel, duodenal aspiration is necessary. If these tests are negative, barium enema or colonoscopy and biopsy are appropriate.

Barium enema or colonoscopy will reveal inflammation, mucosal abnormalities, and villous adenoma with excessive secretion. Colonoscopy can uncover the terminal ileum regional enteritis or, rarely, tuberculous enteritis. Gastrocolic fistula, blind-loop syndrome, or inadvertent gastroileostomy may be diagnosed by upper gastrointestinal series. CT scanning may reveal chronic pancreatitis and pancreatic carcinoma. Laparotomy or laparoscopy are considered a diagnostic procedure. Small-bowel biopsy is indicated in the differential diagnosis of malabsorption syndrome, regional enteritis, Whipple's disease, or lymphoma.

Other diseases that can result in diarrhea include partial intestinal obstruction, Zollinger-Ellison syndrome (in which 20 percent of the patients have diarrhea), endocrine tumors, including those that may secrete vasoactive intestinal peptide, glucagon (in which there is a characteristic migrating rash), somatostatin, and substance P. In the Zollinger-Ellison syndrome, a nasogastric tube will reveal high acid output, and nasogastric suction should result in cessation of the diarrhea.

Treatment

Treatment of diarrhea depends on an accurate diagnosis. *C. difficile* should be treated with vancomycin or high doses of metronidazole. Crohn's disease and ulcerative colitis should be treated with asulfidine or its congeners, steroids, antibiotics, or anti-inflammatories. Nonspecific therapy includes opiate drugs, such as atropine with diphenoxylate, loperamide, paregoric, and probably the most potent of all, codeine. Kaopectate may be helpful. Bismuth subsalicylate (Pepto-Bismol) is effective against *Helicobacter pylori*, which probably does not cause diarrhea, and against some cholera, *E. coli*, and shigella species, which do. Other agents include prostaglandin synthetase inhibitors, such as indomethacin for some of the peptide-secreting tumors, verapamil, and somatostatin, as well as the new families of α_2 -adrenergic antagonists such as clonidine.

Intestinal Obstruction

Etiology

Intestinal obstruction accounts for at least 20 percent of all admissions to a surgical service. More than 9,000 deaths result from intestinal obstruction and operation annually. There are many ways to classify intestinal obstruction, among them extraluminal (including adhesions and neoplastic disease), intraluminal (such as gallstone ileus or stricture), and intramural (such as a Crohn's disease process). It is easier to divide intestinal obstruction into those patients who have had no previous abdominal surgical procedures, and those who have. The importance of previous abdominal procedures is that obstruction of the small bowel secondary to postoperative adhesions makes up 60 to 80 percent of admissions for intestinal obstruction; hernias are a distant second at 15 to 20 percent, and malignant tumors, intramural (which are rare) and extramural (which are far more common), make up 10 to 15 percent of admissions. Among patients who have had no previous operation, obstruction of the esophagus is possible by inflammatory stricture and neoplasm, or functional obstruction, such as achalasia. Obstructing duodenal ulcer comprises a large category of patients. Pancreatic cancer sometimes results in obstruction of the duodenum, although this is rarely a presenting symptom. Small-bowel carcinoma and carcinoma of the duodenum also are relatively rare. In the elderly patient with no previous operation who presents with intestinal obstruction, there is a possibility of gallstone ileus. Crohn's disease results in obstruction from the mass, from acute inflammation, and from an inflammatory or fibrotic stricture. The small bowel or, rarely, the colon may become incarcerated or strangulated in a hernia, and small-bowel volvulus can occur because of a narrowly based mesentery, although this is exceedingly rare. Colonic obstruction most often arises from cancer (60 percent), diverticulitis (15 percent), or volvulus, especially in the elderly.

Patients with previous intraabdominal procedures have a 5 percent incidence of intestinal obstruction secondary to adhesions. This does not increase with the number of previous procedures for lysis of adhesions unless the operation has been unsuccessful or incomplete. Postoperative obstruction usually occurs 4 to 5 days postoperatively, although most of these patients have difficulty from the first postoperative day.

The causes of intestinal obstruction are outline in Table [22-11](#).

Partial Obstruction and Pseudoobstruction

When obstruction is not total, chronic nausea and vomiting can occur. Distention often is present. Localized distention also may occur, such as left upper quadrant distention with gastric outlet obstruction. Bowel sounds reflect amphoric fluid-filled bowel loops when the obstruction is intestinal, or a succussion spasm when the obstruction is at the pylorus or duodenum. Other causes of chronic intestinal obstruction include Crohn's disease and motility disorders, of which the most common are diabetes and postvagotomy gastroparesis, resulting in a functionally obstructed gastropathy and bezoar formation.

Chronic intestinal pseudoobstruction is a heterogeneous clinical syndrome with a variety of causes and clinical presentations. The underlying process involves intestinal motility that is impaired from abnormalities in the intestinal muscle (myopathic) or the nervous system (myenteric neuropathic). Various disease processes may affect muscles and nerves. These diseases usually are diffuse, extending throughout the gastrointestinal tract.

The spectrum of altered motility includes a loss of normal peristalsis and transport, loss of the migrating motor complex, and spontaneous or uncoordinated contractile activity. Myopathic conditions cause hypomotility, with low-amplitude pressure activity. Neuropathic pseudoobstruction, which is far less common, is manifested by excessive uncoordinated activity. The presentation is varied. It may complicate a series of diseases or situations, including endocrine disorders such as myxedema and diabetes mellitus, chronic infection such as cytomegalovirus (especially common in immunocompromised patients) and Chagas' disease, infiltrative diseases such as amyloidosis, vasculitides such as dermatomyositis, progressive systemic sclerosis, systemic lupus erythematosus, various neurologic disorders such as multiple sclerosis, and muscular dystrophy, porphyria, various injuries, paraneoplastic syndromes, and a familial form. In the largest group of patients the cause is not known.

The presentation may be spontaneous or postoperative. Recurrent attacks of nausea, vomiting, crampy abdominal pain, and abdominal distention initially are variable in frequency, intensity, and duration. Constipation and diarrhea may occur. The recurrent episodes become more common until they are chronic, with abdominal distention, bacterial overgrowth, malabsorption, diarrhea, and weight loss. Anorexia may complicate the syndrome. In the postoperative situation, which is potentially the most dangerous, this usually involves patients who have undergone vagotomy or another type of intestinal operation and are thought to be obstructed in the postoperative period. The point of obstruction is never found at reoperation or in radiographic investigations. In the absence of recognition of the true cause, multiple reoperations may take place, often with disastrous results, with multiple enterotomies, bowel resections, leakage, and fistulas, leaving the patient as a small-bowel cripple on permanent total parenteral nutrition, habituated to drugs, and with chronic psychiatric abnormalities.

Early recognition of pseudoobstruction is critical to proper therapy. No specific laboratory studies aid in the diagnosis. One should rule out hypokalemia, hypomagnesemia, and hypoalbuminemia. Other tests include thyroid function tests, antinuclear antibodies, serologic markers, and studies for vasculitis, such as lupus. Manometric studies of the upper and lower intestine can be helpful if the capacity to carry them out exists. Infrequent low-amplitude contractions are characteristic of smooth muscle degeneration, and disordered cluster contractions are characteristic of nervous degeneration. Biopsy of the intestine may or may not reveal loss of myenteric plexi. Full-thickness rectal biopsy may be helpful if Hirschsprung's disease is suspected.

On radiographic examination there is prolonged transit time of barium to the stomach and colon, often as long as 5 h. Fluoroscopy will confirm decreased motility. Radionuclide-labeled pellets confirm the slow transit time and can be used to assess the effects of therapy.

Various drugs can cause pseudoobstruction. Tricyclic antidepressants, opiates, antihistamines, beta-adrenergic agonists, quinidine, and especially amitriptyline should be avoided if possible. Total parenteral nutrition usually is necessary for patients with chronic pseudoobstruction. Operation should be avoided unless a specific site of obstruction has been identified or there is a short segment amenable to resection.

A variety of operations have been proposed, including partial duodenectomy, tailoring of massively dilated segments, total colectomy, imbrication or excision of the antimesenteric border of massively dilated small bowel, and total abdominal colectomy. None of these has shown much efficacy. Radical resection and transplantation have been performed but have not been applied in large numbers of patients. The most successful treatment of idiopathic pseudoobstruction is subtotal colectomy with ileorectal anastomosis at the peritoneal reflection, which has been useful for patients in whom the pseudoobstruction is identified as primarily colonic. Palliative procedures include tube enterostomy, which may be opened periodically, and tube gastrostomy. If a site of obstruction cannot be identified, pseudoobstruction should be diagnosed; the patient should be placed on total parenteral nutrition, bacterial overgrowth (when suspected) should be treated, and no operation should be performed. Cisapride or metoclopramide may provide some relief.

Mechanical Obstruction Pathophysiology

The principal physiologic derangements of the mechanically obstructed intestine with intact blood supply are accumulation of fluid and gas above the point of obstruction and altered bowel motility, which leads to systemic derangements. Death from intestinal obstruction was for many years attributed to "toxins" that were absorbed from the intestine. In 1919 Hartwell and Hoguet were able to prolong the life of dogs with high intestinal obstruction by daily parenteral administration of physiologic saline solution. Gamble later demonstrated that the "toxic" factor in simple mechanical obstruction was the loss of fluid and electrolytes from the body by vomiting and by sequestration in the obstructed bowel.

Accumulation of large quantities of fluid and gas within the lumen of the bowel above an obstruction is progressive. The net movement of a substance across the intestinal mucosa is equal to the difference between the unidirectional flux from intestinal lumen to blood (absorption) and the opposite flux from blood to lumen (secretion). Accumulation of fluid within the bowel, a negative net flux, results if the flux from lumen to blood (absorption) is decreased or if the flux from blood to lumen (secretion) is increased. After 48 h of obstruction, the rate of entry of water into the intestinal lumen increases as a consequence of blood to lumen flux. The findings for sodium and potassium are parallel to those of fluid.

Normal fluxes occur in the direction of blood to lumen, but fluxes from lumen to blood are depressed or abolished in an obstructed ileal segment. As a result, water, sodium, and chloride (and presumably other ions) move into the obstructed intestinal segment but not out of it, distending it with fluid that has approximately the electrolyte composition of plasma.

Wright and associates studied net flux in patients with mature ileostomies. Closed loops were produced by proximal and distal obstructing balloons inserted through the ileostomy. Absorption of a test solution was found to increase at moderate elevations of pressure, but fell below normal at pressure three or four times normal. Conversely, secretion of fluid into the lumen increased progressively as pressure rose. They concluded that increased secretion is the primary cause of fluid loss and distention in intestinal obstruction, with decreased absorption playing a lesser role. Prostaglandin release in response to bowel distention is thought to be the mechanism by which secretion into obstructed loops is increased.

The bowel immediately above the obstruction is the most affected initially. It becomes distended with fluids and electrolytes, and circulation is impaired. With increasing intraluminal pressure, the fluid is dispersed orad until it reaches bowel that is still capable of absorption. When obstruction has been present for a long time, the proximal portions of the intestine also lose their ability to handle fluid and electrolytes, and the entire bowel proximal to the obstruction becomes distended.

A second route of fluid and electrolyte loss is *into* the wall of the involved bowel, accounting for the boggy edematous appearance of the bowel often seen at operation. Some of this fluid exudes from the serosal surface of the bowel, resulting in free peritoneal fluid. The extent of fluid and electrolyte loss into the bowel wall and peritoneal cavity depends on the extent of bowel involved in venous congestion and edema, and the length of time before the obstruction is relieved.

The most obvious route of fluid and electrolyte loss is by vomiting or gastrointestinal tube after treatment is initiated. The aggregate of these losses (1) into the bowel lumen, (2) into the edematous bowel wall, (3) as free peritoneal fluid, and (4) by vomiting or nasogastric suction, rapidly depletes the extracellular fluid space, leading progressively to hemoconcentration, hypovolemia, renal insufficiency, shock, and death unless treatment is prompt.

Much of the distention of the bowel above a mechanical obstruction can be accounted for by the fluid sequestered within the lumen. Intestinal gas also is responsible for distention. The approximate composition of small intestine gas (Table [22-12](#)) shows that the basic composition is swallowed air to which small amounts of gases not found in the atmosphere have been added.

Gases are absorbed from the intestine at rates that are directly related to the partial pressure of the particular gas in the intestine, in the plasma, and in the air breathed. In the case of nitrogen, there is little diffusion, since the partial pressures of the gas are virtually the same in intestine, plasma, and air. On the other hand, carbon dioxide diffuses very rapidly, because the partial pressure of carbon dioxide is high in the intestine, intermediate in plasma, and very low in air. For this reason, though carbon dioxide is produced in large amounts in the intestine, it contributes little to gaseous distention.

With obstruction of the lumen, peristalsis increases in an attempt to overcome the obstruction. After a short time, continuous peristalsis above the obstruction gives way to regularly recurring bursts of peristaltic activity interspersed with quiescent periods. The duration of the quiescent period is related to the level of the obstruction in the gastrointestinal tract; it is 3 to 5 min with high obstruction, and 10 to 15 min with lower ileal obstruction. These muscular contractions may be of sufficient magnitude to traumatize the bowel and contribute to the swelling and edema of the bowel wall. As bowel above the obstruction distends, bowel below the obstruction becomes progressively more quiet. This results from an inhibitory reflex initiated by distention of the bowel above.

Strangulated Obstruction

Occlusion of the blood supply to a segment of bowel in addition to obstruction of the lumen usually is referred to as strangulated obstruction. Interference with the mesenteric blood supply is the most serious complication of intestinal obstruction. This frequently occurs secondary to adhesive band obstruction, hernia, and volvulus.

The accumulation of fluid and gas in obstructed loops and the altered motility seen in simple mechanical obstruction are rapidly overshadowed by the consequences of blockage of venous outflow from the strangulated segment—extravasation of bloody fluid into the bowel and bowel wall. In addition to the loss of blood and plasmalike fluid, the gangrenous bowel leaks toxic materials into the peritoneal cavity. These have been variously identified as exotoxins or endotoxins, or toxic hemin breakdown products.

The consequences of strangulated obstruction are related to several factors. The toxic contents are produced by bacteria. In order for the toxins to have an effect, the mucosa must be disrupted and the toxins must pass into the circulation. Damage to intestinal vessels expedites absorption. Symptoms are a manifestation of the absorbed toxins.

Closed-Loop Obstruction

When both afferent and efferent limbs of a loop of bowel are obstructed, it is referred to as closed-loop intestinal obstruction. This is a dangerous form of obstruction because of the propensity for rapid progression to strangulation of the blood supply before the usual manifestations of intestinal obstruction become obvious. Interference with blood supply may

occur from the same mechanism that produced obstruction of the intestine—e.g., twist of the bowel on the mesentery or extrinsic band—or from distention of the obstructed loop. The secretory pressure in the closed loop rapidly reaches a level sufficient to interfere with venous return from the loop. Widespread distention of the intestine usually does not occur, and so neither does abdominal distention.

Colon Obstruction

The effects on the patient with colon obstruction usually are less dramatic than the effects of small bowel obstruction. Colon obstruction, with the exception of volvulus, usually does not strangulate. Also, because the colon is principally a storage organ with relatively minor absorptive and secretory functions, fluid and electrolyte sequestration progresses more slowly. Systemic derangements are of less magnitude and urgency than in small bowel obstruction.

Progressive distention is the most dangerous aspect of nonstrangulated colon obstruction. If the ileocecal valve is incompetent, partial decompression of the obstructed colon may occur by reflux into the ileum, but if the ileocecal valve is competent, the colon becomes essentially a closed loop. If the obstruction is not relieved, distention progressing to rupture of the colon is possible. The cecum is the usual site of rupture, because it is the segment of the colon with the largest diameter. According to the law of Laplace, the pressure required to stretch the walls of a hollow viscus decreases in inverse proportion to the radius of curvature. Thus, given an equal pressure throughout the colon, the greatest distention will occur in the portion of the colon with the largest radius, the cecum (Fig. [22-5](#)).

Clinical Manifestations

Acute intestinal obstruction, when it is due to volvulus around a mechanical adhesion, usually begins with a sudden onset of abdominal pain. The distance of the obstruction from the ligament of Treitz can be ascertained by determining how long after the onset of pain vomiting takes place, and by the nature of the vomitus: the more distal the obstruction, the more feculent its character. The frequency of crampy abdominal pain and the periods between the abdominal pain may give some hint of the level of obstruction. In the proximal small intestine, the serial cramps may be 3 to 5 minutes apart, whereas more distally they may be at 10- to 15-min intervals.

The four cardinal symptoms and signs of intestinal obstruction are crampy abdominal pain, nausea and vomiting, obstipation, and abdominal distention. Simple mechanical obstruction results in the accumulation of succus entericus above the obstruction, with deranged motility as the bowel distends. Fluid that usually is absorbed in the small bowel tends to be secreted, accumulates, and is lost to the circulation. Vomiting or nasogastric suction may result in hypovolemia. In the absence of resuscitation, renal insufficiency, at times shock, leading to cardiovascular collapse and death can occur.

The principal findings of localized tenderness progressing to rebound tenderness accompanied by fever, tachycardia, and leukocytosis suggest that bowel viability is compromised. Initially abdominal pain is crampy, coincides with peristaltic rushes, and is poorly localized. If progressive loss of bowel viability occurs, the pain evolves to a steady, generalized abdominal discomfort with a local point of rebound tenderness. There is no absolute way to distinguish between strangulated and nonstrangulated obstructions.

In early postoperative obstruction, abdominal distention may not be apparent, particularly in high obstruction. In most other patients with intestinal obstruction, progressive abdominal distention occurs. Tenderness is mild and there usually is little voluntary muscle guarding. Percussion reveals the absence of fluid, and on auscultation loud, high-pitched, and amphoric rushes of peristalsis are noted, with the frequency depending on the location of the obstruction. As mechanical obstruction becomes progressive, hyperperistalsis may give rise to less frequent peristaltic rushes, and the abdomen becomes quieter.

Laboratory Findings

Intestinal obstruction usually results in losses of 4 to 8 L of intravascular and extracellular fluid into the small bowel, with consequent compromise of the cardiovascular system. Elevation of blood urea nitrogen and creatinine levels, hemoconcentration, hyponatremia, and hypokalemia are the most common early laboratory signs. Urinary specific gravity of 1.025 to 1.030 is the rule; proteinuria or mild acetonuria may be present. Sodium-free water from proteolysis and catabolism of fat will partially replace the intravascular deficit, but hyponatremia, hypochloremia, and hypoosmolality will result. Urine volume gradually increases, though not to normal, with excretion of potassium, including the potassium freed by cellular catabolism. The previously noted progressive increase in the hematocrit concentration is halted or reversed by the ingress of endogenous water. Acid–base effects are determined by the nature of the fluid lost. Metabolic acidosis from the combined effects of dehydration, starvation, ketosis, and loss of alkaline secretions is most common. Metabolic alkalosis occurs infrequently, principally because of loss of highly acid gastric juice. With great distention of the abdomen, the diaphragm may be sufficiently elevated to embarrass respiration, resulting in carbon dioxide retention and respiratory acidosis.

Simple mechanical obstruction is accompanied by a modest increase in the number of leukocytes with some shift to the left. White blood cell counts of 15,000 to 25,000/mm³, and marked polymorphonuclear predominance with many immature forms, strongly suggest that the obstruction is strangulated, but this is not a sensitive indicator. Very high white cell counts, such as 40,000 to 60,000/mm³, suggest primary mesenteric vascular occlusion.

Serum amylase level elevations may occur in intestinal obstruction. Amylase gains entry to the blood by regurgitation from the pancreas because of back pressure in the duodenum, or by peritoneal absorption after leakage from dying bowel. No preoperative clinical parameters, including the presence of continuous abdominal pain, fever, peritoneal signs, leukocytosis, acidosis, hyperamylasemia, or any combination of these, are sensitive indicators of vascular compromise. The preoperative assessment of the presence or absence of strangulation is correct in only about 70 percent of cases.

Management

When a patient with intestinal obstruction is admitted, a complete set of blood chemistry tests, including electrolytes, amylase, calcium, phosphorus, magnesium, and lactic acid, should be drawn. After the nasogastric tube has been placed and resuscitation begun, anteroposterior, lateral, and upright and chest films, and abdominal films are obtained. The presence or absence of free air, the bowel gas pattern, whether there is stool in the colon, and whether there is air in the rectum can be evaluated. If a rectal examination has been performed, air in the rectum may have no significance. Many of these patients have had a reflexive bowel movement early after the onset of intestinal obstruction, and a bowel action means relatively little prognostically. If there is air throughout the colon, the diagnosis is likely partial intestinal obstruction, which may be treated expectantly with suction, as the condition often resolves spontaneously.

While placement of a Miller-Abbott tube (a long tube) in intestinal obstruction usually results in the amelioration of postoperative obstruction, there is controversy about whether a long tube offers any advantage other than in the postoperative period in the presence of acute small-bowel obstruction. The patient can be observed carefully to allow nasogastric or, when appropriate, long-tube suction to take effect so that, if possible, operation can be avoided. At the same time, operation should not be delayed if there are symptoms or signs of vascular compromise or the patient is getting sicker.

Once the diagnosis of mechanical obstruction is made, additional workup is indicated in the form of a barium enema to make certain that the obstruction is not in the colon. If barium can be refluxed into a collapsed distal ileum, it will aid in the diagnosis. If acute diverticulitis is suspected, contrast examination of the colon is contraindicated. It is recommended that a large amount of barium should not be used in an adynamic small bowel above an obstruction. Water-soluble contrast examinations of large or small intestine are of limited value because of lack of mucosal detail. Because they are hypertonic and draw the fluid in from the intravascular and extravascular space, water-soluble contrast materials can further distend the bowel and may add to the intravascular fluid deficit in the patient with intestinal obstruction.

Only 10 to 20 percent of patients with partial small-bowel obstruction come to operation. Nasogastric decompression suffices except in metastatic neoplastic disease, in which long-tube decompression may be more effective. In early postoperative obstruction, reoperation is necessary in only about 20 percent of patients, especially when a long tube is used. An average decompressive period of 4 to 6 days is necessary. When the patient's postoperative obstruction clears, the suction should be discontinued and the patient fed around the tube, only removing it when it is clear that the obstruction has totally resolved.

In the case of neoplastic disease and suspected metastasis as the cause of intestinal obstruction, operation is indicated only in those patients with a good performance status. Simple nasogastric or long-tube decompression is appropriate for approximately 3 days. If no benefit is achieved, then operation is indicated. A mortality rate of at least 10 percent can be expected; the morbidity rate approaches 50 percent. A gastrostomy tube should be placed if it is believed that the obstruction will recur.

Patients with simple mechanical obstruction who are operated on within the first 24 h of the disease do not need extensive preoperative preparation, because water and salt depletion and distention usually are not serious at this stage. After the history and physical examination have established the presumptive diagnosis, laboratory studies should be done, intravenous repletion completed, decompression started with a nasogastric sump tube in the stomach, and abdominal and chest x-ray films taken on the way to the operating room. This whole process should take less than 2 h. The mortality rate is less than 1 percent for patients with simple mechanical obstruction who are operated on within the first 24 h.

If the obstruction has been present for more than 24 h when the patients is first seen, depletion and distention may be severe. If strangulation or closed-loop obstruction seems unlikely, the patients is best served by a period of preparation before the obstruction is surgically relieved. In general, the longer the obstruction has existed, the longer it will take to prepare the patient for surgical treatment. To prepare patients with moderate derangements, particularly hypokalemia, usually requires 6 to 12 h; in patients with severe problems, up to 24 h may be necessary, because of the hazard of giving intravenous potassium ion faster than it can equilibrate.

With the possible exception of patients with early simple mechanical obstruction, all patients with intestinal obstruction should have a venous catheter threaded into the superior vena cava or pulmonary artery for frequent measurements of the *central venous pressure (CVP)* or wedge pressure. This line also is used for rapid administration of fluid, and an indwelling catheter should be inserted into the bladder for accurate measurement of the urinary output.

The initial hematocrit reading may be used to estimate the extent of extracellular fluid loss and the volume necessary for restoration. For example, if the hematocrit level has risen to 55 percent, this indicates a loss of approximately 40 percent of the plasma and extracellular fluid volume.

If acid gastric juice loss is prominent, normal saline solution is used; otherwise lactated Ringer's solution and 5% dextrose in water in about equal proportions are preferred to replace the lost fluid and to cover maintenance fluid needs. Potassium chloride also will be necessary but should not be given until a good urinary output is established. Antibiotics should be administered in generous dosage and may be added to the intravenous fluids.

The rate of fluid administration is best controlled by monitoring the CVP or pulmonary artery wedge pressure. Fluids may be given rapidly as long as the CVP remains below 10 to 12 cmH₂O. The end point of volume replacement is indicated by a sudden rise in the CVP or left atrial pressure. Other guides are return of skin turgor and the hourly rate of urine output. The goal of electrolyte concentration and acid-base balance is restoration of these to, or close to, the normal range by the time the volume deficit has been repaired. This usually is possible in patients with reasonably normal renal and pulmonary function.

When the possibility of strangulation exists, preoperative restoration to fluid-electrolyte normality is not possible or advisable. This is an emergency situation requiring vigorous preparation with fluids and electrolytes, antibiotics, nasogastric suction, and an operation at the earliest possible moment to remove the cause of the strangulation or nonviable bowel. Despite application of these principles, the mortality rate in strangulated obstruction is about 25 percent.

Operative Procedure

If operation is required, the critical aspect of the treatment of intestinal obstruction is that it be carried out at a time when obstruction is *simple*, and a lysis of adhesions, reduction of the volvulus, or other therapeutic maneuvers are sufficient. If one delays operation and compromised small bowel or closed-loop obstruction develops, the situation becomes much more dangerous to the patient, and mortality rates rise precipitously. Proper timing of the operation for intestinal obstruction is essential. There are four types of obstruction in which the operation should be done as soon as possible after admission: strangulation, closed-loop obstruction, colon obstruction, and early simple mechanical obstruction.

General anesthesia is safest for the patient. Endotracheal intubation, occasionally performed under local anesthesia, is particularly indicated to prevent aspiration for regurgitated gastric content. Local anesthesia should be used only when the surgeon knows the cause of the obstruction and plans a limited procedure, such as transverse colostomy.

A generous incision should be used to enable the surgeon to lyse all of the adhesions, including those that might be present in the pelvis. A long midline incision is preferred by most surgeons, but it may be possible to carry out a long transverse incision if that has been the previous approach, provided access can be gained to the pelvis. If the adhesions are extensive, all adhesions must be taken down; after they are all taken down, the patient must be checked carefully for inadvertent enterotomies. In compromised bowel, bowel resection and end-to-end anastomosis should be performed unless the disparity in the lumina is too great, in which case an end-to-side anastomosis should be undertaken. If there is encased bowel in the pelvis, colostomy or ileostomy may be necessary; the preoperative discussion with the patient should include this possibility.

Surgical procedures for the relief of intestinal obstruction are divided into five categories:

- (1) Procedures not requiring opening of bowel —lysis of adhesions, manipulation-reduction of intussusception, reduction of incarcerated hernia
- (2) Enterotomy for removal of obturation obstruction —gallstone, bezoars
- (3) Resection of the obstructing lesion or strangulated bowel with primary anastomosis
- (4) Short-circuiting anastomosis around an obstruction
- (5) Formation of a cutaneous stoma proximal to the obstruction —cecostomy (rarely performed), transverse colostomy

On opening the peritoneum, the presence or absence of free peritoneal fluid should be noted as well as the appearance of the fluid. Bloody fluid denotes strangulation, and clear straw-colored fluid is found with simple obstruction. The point of obstruction is best found by starting in the right lower quadrant. If the cecum is grossly distended, the obstruction is in the colon. If collapsed small bowel is found, this is followed back to the point of obstruction, avoiding evisceration of the proximal distended loops.

The surgeon is sometimes faced with the difficult decision of whether or not to resect a loop of intestine of questionable viability. Before release, the strangulated loop of viable bowel has a dull purple-red appearance and is devoid of motion. After release, there is a dramatic color change to bright red in an obviously viable loop, as well as a return of peristalsis. In obviously dead bowel there is no color change and no motion after release of the strangulating obstruction. The loop that only partially “pinks up” and has little or no motion is a problem. It is usually best to wrap the questionable segment in moist laparotomy pads and leave it completely undisturbed for 10 min. If the circulation is obviously better at the end of this time, the loop is replaced in the abdomen. If the viability of the segment is still in doubt, resection should be done. Fluorescein staining and Doppler evaluations have been used to distinguish between viable and dead bowel, but results are not consistent. If a very long segment of bowel is involved, then an attempt should be made to restore flow in the larger vessels supplying the segment. If this is unsuccessful, the surgeon probably should accept the risk of replacing nonviable bowel. After operation the patient is closely observed; if evidence of progressive toxicity develops, reoperation and resection are done. Reexploration and reevaluation of the status of the bowel about 24 h later may be advisable.

Decompression of grossly distended intestine during the operative procedure sometimes is necessary, particularly in late simple mechanical obstruction. Operative decompression is still a contentious point. The site of obstruction is more easily found, the uncontrolled eventration of distended loops through the incision is avoided, the bowel can be returned to the peritoneal cavity without the kinks that may cause segmentation and postoperative obstruction, and closure of the incision is facilitated. Relief of distention also improves the blood supply to the intestine, and peristalsis soon returns. Normal mucosa is impermeable to these toxins, but permeability is affected by impairment of the blood supply, and absorption can occur in compromised bowel.

Operative aspiration can be done in a variety of ways. Multiple needle aspirations are ineffective and increase the morbidity; studies have shown a wound infection rate of 20 percent versus a rate of 4 percent in a comparable group of patients without needle aspiration. Decompression with an ordinary suction tip through multiple enterotomies, though effective, is attended by an increased infection rate, plus the risk of small-bowel fistula. An effective, safe method of decompression is passing a tube from above downward, so that the entire gastrointestinal tract proximal to the obstruction is decompressed. A firm tube with a generous lumen (Baker tube) is introduced through a proximal jejunostomy. It may be preferable to pass the tube transnasally or through a gastrostomy and thread it into the small intestine. The tube is advanced by manipulation through the intact bowel until the entire length of involved bowel down to the obstructed segment is pleated on the long tube. The tube is secured in that position for postoperative decompression.

Postoperative Care

The principles of postoperative care are the same as the preoperative preparation of the patient with obstruction: fluids and electrolytes, antibiotics, gastrointestinal decompression, and, if postoperative ileus is prolonged, parenteral nutrition.

Fluid and electrolyte management is more difficult in the postoperative intestinal obstruction patient than in the usual postoperative abdominal surgical patient because of the large third space of sequestered isotonic fluid. There is continued loss in the immediate postoperative period into the sequestered fluid space. The rate of loss slows and is reversed after a variable period, usually about the third postoperative day. This large autotransfusion as fluid is picked up by the vascular compartment from the sequestered fluid must be allowed for in planning the daily ration of intravenous fluid therapy, because the patient might be watered into congestive failure. Serum sodium, potassium, and magnesium levels must be closely watched and kept in the normal range. A deficit of any of these ions, especially potassium and magnesium, is associated with prolonged paralysis of the gastrointestinal tract.

Decompression of the gastrointestinal tract also is more difficult than in the usual postoperative patient, because restoration of normal propulsive intestinal motility usually is significantly delayed after release of intestinal obstruction. Bowel function usually resumes about the third day after abdominal operation, but after intestinal obstruction it is often 5 or 6 days before gastrointestinal decompression can be discontinued. Because of this, gastrostomy is often desirable for patient comfort.

Ileus

Ileus can be divided into three types. The most common is *adynamic or inhibition ileus*, in which motility is diminished or absent because of inhibition of normal neuromuscular activity. *Spastic ileus* refers to bowel that is tightly contracted without coordinated propulsive activity. *Ischemic ileus* is the result of either low-flow (nonocclusive ischemia) or vascular occlusion, in which coordinated motility is impossible because of the dying musculature.

Spastic ileus is seen in heavy-metal (lead) poisoning and in porphyria. Coordinated hyperactivity of the intestine may return in crampy or colicky pain, particularly in children. Therapy is directed at the primary disease. Adynamic ileus, the most common form of ileus, occurs after a variety of abdominal operations. In the normal postoperative course, the small bowel continues to function. This is the basis of early small-bowel feeding using catheter jejunostomy. Gastric ileus and colonic ileus last for approximately 2 days and 3 to 4 days, respectively, and intake can be resumed on the second or third postoperative day, or sooner after minimally invasive surgery.

If ileus is prolonged, possible causes should be evaluated: metabolic, septic, mechanical, or inflammatory. Of the metabolic causes, drug-induced ileus resulting from narcotics is the most common. Morphine or meperidine given intravenously, by patient controlled analgesia (PCA), or even into epidural catheters, can result in prolonged ileus. Ileus may be prolonged by the use of epidural opiates. Opiates should be limited to the first 24 h; thereafter pure bupivacaine will give adequate analgesia.

Metabolic causes of ileus include hypokalemia, hyponatremia, and hypomagnesemia. Mechanical causes include hematoma at the anastomosis and within the mesentery, a leak from the anastomosis, or an infection adjacent to the anastomosis or in the abdomen. Most surgeons believe that the more extensive the laparotomy and the longer the operation, the more profound the ileus, but several studies have failed to correlate the degree of intraoperative bowel manipulation, the duration of operation, and the duration of ileus. Intraperitoneal and retroperitoneal inflammations such as pancreatitis or peritonitis, or any intraabdominal infection, prolong ileus. Other situations can contribute to an adynamic ileus. A retroperitoneal hematoma, spinal fractures, rib fractures, and pelvic fractures also may contribute, the latter presumably because of the peripelvic hematoma.

Clinical Manifestations

There is no specific period for duration of “normal” postoperative ileus. Instead of passing flatus and becoming hungry on about the fourth postoperative day, abdominal distention persists and there is tympany and scattered occasional bowel sounds. Metabolic and treatable causes such as anemia, sepsis, hyponatremia, hypokalemia, hypomagnesemia, hypoalbuminemia, and hypoosmolality should be identified and corrected.

Radiologic examinations can help to outline the differential diagnosis between postoperative ileus and mechanical obstruction (Table 22-13). Enteroclysis is especially useful. In adynamic ileus some contrast material should reach the cecum in 4 h; if the barium column is stationary for 3 to 4 h, mechanical obstruction is present. Postoperative mechanical obstruction usually responds to long-tube decompression.

Management

Treatment of ileus is supportive, with nasogastric suction or gastrostomy drainage, electrolyte replacement, and, if prolonged, parenteral nutrition. A long (Miller-Abbott or Kaslow) tube may be useful if there is confusion between ileus and obstruction. A gastrostomy tube after an extensive operation, such as lysis of adhesions and fistula resection, is very

useful. A catheter jejunostomy (usually a #12 latex catheter) gives ample latitude with respect tot the type and consistency of the enteral diet. Stimulating agents are not of much value. Agents that have been used to hasten the return of small-bowel activity include erythromycin (a motilin agonist), bethanechol, vasopressin, metoclopramide, and cisapride. Hydration and nutritional support remain the mainstay of therapy.

Gastrointestinal Bleeding

Gastrointestinal bleeding has high mortality and morbidity rates, particularly because gastrointestinal bleeding in the elderly has become increasingly common. Accurate diagnosis is critical, particularly in the elderly, who have limited reserves. A careful history should be taken, a directed physical examination performed, and pertinent studies obtained in these patients before a significant amount (4 to 6 units) of blood has been lost.

Bleeding can arise anywhere along the entire gastrointestinal tract, from oropharynx to anus; it may be occult, presenting late with weakness, anemia, and orthostasis, or it may be massive, presenting with sudden, rapid loss of blood. Bleeding represents the initial symptom of gastrointestinal disease in more than one-third of patients, and in 70 percent there is no history of a previous bleeding episode. Although the majority of gastrointestinal bleeding (80 percent) will stop spontaneously without intervention, the other 20 percent of patients present a diagnostic and therapeutic challenge. Flexible endoscopy and arteriography play an increasing role in the diagnostic and therapeutic management of these patients.

Definitions

Hematemesis is the vomiting of blood, which may be either digested in the stomach or fresh and unaltered. It may be painful or painless, and it indicates upper gastrointestinal bleeding. Usually it is the result of bleeding between the oropharynx and the ligament of Treitz. Hematobilia occasionally results in hematemesis that may be associated with melena. Vomiting gross blood and clots indicates rapid bleeding, while “coffee-ground” emesis usually is associated with slower rates of bleeding, with the blood retained in the stomach altered to form acid hematin.

Melena is the peranal passage of stools with altered blood that are black and tarry and have a distinctive odor. Hematochezia is the passage of liquid blood or blood clots of varied brightness or color, maroon to bright red if the bleeding is vigorous enough. While acute bleeding can be dramatic, chronic bleeding may be equally troublesome, and in the absence of diagnosis might require repeated transfusions. Melena almost always occurs from the upper gastrointestinal tract, but it can occur from the distal small intestine or from the right side of the colon. Melena without hematemesis usually indicates a lesion distal to the ligament of Treitz.

As little as 50 to 60 mL of blood in the intestinal tract produces a melena stool. Melena can persist for 5 to 7 days after a significant, 2-unit bleed. Occult blood (positive guaiac test) may persist in the stool for up to 3 weeks after a good-sized bleed with associated hematemesis or melena; it does not imply continued bleeding. The amount required to convert the stool from guaiac-negative to guaiac-positive is approximately 10 mL of bleeding per day. While a generous portion of red meat occasionally can result in a guaiac-positive stool, this is infrequent. Iron does not result in a guaiac-positive stool, but it can result in a greenish-black stool that can be confused with melena. The distinctive odor of melena is the result of production of sulfide from heme by the action of hydrogen sulfide on the iron of the heme molecule.

Consequences of Gastrointestinal Bleeding

Most patients can withstand a sudden loss of 10 to 15 percent of their intravascular volume (class I hemorrhage) without shock. Hypotension occurs with greater intravascular losses, depending on the rate of bleeding and the patient response. Blood counts and hematocrit determinations are unreliable estimates of the amount of blood that has been lost, since equilibrium takes approximately 24 to 48 h. Initially a vasovagal response results in bradycardia; continued bleeding results in hypotension, cardiovascular instability, oliguria, decreased myocardial perfusion and decreased cerebral perfusion, renal failure, or multiple organ failure syndrome. With upper gastrointestinal blood loss, blood urea nitrogen (BUN) levels may be elevated to 30 to 50 mg/dL. The BUN:creatinine ratio may have a prognostic and diagnostic value; if the ratio is greater than 36:1, the bleeding arises from an upper gastrointestinal source. There is little azotemia seen with a lower gastrointestinal bleed, the BUN:creatinine ratio is less than 20:1, provided adequate renal function is maintained. Azotemia usually clears in 3 days after cessation of bleeding, without renal damage. Hypotension and renal ischemia may result in acute tubular necrosis.

Upper Gastrointestinal Bleeding

Etiology

The demographics of upper gastrointestinal bleeding depend largely on socioeconomic factors. Peptic ulcer is more common in suburban areas; gastritis and bleeding esophageal varices are more common in urban hospitals. The incidence of gastritis reflects the number of patients taking nonsteroidal anti-inflammatory medications. Duodenal ulcers bleed four times as often as gastric ulcers. Ten to 15 percent of patients with peptic ulcer bleed massively.

Upper gastrointestinal bleeding can be caused by nosebleed, esophageal and gastric varices, Mallory-Weiss tears, hiatus hernia and reflux esophagitis, gastritis (associated with portal hypertension, nonsteroid anti- inflammatory medications, or other gastric irritants), gastric ulcers, gastric neoplasms, duodenal ulcers, and hematobilia. There may be significant bleeding from duodenal polyps, but this is uncommon.

The proportion of patients presenting with upper gastrointestinal bleeding in the over-sixty age group has increased from 10 percent in the 1920s to almost 60 percent today. The mortality rate from upper gastrointestinal bleeding in patients over 60 years of age is 20 to 25 percent; in patients under 60 years old, the mortality rate is 4 percent. Although elective gastric operations for duodenal ulcer has decreased because of the widespread use of histamine H₂-receptor antagonists and proton-pump inhibitors, the incidence of surgery for bleeding duodenal ulcers has not decreased. Sudden cessation of histamine H₂-receptor antagonists or proton-pump inhibitors may result in a sudden rebound in acid secretion, resulting in gastrointestinal bleeding.

Nosebleeds rarely are the source of major bleeding but must be ruled out by a careful examination of the posterior pharynx to make certain that blood is not running down the pharynx into the esophagus and stomach, only to be vomited up later.

Esophagitis

Hiatus hernia and reflux esophagitis are not commonly the causes of massive bleeding. Reflux esophagitis is more likely to result in chronic occult bleeding. Significant acute bleeding is more often associated with paraesophageal hernias. The bleeding usually is associated with grade II or III esophagitis with friable mucosa.

Varices

Bleeding esophageal varices or gastric varices in the presence of liver disease account for about 10 percent of upper gastrointestinal bleeding and are a life-threatening situation with a high mortality rate. Alcoholism is the most common cause of portal hypertension, but hepatitis B and hepatitis C are increasingly seen as causes of posthepatic cirrhosis. Hepatocellular carcinoma may complicate hepatitis B and result in sudden onset of portal hypertension with portal vein thrombosis and bleeding. In the pediatric population, 95 percent of all upper gastrointestinal bleeding is because of variceal hemorrhage, usually as a consequence of extrahepatic portal venous obstruction. In patients with cirrhosis and portal hypertension, variceal hemorrhage accounts for 50 to 75 percent of all episodes of upper gastrointestinal bleeding, with gastritis and gastric or duodenal ulcer constituting the remainder. Variceal hemorrhage usually is precipitated by ulceration of the varix secondary to reflux esophagitis or increased pressure within the varix. Recurrent bleeding and mortality follow the inability of the failing liver to synthesize reparative proteins and proteins necessary for coagulation. Initial therapy should include prompt sclerotherapy or variceal ligation accompanied by vasopressin infusion. Variceal ligation is as effective in controlling bleeding as sclerotherapy and is associated with fewer treatment-related complications and better survival rates. If ligation or sclerotherapy is unsuccessful, emergency shunting procedures or hepatic transplantation must be considered.

Mucosal Tear

Esophagogastric mucosal, or Mallory-Weiss tears account for 5 to 10 percent of all upper gastrointestinal bleeding. The Mallory-Weiss tear presents in a classic pattern. Initially vomiting is without blood, but after retching and vomiting with epigastric pain, bleeding and hematemesis begin. More than 90 percent of the patients stop bleeding spontaneously and require no specific therapy. If bleeding persists, endoscopic therapy, with injection of a vasoconstrictive agent into the surrounding area, or intravenous infusion of vasopressin, usually stops the bleeding. Occasionally, balloon tamponade with a Sengstaken-Blakemore (gastric balloon only), or Linton tube is necessary to control the bleeding.

Gastritis

Depending on the patient population, up to one-third of upper gastrointestinal bleeding can be caused by the acute mucosal lesions of diffuse gastritis. These erosions often are multiple and are found primarily in the fundus and body of the stomach. Chronic gastritis is most commonly associated with *H. pylori* infection and presents with pain and chronic slow blood loss. Acute gastritis frequently bleeds and usually is associated with ingested substances that are damaging to the gastric mucosa. These include nonsteroidal anti-inflammatory drugs, alcohol, corticosteroids, and oral potassium supplements.

Vasopressin infusion, iced saline or lactated Ringer's solution lavage, the use of sucralfate, histamine H₂-receptor antagonists, or proton-pump inhibitors may be useful in stopping the hemorrhage. Endoscopic electrocoagulation or laser therapy have been used with mixed success. If the hemorrhage does not stop with medical or endoscopic therapy, vagotomy and antrectomy is the procedure of choice, with ligation of individual bleeding points. Total gastrectomy occasionally is necessary. While vagotomy and pyloroplasty, or gastrojejunostomy, are advocated by some, our experience with these procedures has been disappointing. If gastritis complicates portal hypertension, portal decompression may be necessary to stop repeated bleeding.

Peptic Ulcer

Peptic ulceration is the most common cause of upper gastrointestinal bleeding, present in one-half to two-thirds of patients with upper gastrointestinal bleeding. Bleeding may be the initial presenting symptom in up to 10 percent of patients with peptic ulcer. Duodenal ulcer bleeding is four times more common than gastric ulcer bleeding. Benign gastric ulcers tend to bleed more often and to a greater extent than gastric neoplasms, which rarely bleed significantly but may be associated with chronic blood loss. Bleeding duodenal ulcers usually are posterior and generally involve erosion in a branch of the gastroduodenal artery. Significant bleeding occurs in 10 to 15 percent of all peptic ulcer patients, and surgical intervention is required in up to 20 percent of these patients. The elderly are more likely to bleed persistently, because atherosclerotic vessels contract less, and operation may be necessary in a higher percentage of patients over 60 years old. if operation is required, it should not be delayed until these patients have received so much blood that they have developed a coagulopathy.

Peptic ulcer at the stoma of a gastrojejunostomy should be considered in any patient who has had previous gastric surgery. Inadequate resection, retained antrum, or inadequate vagotomy may predispose to stomal ulceration.

Stress Ulcer

This term refers to the acute gastroduodenal lesions that arise after or during episodes of shock, sepsis, surgery, trauma, burns (Curling's ulcers), and intracranial pathology or

surgery (Cushing's ulcers). The current hypothesis about stress ulceration is that it is the product of bile reflux damage to the gastric protective barrier combined with decreased gastric blood flow secondary to splanchnic vasoconstriction, but sepsis, coagulopathy, and activation of cytokines and other mediators also may play a role. Stress ulceration is associated with eight documented specific risk factors: multiple system trauma, hypotension, respiratory failure, renal failure, sepsis, jaundice, recent surgery, and burns. The greater the number of risk factors, the greater the incidence of ulceration and bleeding. Curling's ulcers may be multiple and increase in frequency as the percentage of body surface area burned increases. They occur in 12 percent of patients with burns. Cushing's ulcers probably result from the same pathophysiology as stress ulcerations, although significant gastric hypersecretion may occur after certain neurosurgical procedures. In the past, stress ulceration was common in the intensive care unit, but as resuscitation and postoperative care have improved, stress ulceration has become less common. Gastric acid neutralization, with sucralfate and antacids, also has contributed to reducing the incidence.

Other Causes

Miscellaneous disorders may lead to upper gastrointestinal bleeding in 8 to 18 percent of cases. Gastric neoplasms, such as adenocarcinoma, adenoma, angioma, leiomyoma, leiomyosarcoma, lymphoma, and leukemia all may bleed. Dieulafoy's vascular malformations are rare submucosal dilated arterial lesions that may occur anywhere in the gastrointestinal tract but are most commonly found in the stomach. The cause of these lesions is unknown, and they may present with massive gastrointestinal hemorrhage. Bleeding from Dieulafoy's lesions usually can be managed by endoscopic injection treatment, with good long-term outcomes. Aortoenteric fistulas usually present as a sentinel or herald bleed that is followed by a massive bleed. Aortoenteric fistulas are seen in patients with prior aortic reconstructive surgery; low-grade sepsis may involve the graft at the upper suture line, which usually is contiguous with the posterior aspect of the third portion of the duodenum. Primary aortoenteric fistulas, although rare, also have been reported. Hematobilia rarely is seen and usually follows hepatic injury or manipulation (e.g., percutaneous transhepatic procedures and catheterization) in which there has been arterial damage adjacent to the biliary system. Colicky pain and jaundice sometimes are present, aiding in the diagnosis. Bleeding can be massive, requiring urgent intervention.

Clinical Manifestations

A careful history should be taken, with inquiries about peptic ulcer disease, medications, alcoholism, previous cirrhosis, heartburn, and reflux esophagitis. If the initial vomitus was without blood and the subsequent vomiting was with pain and blood, this suggests a Mallory-Weiss tear. Peptic ulcer pain, which disappears with bleeding, strongly suggests duodenal ulcer bleeding. Twenty percent of patients with bleeding ulcer have no previous history of ulcer. While patients with cirrhosis have an increased ulcer diathesis, esophagogastric varices accounts for more than one-half of the bleeding episodes in cirrhotics.

Patients should be questioned about medications, particularly ingestion of drugs implicated as causes of gastrointestinal bleeding: salicylates, alcohol, steroids, and anticoagulants. Heartburn and epigastric substernal pain suggest reflux esophagitis.

Physical examination should focus on the various diagnostic possibilities and should include careful palpation of the neck for the occasional rare palpable parathyroid adenoma and examination for spider angiomas, rhinophyma, palmar erythema, and Dupuytren's contracture of the hands, a prominent abdominal venous pattern, and other stigmata of cirrhosis, including testicular atrophy, caput medusa, ascites, malnutrition, and jaundice. The mucous membrane should be investigated for melanin spots of the Peutz-Jeghers syndrome. Hereditary telangiectasis lesions are most common on the lips, tongue, and ears. Virchow's (left supraclavicular) node suggests a malignant intraabdominal process. A rectal examination and stool guaiac test should be performed. Auscultation of the abdomen will reveal hyperactive bowel sounds secondary to the cathartic effect of blood, especially in upper gastrointestinal bleeding.

Management

Large-bore intravenous lines should be placed and a type and crossmatch for 4 to 6 units initiated. Crystalloid resuscitation proceeds until blood is available. The initial priority is to restore blood volume. A Foley catheter is essential to monitor the resuscitation; if resuscitation is insufficient, a central venous catheter or, preferably, a pulmonary artery catheter may be required. A nasogastric tube should be placed in all patients with gastrointestinal bleeding. Even if upper gastrointestinal bleeding appears unlikely, it cannot be excluded until guaiac-negative bile is seen in the nasogastric tube.

Once the patient is stabilized, appropriate diagnostic studies can be undertaken. The initial laboratory evaluation should include a complete blood and platelet count, determinations of prothrombin and partial thromboplastin time, measurement of electrolyte levels, and liver chemistry studies. The progress of resuscitation can be checked by determining whether the patient has orthostatic hypotension. If the patient has a blood pressure that falls 20 mmHg, a decreased pulse pressure, or the pulse rises more than 25 beats/min when sitting, it can be assumed that there has been a loss of at least 20 percent of circulating blood volume (class II hemorrhage), and transfusion is required.

If upper gastrointestinal bleeding is suspected, the stomach is lavaged with iced saline solution with a large-bore tube, and emergency esophagogastroduodenoscopy is performed. This should be the first diagnostic test performed as soon as the patient is hemodynamically stable. The accuracy of upper endoscopy for the determination of the cause of bleeding ranges from 74 percent to 96 percent and is approximately 90 percent accurate when performed within 48 h of initial presentation. Bleeding from a duodenal or gastric ulcer, acute gastritis, Mallory-Weiss tear, and varices may be controlled endoscopically by various methods of coagulation, injection, or ligation. Barium studies and upper gastrointestinal series are contraindicated, because the presence of barium in the gastrointestinal tract precludes radionuclide and angiographic studies. Radionuclide scanning with technetium-labeled red blood cells may identify bleeding at rates as low as 0.1 to 0.2 mL/min (Fig. 22-6), but it is not helpful in localizing the site of bleeding, particularly in the upper tract. It is most useful as a precursor to arteriography. Arteriography may localize acute bleeding points if blood loss exceeds 1 mL/min (Fig. 22-7). Embolization of specific lesions may be carried out. Embolization of hepatic arteries may result in total thrombosis of hepatic arteries and should be considered carefully. In the case of bleeding esophageal varices, no lesion will be seen on arteriogram, and there is no indication for arteriography; intraarterial vasopressin has no advantage over intravenous vasopressin.

If the patient with bleeding esophageal varices fails to stabilize after sclerotherapy, ligation, or intravenous vasopressin, sclerotherapy or ligation can be attempted a second time. The Sengstaken-Blakemore tube should be tested and then placed through the nostril down to the farthest point; the gastric balloon should be blown up to approximately 50 to 100 mL and its position verified by radiograph. The balloon is inflated to 300 mL and snugged against the gastroesophageal junction. One or two pounds of traction is added, and this may be sufficient to stop the hemorrhage; if not, the esophageal balloon is inflated to 40 mmHg. The balloon can be left in place for 24 h. If bleeding resumes when the balloon is deflated, this is an indication for operation. Additional temporization will only increase the risk to the patient.

Some have advocated *transjugular intrahepatic portosystemic shunting (TIPS)* as a procedure for the acutely bleeding patient with portal hypertension, especially those who are candidates for transplantation. Access is gained through a major intrahepatic portovenous branch by puncture through a hepatic vein. A parenchymal track between hepatic and portal veins is then created with a balloon catheter, and a 10-mm expandable metal stent is inserted, creating the shunt with a pressure gradient of less than 12 mmHg. The initial experience reported for acutely bleeding patients is that a shunt is obtained successfully in more than 90 percent. An in-hospital mortality rate of 35 to 56 percent has been reported, and incomplete control of bleeding occurred in 26 percent (Fig. 22-8). In patients who survive, the incidence of encephalopathy is approximately the same as it is in patients receiving portacaval shunts. Another disadvantage of the procedure is that shunt stenosis or occlusion develops in as many as 50 percent of the patients within 1 year of TIPS insertion, invariably caused by intimal hyperplasia. While stenosis can be remedied by balloon dilation, it requires repeated angiograms or Doppler ultrasonographic examinations. By 5 months, 50 percent of stents had a significant gradient, and 80 percent required angioplasty within 1 year. After an initial wave of enthusiasm, TIPS now is being reserved for patients who have failed other forms of nonoperative therapy such as sclerotherapy, vasopressin, or variceal ligation and in whom transplantation is being considered. It is used as a bridge for transplantation.

TIPS is the method of choice for the treatment of refractory complications of portal hypertension in the liver transplant candidate. In patients undergoing orthotopic liver transplantation, TIPS is associated with reduced operative transfusion requirements, operative time, and length of intensive care unit and hospital stays compared to surgical portosystemic shunts. Because of the low morbidity of an elective surgical shunt and the high cost of maintenance of TIPS, elective surgical shunt is still recommended for those stable patients considered unsuitable for orthotopic liver transplantation.

Orloff has proposed another approach, in which patients with demonstrated bleeding esophageal varices are taken to the operating room for emergency portacaval shunt within 8 h of presentation, after a rapidly carried out diagnostic series. An impressive mortality rate of 17 percent has been reported. While some authorities disagree with this approach, temporization of patients with bleeding esophageal varices leads to aspiration, pneumonitis, coagulopathy, and deterioration of hepatic function.

If upper gastrointestinal bleeding requires 4 units of blood within the first 24 h and continues, it is an indication for operation. This is particularly true in the elderly, in whom gastrointestinal bleeding is less well tolerated, and in whom underlying arteriosclerosis with vascular compromise more often leads to cerebrovascular accidents or myocardial infarctions. The surgical approach should be tailored to the underlying pathology, with the primary goal to control the bleeding and then to institute more definitive therapy, depending on the patient's general condition and ability to tolerate surgical stress.

Lower Gastrointestinal Bleeding

Small Intestine

Bleeding from the small intestine is uncommon, occurring in 10 to 15 percent of all lower gastrointestinal bleeding, and rarely is massive or life-threatening. Small-intestine bleeding is one of the more difficult diagnostic situations and usually is a diagnosis of exclusion after upper gastrointestinal and colonic sources of bleeding have been ruled out. Causes of small-intestine bleeding include Meckel's diverticula, Crohn's disease, intussusception, neoplasms, vascular malformations (such as hereditary telangiectasia, microaneurysm, and hemangioma), intestinal varices, blood dyscrasias, non-Meckel's diverticular, mesenteric thrombosis, drug reactions, enteric infections, and polyps, some of which are associated with the Peutz-Jegher syndrome.

A thorough endoscopic examination of the esophagus, stomach, entire duodenum, and colon is essential before assuming a small-bowel source. In approximately 10 percent of cases of small-bowel bleeding, blood is seen exiting the ileocecal valve on colonoscopy. Bleeding associated with a Meckel's diverticulum usually is seen in the pediatric population and is related to ectopic gastric mucosa within the diverticulum. Technetium scanning can be used to demonstrate a Meckel's diverticulum. Crohn's disease presents with melena or hematochezia in 20 percent of patients; the bleeding rarely is of significant quantity. Intussusception generally presents with a classic history and symptom complex of intermittent colicky pain with signs of partial obstruction along with the passage of the characteristic currant-jelly stool. This most commonly affects infants; the cause usually is not determined, and diagnosis and treatment are implemented together as a barium enema. Operative reduction or resection may be necessary. The adult who present with intussusception generally has a neoplasm at the lead point and should undergo operative exploration and resection.

A radionuclide ^{99m}Tc-tagged erythrocyte scan during the acute bleeding may be helpful in identifying the area of origin of the bleeding. Although it is advocated as capable of detecting blood loss as slow as 0.1 to 0.2 mL/min, a loss of 0.5 mL/min is more realistic; the scan can be repeated for up to 30 h when bleeding is intermittent (see Fig. 22-6). Arteriography will identify a bleeding site when the bleeding rate is higher, e.g., 1 to 2 mL/min (see Fig. 22-7). Arteriography also may identify a tumor blush or an arteriovenous malformation or angiodysplasia. Tagged-erythrocyte scans and arteriographic investigations may demonstrate a small-intestinal source and may localize it to the upper or lower small intestine. If the specific arterial arcade affected can be identified on angiography, it is helpful to select the specific branch and leave the catheter in place for intraoperative

localization. At operation, it may be necessary to transilluminate the small bowel or to carry out intraoperative enteroscopy, which is an effective tool in selected patients with occult gastrointestinal bleeding and correctly identifies a treatable source and prevents rebleeding in 41 percent of patients.

Colon Bleeding Etiology

Colonic bleeding can be acute and massive, or chronic. Chronic bleeding most often results from polyps or neoplastic disease (Table 22-14). Right-sided neoplastic lesions tend to be more silent and endocolonic rather than scirrhous and obstructive, and they come to attention because of anemia and a guaiac-positive stool. Melena occasionally is present, but usually the presentation is that of unexplained anemia, and perhaps orthostatic hypotension, fatigue, and weight loss. Diverticulosis rarely is the cause of chronic colon bleeding. Massive lower gastrointestinal bleeding is caused by diverticulosis or angiodysplasia. Angiodysplastic lesions are small and occur most frequently in the cecum and the right colon. These vascular lesions have the following characteristics: (1) they are not congenital or neoplastic, but degenerative; (2) they are not associated with other vascular lesions of viscera or skin; (3) they increase with age; (4) they usually are small (less than 5 mm in diameter) and pathologically may require injection and clearing techniques for demonstration; and (5) they can be diagnosed by colonoscopy, but most accurately by angiography.

Colonic bleeding from angiodysplasia is best diagnosed by arteriography or colonoscopy. Almost 80 percent of these patients will stop bleeding spontaneously, but the risk of subsequent rebleeding is significant; approximately 50 percent of patients rebleed within three years after the initial bleed. An operation offers the best potential cure, but rebleeding from other angiodysplastic sources has been reported. The risk of rebleeding after laser ablation or electrocoagulation of angiodysplasia, is great.

Ulcerative colitis rarely bleeds massively and is characterized by chronic bloody diarrhea in most cases, but occasionally massive bleeding occurs. Polyps, other neoplastic disorders, and other ulcerated lesions also may bleed. Cytomegalovirus infection is another possible cause of lower gastrointestinal bleeding and is becoming more common, especially in immunocompromised patients. Anorectal disorders, including hemorrhoids, particularly in patients with liver disease, occasionally result in massive hemorrhage. These usually are within easy reach of the anoscope or rigid sigmoidoscope.

Management

Massive lower gastrointestinal bleeding is a surgical emergency and the patient should be resuscitated rapidly. A Foley catheter should be placed, and upper gastrointestinal bleeding ruled out with the placement of a nasogastric tube. Crystalloids should be given rapidly through two large-bore intravenous lines, and type-specific blood should be given as soon as available. Anoscopy and proctosigmoidoscopy are performed as part of the initial evaluation. The technetium-labeled red blood cell scan is the least invasive method available for detecting bleeding. This technique can identify slow bleeding and may pick up intermittent bleeding, since the patient can be rescanned over a 30-h period as frequently as necessary once the tagged cells have been injected. Radionuclide scanning is 91 percent sensitive and 100 percent specific for bleeding. Although its accuracy in pinpointing a specific site of bleeding is poor, tagged red cell scintigraphy appears to be a useful screening examination for patients with lower gastrointestinal bleeding who are hemodynamically stable, those with recurrent or prolonged bleeding, those with inconclusive endoscopy or barium studies, and those who are high-risk surgical candidates. This may avoid the potential morbidity associated with an unnecessary angiographic study and can help the radiologist in focusing the angiogram on the area of greatest concern.

Angiography is the definitive method of locating the point of bleeding. A large proportion of scans and arteriograms define bleeding from diverticula or angiodysplasia in the hepatic flexure or right colon. The use of CT and MRI for identifying gastrointestinal hemorrhage has been reported. CT performed without oral contrast agents, immediately after a negative angiographic result, may identify a region of the bowel that has focally dense intraluminal contrast. If a long circulating macromolecular contrast agent is used to decrease the T₁ signal of extravasated blood, gastrointestinal hemorrhage can be easily detected with MRI. While colonoscopy often is performed during massive gastrointestinal bleeding, it rarely provides useful information, because the patient is poorly prepared, and the bleeding, particularly if it is right-sided, cannot be localized. Barium investigations are not recommended for patients with lower gastrointestinal bleeding. Once barium is in the colon, angiographic localization is impossible.

If operation is required, when angiographic localization reveals well defined left- or right-sided colon lesions, local resection is indicated. If bleeding cannot be localized, colectomy and ileoproctostomy is indicated.

Chronic lower gastrointestinal bleeding should be evaluated with a careful history with questioning on weight loss, abdominal pain, family history of carcinoma, and the characteristic crampy abdominal pain and altered bowel habits of diverticular disease. Physical examination should include careful abdominal examination, palpating for a mass or the rigid or spastic sigmoid of diverticulitis, and pelvic and rectal examinations. Anoscopy and proctosigmoidoscopy are carried out in conjunction with barium enema, but colonoscopy is preferable. Barium enema cannot identify angiodysplastic lesions, but colonoscopy is successful in identifying them in 70 percent of patients. Double-contrast barium enema is helpful if colonoscopy is difficult or if a poor preparation limits the information to be gathered by colonoscopy. In rare instances, when colonoscopy or double-contrast barium enema fail to reveal any lesions, angiography may be indicated.

Rectal and Anal Bleeding

Fresh, unchanged blood on the exterior of the stool usually is from hemorrhoids, fissure, or proctitis. Pain associated with bleeding is usually pathognomic of fissure. Bleeding that drips into the bowl water is most likely the result of hemorrhoids or fissure. Any rectal or anal bleeding should be fully investigated; it should not be assumed that the bleeding is caused by hemorrhoids, because carcinomas can occur in young patients and a proximal tumor can cause engorgement of hemorrhoidal veins.

Jaundice

The term *jaundice* is derived from the French word meaning yellow, and it refers to excess bile pigments in the tissues and serum. Jaundice is clinically recognizable when the level of bilirubin is approximately 2.0 to 3.0 mg/dL. For the surgeon, the most important question is whether the patient has “medical jaundice” or “surgical jaundice,” the latter caused by obstruction that must be relieved by a surgical procedure. With careful history taking and physical examination, and simple laboratory tests (including liver chemistry studies), an accurate diagnosis can be achieved in approximately 85 percent of patients.

Normal Bilirubin Metabolism

Bilirubin, a tetrapyrrole, is produced by the oxidation of heme and reduction of the resultant biliverdin (Fig. 22-9). The majority (80 percent) of bilirubin formed is derived from the catabolism of heme released from the destruction of senescent red cells, principally in the spleen. The other 20 percent of daily bilirubin production is derived from the rapid turnover of hepatic heme-proteins and from ineffective erythrocytes in the bone marrow. Bilirubin is water insoluble, and its elimination, carried out by the hepatocyte, is largely dependent on the liver, which conjugates it with sulfate and glucuronide and excretes it into the bile. Unconjugated bilirubin is rapidly removed from the circulation by the liver. Wherever it originates, bilirubin is bound to albumin in the systemic circulation and dissociates from albumin when transported into the hepatocyte.

Inside the hepatocyte, the bilirubin binds to the cytosolic proteins ligandin and Z protein. Bilirubin conjugation, primarily with glucuronic acid, occurs in microsomes and is mediated by a specific bilirubin uridine diphosphate glucuronic acid (UDPGA) transferase. The transfer of one glucuronic acid to either of the two linked carboxyethyl side chains yields two isomeric bilirubin monoglucuronides (BMG). Transfer of a second glucuronic acid to the remaining carboxyethyl side chain forms bilirubin diglucuronide. In human bile more than 85 percent of the secreted bilirubin is in the form of bilirubin diglucuronide. Deficiency of the enzyme or the nucleotides may result in impaired bilirubin conjugation and unconjugated hyperbilirubinemia.

Bilirubin in normal bile is at least 98 percent conjugated, but unconjugated bilirubin is present in small concentrations. The major source of the unconjugated bilirubin is hydrolysis and enzymatic activity on excreted bilirubin conjugates that may occur at alkaline pH. The sick or failing liver may conjugate only one glucuronide on bilirubin, and this may be partially excreted into the bile. On a routine laboratory test, this will be read as direct bilirubin and can make liver chemistry studies difficult to interpret.

Once conjugated and dissolved, the bilirubin is secreted across the canaliculus by a carrier-mediated transport system. The bile salts form a “micellar sink,” which augments passive diffusion of the pigment from the cell into the bile. The conjugated bilirubin is poorly absorbed and usually fecally excreted. Conjugated bilirubin may be deconjugated by bacterial or intestinal glucuronidases. Bacteria further metabolizes the conjugated bilirubin to urobilinogen and urobilin (Figs. 22-9, 22-10). The majority of the unconjugated bilirubin and urobilinogen is excreted in the feces. In adults, a minor amount undergoes an enterohepatic circulation, while in the newborn, absorption of unconjugated bilirubin may cause “physiologic jaundice.” Any conjugated bilirubin and urobilinogen that is reabsorbed may be excreted through the kidney. Normally, almost all conjugated bilirubin is in the bile and is not present in the plasma, and the 0.1 or 0.2 mg/dL of conjugated bilirubin reported in most normal liver chemistry studies is an artifact.

Etiology

A number of causes are responsible for jaundice (Fig. 22-11). These include:

(1) Congenital causes

(A) Enzymatic (glucuronyl transferase) deficiencies, e.g., Crigler-Najjar syndrome and Gilbert's syndrome

(B) Familial conjugative disorders, e.g., Dubin-Johnson syndrome and Rotor's syndrome

(C) Overproduction of bilirubin because of hemolytic diseases ineffective erythropoiesis

(2) Physiologic causes

(A) Neonatal jaundice secondary to “immature” or inadequate glucuronyl transferase

(B) Increased pigment production secondary to tissue infarction or large collections of blood in tissues

(3) Inflammatory disease

(A) Hepatitis —toxic, alcoholic, or infectious

(B) Infection elsewhere in a patient with cirrhosis, e.g., spontaneous bacterial peritonitis or subacute bacterial endocarditis

- (C) Appendicitis, complicated by coliform organisms, can give jaundice without pylephlebitis
- (4) Metabolic/nutritional deficiencies
- (A) Drug-impaired uptake of bilirubin and drug-induced hemolysis and cholestasis
- (B) Alcoholic malnutrition –induced hepatitis and cirrhosis
- (C) Gallstones, with manifestations, including common hepatic duct stones and common bile duct stones
- (5) Neoplastic disease
- (A) Primary tumors of the liver
- (B) Metastases from other primary tumors
- (C) Extrahepatic obstruction

((1))

Intrahepatic obstruction of cholangiocarcinoma or adenocarcinoma

((2))

Klatskin (common hepatic duct bifurcation obstruction) types of tumors

- (D) Carcinoma of the gallbladder, with obstruction of the common bile duct
- (E) Ampullary carcinoma, or carcinoma of the duodenum
- (F) Carcinoma of the pancreas
- (G) Obstruction of bile ducts by portal nodes secondary to lymphoma or metastases

Congenital or familial types of hyperbilirubinemia are characterized by increases in unconjugated or conjugated bilirubin. In the unconjugated type of hyperbilirubinemia, an inherited partial deficiency of glucuronyl transferase (or Gilbert's syndrome) is common and affects 3 to 6 percent of the population, more commonly males. Serum bilirubin is low, usually below 3 mg/dL and rarely exceeding 5 mg/dL. The jaundice may fluctuate and is exacerbated after prolonged fasting, surgery, fever or infection, excessive exertion, or alcohol ingestion. The diagnosis of this disorder usually is made by exclusion in a patient with low-grade unconjugated hyperbilirubinemia with no systemic symptoms, no obvious hemolysis, normal liver function tests, and a normal liver biopsy. The history will reveal mild intermittent jaundice. Crigler-Najjar syndrome can be manifested in two types: an absence of glucuronyl transferase occurs in the autosomal recessive type I, and a partial deficiency occurs in the autosomal dominant type II. Type I is a rare disorder that presents in infancy; type II is less severe and may not appear until adolescence. Acquired deficiencies of glucuronyl transferase occur when there is an inhibition of the enzyme by certain agents. Drugs such as chloramphenicol, novobiocin, or vitamin K are known to inhibit the enzyme. Because the enzyme activity is decreased in the neonatal state, inhibition is more evident at that time. Pregnan-3b, 20 a-diol in breast milk is a glucuronyl transferase inhibitor and may lead to the so-called breast-milk jaundice. Hypothyroidism delays the normal maturation of glucuronyl transferase and may prolong neonatal jaundice if unrecognized.

Hepatic uptake of bilirubin can be impaired by flavaspidic acid and novobiocin as well as some cholecystographic dyes. Disorders associated with hemolysis, such as the hemolytic anemias and increased pigment production, can lead to an increased rate of bilirubin production that may exceed the amount that can be cleared by a normal liver. Ineffective erythropoiesis in patients with thalassemia, pernicious anemia, and congenital erythropoietic porphyria can lead to an increased destruction of red blood cells or their precursors in the marrow. The resultant hyperbilirubinemia in all of these conditions is primarily unconjugated with small increases in conjugated bilirubin and serum bilirubin in the range of 3 to 5 mg/dL. There are some situations in which there is a combination of hemolysis and impaired liver function (cirrhosis, sepsis, neoplasm), and serum bilirubin levels may be higher. Unconjugated hyperbilirubinemia is rarely a surgical condition. The serum bilirubin tends to be low, below 4 mg/dL, because the production of bilirubin exceeds the liver's capacity to conjugate with glucoronide. While hemolytic disorders frequently result in gallstone formation, which may produce mechanical obstruction, uncomplicated hyperbilirubinemia of the unconjugated variety usually is a benign condition.

When jaundice is caused by primary liver disease, there usually are elevated plasma levels of conjugated and unconjugated bilirubin, and the urine contains bilirubin. In many familial hepatic abnormalities, in some forms of liver injury, and in extrahepatic biliary obstructive jaundice, the increase is in the conjugated bilirubin. The levels or proportions of unconjugated and conjugated bilirubin in the serum do not help in the differentiation between intrahepatic and extrahepatic causes of jaundice, but only differentiate parenchymal and obstructive disease from disorders associated with predominately unconjugated hyperbilirubinemia (Table 22-15).

Predominantly conjugated jaundice may be caused by the familial defects in hepatic excretory function of Dubin-Johnson syndrome and Rotor's syndrome. In the autosomally recessive Dubin-Johnson syndrome, also called chronic idiopathic jaundice, there is a cellular defect in biliary excretion of amphophilic anionic conjugates (i.e., bilirubin, cholephilic iodinated dyes, sodium sulfobromophthalein, epinephrine metabolites, porphyrins) from hepatocytes into the bile. The mechanism of this disorder is unknown, but it may be a result of the absence of the canalicular isoform of the multidrug-resistant protein (MRP) gene encoded conjugate export pump. The serum bilirubin level in these patients is typically between 3 and 15 mg/dL; it is predominantly conjugated, with the serum containing more deconjugated than monoconjugated bilirubin. Homozygous patients typically have significant levels of unconjugated bilirubin. There is an accumulation of a black pigment, believed to represent the accumulation of epinephrine metabolites in the liver, which leads to the characteristic "black liver."

Dubin-Johnson patients may be asymptomatic or may present with vague gastrointestinal or constitutional complaints. There may be slight hepatomegaly and tenderness. The biliary tract does not visualize with oral or intravenous cholangiography. Sodium sulfobromophthalein elimination rises late (90 min) and there is no secondary rise in the serum of dyes that are not conjugated in the liver (e.g., indocyanine green). Urinary excretion of coproporphyrin I is increased, while the normal urinary metabolite coproporphyrin III is decreased. Oral contraceptive use may accentuate the hyperbilirubinemia. The overall prognosis in Dubin-Johnson syndrome usually is excellent.

Rotor's syndrome is a rare autosomal recessive trait that results in an impairment of hepatic storage capacity of conjugated bilirubin. In contrast to Dubin-Johnson syndrome, the liver is not pigmented, serum conjugated bilirubin is predominantly the monoconjugate, the biliary tree visualizes on oral and intravenous cholangiography, there is no secondary rise in sodium sulfobromophthalein excretion, and total urinary coproporphyrin excretion is increased.

Other are causes of hyperbilirubinemia include benign familial recurrent cholestasis, recurrent jaundice of pregnancy, and drug-induced cholestasis. Benign familial recurrent cholestasis is believed to be congenital and is first manifested at a young age. This disorder is characterized by recurrent attacks of pruritus and jaundice; liver biopsy reveals a cholestatic morphology, but cirrhosis does not develop, and the course usually is benign. Recurrent jaundice or intrahepatic cholestasis of pregnancy usually occurs in the third trimester but may develop any time after the seventh week of gestation. The serum bilirubin level is elevated (usually > 6 mg/dL), as are alkaline phosphatase and cholesterol levels. Patients may complain of pruritus that responds to cholestyramine treatment. The cause is thought to be related to an increased sensitivity to the hepatic effects of estrogens and progesterone. The symptoms usually subside within 1 to 2 weeks after delivery but may recur with subsequent pregnancies.

Drug-induced cholestasis is commonly associated with oral contraceptive use and resolves with the discontinuation of these medications. The administration of certain testosterone analogues also can lead to jaundice but is associated with chronic liver disease and biliary cirrhosis. Many other drugs may cause cholestasis as well as liver injury along with other systemic features, including fever, rash arthralgia, and eosinophilia. Some of the common drugs associated with hepatic toxicity or jaundice are methimazole, erythromycin estolate, chlorpropamide, chlorpromazine, tetracycline, halothane, phenytoin, methyldopa, isoniazid, chlorothiazide, and acetaminophen.

The most common disorders associated with jaundice are hepatitis and cirrhosis. Inflammation result sin hepatocellular damage or dysfunction, and it is important to recognize these conditions, because operative intervention, especially requiring general anesthesia, can be fatal. Impairment of hepatic function occurs form viral hepatitis or hepatitis of other viral causes, including cytomegalovirus in the immunosuppressed patient. Hepatitis A, B, or C may adversely affect hepatic function. Infection elsewhere in the body also can give rise to a hepatitis, which might impair hepatic function, e.g., multiple organ failure syndrome. When patients have cirrhosis, infection may result in impaired hepatocellular function. When hepatitis is clinically significant, serum transaminase and lactate dehydrogenase usually are elevated, with alkaline phosphatase slightly elevated, and bilirubin variously elevated. In patients with jaundice and brown stools, the diagnosis almost always is related to hepatic disease rather than surgically correctable causes.

Metabolic/nutritional causes of jaundice include alcoholic hepatitis and cirrhosis and the hemolytic effects of associated liver disease. If jaundice complicates cirrhosis of the liver, it is a poor prognostic sign. In patients with established cirrhosis, remote infection can result in jaundice and hepatic dysfunction. Starvation, increased alcoholism, and any noxious stimulus, including general anesthesia, also will cause jaundice in a patient with cirrhosis.

Alcoholic hepatitis is the result of excessive ingestion of alcohol over a short period of time, and probably a deficiency of normal nutrients. An enlarged tender liver usually is present, often with fever and leukocytosis. High levels of conjugated serum bilirubin are recorded, and alkaline phosphatase levels may be two to three items normal; the serum transaminase levels rarely exceed 300 units/L, and the ratio of AST to ALT is almost always greater than 2:1. Histologic findings on liver biopsy include fatty infiltration, polymorphonuclear inflammation, cellular necrosis, and Mallory's hyalin in the cytoplasm. Operation on patients with alcoholic hepatitis should be avoided if possible. The therapy of alcoholic hepatitis is abstinence from alcohol, and nutritional support, enteral or parenteral. Conventional nutritional support formulas can be used if the patient tolerates them

without encephalopathy. If these regimens are not tolerated, patients should receive branched-chain enriched amino acid nutritional support, enterally or parenterally.

Extrahepatic biliary obstruction can be caused by gallstones, neoplasm, trauma, or injury at the time of operation. Patients with gallstones usually have a history of right upper quadrant pain, radiating to the back or to the right shoulder, in association with fried, fatty, or greasy foods. Alkaline phosphatase elevation in patients with jaundice is likely to be more significant than the increase in bilirubin, which is not often above 10 mg/dL. It is rare to have elevation of bilirubin higher than 15 mg/dL with gallstones obstructing the common duct. It also is rare to have an elevation of bilirubin above 15 mg/dL with any mechanical obstruction, and values higher than this usually indicate hepatitis or cholestatic jaundice, e.g., hepatitis secondary to medication. Injury to the common duct after cholecystectomy may be manifested immediately or may be delayed.

A host of intrahepatic neoplasms can produce jaundice, including primary cholangiocarcinoma, intrahepatic duct obstruction, and metastases from any gastrointestinal primary carcinoma.

Most jaundice is the result of extrahepatic biliary obstruction. Adenocarcinoma of the common hepatic duct bifurcation (or Klatskin's tumor) is common. Less commonly, adenocarcinoma affects the proximal common bile duct. Carcinoma of the gallbladder, which is a silent lesion, may present as jaundice when the carcinoma obstructs the common bile duct by external compression. Carcinomas at the distal end of the biliary tree, including carcinoma of the pancreas, ampulla, and duodenum, are the most common causes of surgically correctable jaundice. The elevation of alkaline phosphatase tends to be higher, and the bilirubin level rarely exceeds 15 mg/dL with these lesions. Early anorexia is a specific manifestation of carcinoma of the pancreas, even with a small tumor burden. Portal lymph nodes may obstruct the common bile duct from the metastases, e.g., colon carcinoma, or lymphoma.

Clinical Manifestations

History

The history taking should focus questions on previous episodes of jaundice and whether the urine is dark and the stool light. Brown stools with jaundice usually are associated with cholestatic jaundice and nonsurgical jaundice. Inquiries should be made about similar episodes or episodic pain in the right upper quadrant provoked by fried, fatty, or greasy foods. Hepatitis exposure is highly significant. Family history and ethnic background have some bearing on some of the hemoglobinopathies, such as thalassemia or sickle cell anemia. Loss of appetite, weight loss, and the use of alcohol, drugs, or medications should be determined. Although carcinoma of the pancreas is classically painless, 20 to 30 percent of patients complain of backache or epigastric distress. A history of pruritus preceding jaundice occurs more commonly with extrahepatic and intrahepatic obstruction.

Physical Examination

The stigmata of chronic liver disease should be investigated. The liver edge should be palpated, and the total liver span estimated. The presence of a palpable spleen is highly significant. Adenopathies in the axilla or in the neck are significant. A stool guaiac test should be performed. A palpable gallbladder is more commonly associated with extrahepatic ductal obstruction when the obstruction is below the entrance to the cystic duct. This finding, known as Courvoisier's law, is more common in neoplasia than with biliary calculi.

Laboratory Studies

Laboratory studies should include liver function tests, a urinalysis with particular attention to the presence of bilirubin and urobilinogen, a complete blood count, and determinations of serum electrolytes, prothrombin time, and serum amylase. The alkaline phosphatase level may give some indication of the presence and degree of obstruction. An extremely high alkaline phosphatase value is suggestive of an intrahepatic process. The degree of hyperbilirubinemia and direct or indirect fractions may give some evidence about the cause of the jaundice. The presence of bilirubin or urobilinogen in the urine generally indicates good hepatic function and obstruction. Elevation of the transaminases and lactate dehydrogenase can indicate an acute inflammatory pattern. Prothrombin time elevation and hypoalbuminemia provide evidence of impaired synthetic function (as in cirrhosis). Hypokalemia, alkalosis, and a low blood urea nitrogen level may be associated with chronic liver disease. Amylase elevation, if present, provides additional evidence of an extrahepatic (obstructive) cause.

Radiologic Examination

The simplest and most noninvasive method of imaging the biliary tree is an ultrasonographic examination looking for dilated bile ducts or stones in the gallbladder. Hepatobiliary (HIDA) scans are accurate in diagnosing complete cystic duct or common duct obstruction, but offer little anatomic accuracy. HIDA scans are most useful in diagnosing acute cholecystitis and identifying postoperative biliary fistulas. If dilated ducts are found, a percutaneous transhepatic cholangiogram (PTC) or endoscopic retrograde cholangiopancreatography (ERCP) can be performed, depending on local expertise, suspected cause, and therapeutic options. Both studies are highly sensitive and specific in delineating the cause of the obstruction. PTC is associated with a certain incidence of bleeding and bile leakage, particularly in patients with impaired synthetic function. Bile leakage may be minimized by leaving the catheter in for drainage. ERCP may allow for direct biopsy or brushing of a suspected neoplasm and manipulation of the distal duct with removal of stones. The purpose of both examinations is to define the anatomy and, if obstruction is present, to place a stent or decompressive tube. Antibiotics should be given before manipulation to obviate cholangitis. ERCP is the procedure of choice in patients who are jaundiced after cholecystectomy, because endoscopic sphincterotomy permits stone removal.

CT is less sensitive in detecting noncalcified gallstones, but is more reliable in identifying masses in the porta hepatis or head of the pancreas. If a Klatskin's tumor is suspected, the absence of a mass on CT suggests resectability. In carcinoma of the pancreas, a dynamic CT scan can define invasion of the portal vein. Upper gastrointestinal series may reveal partial duodenal obstruction requiring gastrojejunostomy. Partial invasion of the C- loop (E sign) is indicative of carcinoma of the pancreas.

Multiple Organ Failure Syndrome

The *multiple organ failure syndrome (MOFS)* may be associated with gastrointestinal disease. This clinical entity is characterized by the progressive but potentially reversible physiologic dysfunction of two or more organs or organ systems that arises after resuscitation from an acute life-threatening event. The classic patient who develops MOFS has overwhelming infection, multiple trauma, massive burn, or massive ischemia. Risk factors for development of this syndrome are listed in Table [22-16](#).

It is estimated that MOFS is the cause of death in at least one-half of all patients who die in the *surgical intensive care unit (SICU)*. When deaths from MOFS secondary to traumatic injury are excluded from consideration, the proportion of MOFS-related SICU deaths may be 70 percent or higher. MOFS complicates 7 to 22 percent of emergency operations, including 30 to 50 percent of emergent operations for intraabdominal sepsis. Once MOFS is established, the mortality rate exceeds 70 percent, and survival from MOFS involving four or more organ systems is unusual.

MOFS or the sepsis syndrome, high-output respiratory failure, or any of a number of synonyms has its origin in two situations that may complicate gastrointestinal disease. Although approximately 75 percent of cases of MOFS are associated with bacterial infection, systemic inflammation without active infection also may precipitate organ failure. Disease processes involving the gastrointestinal tract typically account for the bacterial source giving rise to MOFS. Typical examples include colonic perforation with fecal peritonitis, acute diverticulitis with abscess, and periappendiceal abscess. Activated host immune cells, particularly neutrophils and macrophages, cause tissue injury similar to that produced by MOFS in clinical and animal models (Table [22-17](#)). Diffuse organ injury can be induced not only by excess activation of proinflammatory mediators but also by the deficiency of counterinflammatory mediators such as transforming growth factor beta (TGF- β) and interleukin-10.

A second hypothesis for MOFS is that the gastrointestinal tract may serve as an unseen generator. This involvement of the gastrointestinal tract may result from the gut's dual role as an occult microbial reservoir and a potent immunoregulatory organ whose normal function is to maintain a tonic counterinflammatory influence on immunologic activity. The critically ill patient has rapid colonization of the oral pharynx and proximal gastrointestinal tract by the endogenous microbial species. This colonization then predisposes to nosocomial infection with the same organisms. These infections may develop as a result of subclinical aspiration in the intubated patient or as a consequence of the passage of viable bacteria through the intestinal wall, a phenomenon termed *bacterial translocation*. Although this phenomenon has been readily observed in experimental animal models in response to the same insults that predispose to MOFS, such as trauma, hemorrhage, massive burn injury, pancreatitis, intraabdominal infection, and endotoxemia, evidence in human beings is largely circumstantial, and occasionally contradictory.

The extensive immunologic activities of the gut and liver normally blunt expression of the immune response. This counterbalance may be impaired in critical illness and permit the expression of a systemic inflammatory response. Mesenteric ischemia and hypoxia cause the intestinal epithelial cells to decrease the normal dampening effect on the release of tumor necrosis factor (TNF) by macrophages in response to endotoxemia. Animal models of hemorrhagic shock and intestinal ischemia reperfusion have shown elevated TNF and interleukin-6 in portal venous blood within minutes of restoring normal gut perfusion. Patients undergoing aortic aneurysm repair have elevated portal venous levels of TNF. This portal endotoxemia with resultant Kupffer cell activation triggers a release of potent immunosuppressive activity from alveolar and splenic macrophages. This activation can be lessened by Kupffer cell blockade.

Another theory of MOFS is the “two-hit hypothesis.” Normally, after an acute insult such as infection, traumatic injury, or shock there is an initial flow phase that is a hyperdynamic hemodynamic state lasting 3 to 5 days. If convalescence continues uninterrupted, the patient's course passes into the ebb phase, in which the metabolic response and the patient's physiologic profile change from stress to downgrading catabolism and provide the basis for anabolism to complete recovery. At any point if the flow phase is again activated by even the most subtle second insult, there is a markedly exaggerated host response that is manifested as MOFS.

Seven organ systems may demonstrate dysfunction or failure as part of MOFS (Table [22-18](#)). A typical clinical course of MOFS begins with the onset of *acute respiratory distress syndrome (ARDS)*, tachypnea, tachycardia, pyrexia, progressive hypoxia, and the characteristic “fluffy whiteout” on chest x-ray. The next organ system to show signs of dysfunction or failure is the kidney. Despite an adequate urine output, laboratory tests will show elevation of the blood urea nitrogen and creatinine levels. The third organ to fail in the classic description of MOFS is the liver. Two clinical syndromes have been described: ischemic hepatitis and jaundice. The former, also called “shock liver,” characteristically follows an episode of hypotension. Biochemical laboratory studies include elevations in amino transferase level and prothrombin time and hypoglycemia. Successful resuscitation results in rapid normalization of these markers of hepatic function. Jaundice, which is more common than ischemic hepatitis, typically evolves many days after the inciting physiologic insult. Conjugated hyperbilirubinemia is a prominent feature, but elevation of amino transferase levels and prothrombin time is less pronounced. The pathogenesis is multifactorial and includes ongoing hepatic ischemia, infection, cholestasis induced by total parenteral nutrition, and drug toxicity.

At this point, if the infection or hemorrhagic shock is controlled there is an opportunity to reverse the trend of organ failure, and the mortality rate is approximately 40 percent. Once progressive hepatic and renal failure supervene, mortality rates in excess of 70 percent are observed. Cardiovascular dysfunction in MOFS is evident clinically as hypotension that is refractory to volume challenge (necessitating inotropic and vasopressor support), tachydysrhythmia, and peripheral edema. Right ventricular dysfunction is prominent and may be a consequence of increased pulmonary vascular resistance secondary to the ARDS. Seventy percent of critically ill patients will have some element of CNS dysfunction that presents as alterations in level of consciousness without localizing signs. The mechanism of this altered mentation is poorly understood. Postulated mechanisms include the direct effects of proinflammatory mediators on cerebral function, the development of vasogenic cerebral edema, areas of infarction related to hypotension, and alterations in the

blood-brain barrier that change the composition of interstitial fluid. Hematologic dysfunction includes anemia, transient leukopenia, lymphopenia, and thrombocytopenia. The most common of these is thrombocytopenia; causes include increased consumption, intravascular sequestration, and impaired thrombopoiesis secondary to suppression of bone marrow function. Heparin-induced thrombocytopenia occurs in 10 percent of patients receiving heparin. Disseminated intravascular coagulopathy (DIC) is characterized by derangements in platelet numbers and clotting times and the presence of fibrin degradation products in the plasma.

Until recently, stress gastrointestinal bleeding was a common problem occurring in up to one-quarter of all SICU admissions, but improved techniques of resuscitation and hemodynamic support, earlier diagnosis of infection, and the widespread use of stress ulcer prophylaxis all have contributed to a reduction in the frequency of this complication. Rates of clinically important bleeding in the contemporary SICU have dropped below 4 percent. Other manifestations of gastrointestinal dysfunction include ileus, intolerance of enteral feeding, pancreatitis, and acalculous cholecystitis. Splanchnic hypoperfusion with mucosal acidosis is a pathologic feature common to all of these.

Diarrhea is present in as many as 40 percent of critically ill patients. It results from a decrease of absorptive capability of the small or large intestine and happens because of one or more of the following pathophysiologic mechanisms: (1) the presence of poorly absorbed osmotically reactive solutes, (2) intestinal secretion of water and electrolytes, (3) structural mucosal damage (inflammatory or surgical), and (4) abnormal intestinal motility.

Osmotic diarrhea is typically caused by a rapid reintroduction of enteral feeding to a patient after a prolonged fast. This most likely is because of incomplete digestion of complex carbohydrate in the enteral formulas. For this reason, when beginning enteral feeding, an initial dilution of the hyperosmolar formulas to hypoosmolar is recommended.

Secretory diarrhea usually is caused by bacterial endotoxins, humoral agents, defective neural control, and bile salt fatty acids. The later is not uncommon in surgical patients undergoing resection of the terminal ileum or total abdominal colectomy. Cholestyramine is a nonabsorbable anion exchange resin that binds bile acids in the intestinal lumen. It may be of benefit if bile salts are implicated in the pathogenesis of diarrhea, particularly for short-bowel syndrome, especially if less than 100 cm of ileum has been resected.

Intestinal mucosal structural damage causes a generalized impairment of absorption and diarrhea. Crohn's disease, ischemic bowel disease, and antibiotic-associated colitis are among the primary concerns. Some series report that as many as 25 percent of hospitalized patients treated with antibiotics develop diarrhea. *C. difficile* has been established as the most important cause of colitis. Almost all antibiotics, with the possible exception of aminoglycosides and vancomycin, have been associated with *C. difficile*. It should be suspected in any patient who develops diarrhea during or within 10 weeks of antibiotic therapy, and it may develop after only a single dose. When unrecognized, complications include toxic megacolon, severe electrolyte imbalance, hypoalbuminemia, and acute migratory polyarthritis. Bedside rigid sigmoidoscopy may be diagnostic, but the preferred method of diagnosis is a tissue culture assay for the *C. difficile* toxin. Fecal leukocytes also are present. Inflammatory and ischemic bowel disease also should be considered; rectal mucosal biopsy can be used to distinguish these causes. The most widely used treatment for antibiotic-associated colitis is vancomycin, but metronidazole appears to provide comparable efficacy and is considerably less expensive. Once therapy is initiated, diarrhea resolves gradually within 4 to 5 days.

Acute pancreatitis in the critical care unit patient can be quite protean in its presentation and manifestations. Pancreatitis in the postoperative setting commonly follows surgery of the pancreas, biliary tract, stomach, or duodenum, but it may occur after nonabdominal surgery, including open- heart surgery, hypotensive shock, endoscopic retrograde cholangiopancreatography, malignant ductal obstruction, transurethral prostate resection, and transplantation. It is an uncommon problem that has a mortality rate of 25 to 40 percent. Other causes of pancreatitis in SICU patients include medications (azathioprine, 6-mercaptopurine, thiazides, sulfonamides, furosemide, estrogens, and tetracycline). Hypercalcemia and pancreatitis have been described in patients receiving total parenteral nutrition. Despite lack of specificity, the serum amylase level remains the test most frequently used for the diagnosis. It should also be determined in patients who suddenly develop shock or anuria with or without abdominal pain, exhibit fever, abdominal pain, or shocklike phenomena in the postoperative states, especially after abdominal surgery, develop diabetic coma with shock, or exhibit clinical features suggestive of a myocardial infarction. Treatment is directed at symptoms and toward preventing or treating complications of the disease. It includes aggressive fluid resuscitation for treatment of shock, nothing by mouth to avoid stimulation of the pancreas, and appropriate use of analgesics for pain relief. Quinolone antibiotics may have a role in the treatment of pancreatitis. CT-guided needle aspiration of pancreatic masses (phlegmons) and cyst/fluid collections should be performed to determine the presence of an infected process. Culture results should help in selecting appropriate antibiotic agents. An infected phlegmon or abscess requires open drainage, because CT-guided catheter drainage is ineffectual.

Acalculous cholecystitis is common after multiple trauma, massive thermal injury, and massive blood transfusions. Histologic examination confirms that mucosal ischemia is the initiating event. Treatment options include radiologically guided percutaneous drainage, open cholecystostomy, or laparoscopic cholecystectomy.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 23 - Esophagus and Diaphragmatic Hernia

Jeffrey H. Peters
Tom R. DeMeester

SURGICAL ANATOMY

The esophagus is a muscular tube that starts as the continuation of the pharynx and ends as the cardia of the stomach. When the head is in normal anatomic position, the transition from pharynx to esophagus occurs at the lower border of the sixth cervical vertebra. Topographically this corresponds to the cricoid cartilage anteriorly and the palpable transverse process of the sixth cervical vertebra laterally (Fig. 23-1). The esophagus is firmly attached at its upper end to the cricoid cartilage and at its lower end to the diaphragm; during swallowing, the proximal points of fixation move craniad the distance of one cervical vertebral body.

The esophagus lies in the midline, with a deviation to the left in the lower portion of the neck and upper portion of the thorax, and returns to the midline in the midportion of the thorax near the bifurcation of the trachea (Fig. 23-2). In the lower portion of the thorax, the esophagus again deviates to the left and anteriorly to pass through the diaphragmatic hiatus.

Three normal areas of esophageal narrowing are evident on the barium esophagogram or during esophagoscopy. The uppermost narrowing is located at the entrance into the esophagus and is caused by the cricopharyngeal muscle. Its luminal diameter is 1.5 cm, and it is the narrowest point of the esophagus. The middle narrowing is due to an indentation of the anterior and left lateral esophageal wall caused by the crossing of the left main stem bronchus and aortic arch. The luminal diameter at this point is 1.6 cm. The lowermost narrowing is at the hiatus of the diaphragm and is caused by the gastroesophageal sphincter mechanism. The luminal diameter at this point varies somewhat depending on the distention of the esophagus by the passage of food, but has been measured at 1.6 to 1.9 cm. These normal constrictions tend to hold up swallowed foreign objects, and the overlying mucosa is subjected to injury by swallowed corrosive liquids due to their slow passage through these areas.

Figure 23-3 shows the average distance in centimeters measured during endoscopic examination between the incisor teeth and the cricopharyngeus, aortic arch, and cardia of the stomach. Manometrically, the length of the esophagus between the lower border of the cricopharyngeus and upper border of the lower sphincter varies according to the height of the individual.

The pharyngeal musculature consists of three broad, flat, overlapping fan-shaped constrictors (Fig. 23-4). The opening of the esophagus is collared by the cricopharyngeal muscle, which arises from both sides of the cricoid cartilage of the larynx and forms a continuous transverse muscle band without an interruption by a median raphe. The fibers of this muscle blend inseparably with those of the inferior pharyngeal constrictor above and the inner circular muscle fibers of the esophagus below. Some investigators believe that the cricopharyngeus is part of the inferior constrictor; that is, that the inferior constrictor has two parts, an upper or retrothyroid portion having diagonal fibers, and a lower or retrocricoid portion having transverse fibers. Keith in 1910 showed that these two parts of the same muscle serve totally different functions. The retrocricoid portion serves as the upper sphincter of the esophagus and relaxes when the retrothyroid portion contracts to force the swallowed bolus from the pharynx into the esophagus.

The cervical portion of the esophagus is approximately 5 cm long and descends between the trachea and the vertebral column from the level of the sixth cervical vertebra to the level of the interspace between the first and second thoracic vertebrae posteriorly or the level of the suprasternal notch anteriorly. The recurrent laryngeal nerves lie in the right and left grooves between the trachea and the esophagus. The left recurrent nerve lies somewhat closer to the esophagus than the right owing to the slight deviation of the esophagus to the left and the more lateral course of the right recurrent nerve around the right subclavian artery. Laterally, on the left and right sides of the cervical esophagus are the carotid sheaths and the lobes of the thyroid gland.

The thoracic portion of the esophagus is approximately 20cm long. It starts at the thoracic inlet. In the upper portion of the thorax, it is in intimate relationship with the posterior wall of the trachea and the prevertebral fascia. Just above the tracheal bifurcation, the esophagus passes to the right of the aorta. This anatomic positioning can cause a notch indentation in its left lateral wall on a barium swallow radiogram. Immediately below this notch the esophagus crosses both the bifurcation of the trachea and the left main stem bronchus, owing to the slight deviation of the terminal portion of the trachea to the right by the aorta (Fig. 23-5). From there down the esophagus passes over the posterior surface of the subcarinal lymph nodes, and then descends over the pericardium of the left atrium to reach the diaphragmatic hiatus (Fig. 23-6). From the bifurcation of the trachea downward, both the vagal nerves and the esophageal nerve plexus lie on the muscular wall of the esophagus.

Dorsally, the thoracic esophagus follows the curvature of the spine and remains in close contact with the vertebral bodies. From the eighth thoracic vertebra downward, the esophagus moves vertically away from the spine to pass through the hiatus of the diaphragm. The thoracic duct passes through the hiatus of the diaphragm on the anterior surface of the vertebral column behind the aorta and under the right crus. In the thorax, the thoracic duct lies dorsal to the esophagus between the azygos vein on the right and the descending thoracic aorta on the left.

The abdominal portion of the esophagus is approximately 2cm long and includes a portion of the *lower esophageal sphincter (LES)* (Fig. 23-7). It starts as the esophagus passes through the diaphragmatic hiatus and is surrounded by the phrenoesophageal membrane, a fibroelastic ligament arising from the subdiaphragmatic fascia as a continuation of the transversalis fascia lining the abdomen (Fig. 23-8). The upper leaf of the membrane attaches itself in a circumferential fashion around the esophagus, about 1 to 2 cm above the level of the hiatus. These fibers blend in with the elastic-containing adventitia of the abdominal esophagus and the cardia of the stomach. This portion of the esophagus is subjected to the positive-pressure environment of the abdomen.

The musculature of the esophagus can be divided into an outer longitudinal and an inner circular layer. The upper 2 to 6cm of the esophagus contains only striated muscle fibers. From there on smooth muscle fibers gradually become more abundant. Most of the clinically significant esophageal motility disorders involve only the smooth muscle in the lower two-thirds of the esophagus. When a surgical esophageal myotomy is indicated, the incision needs to extend only this distance.

The longitudinal muscle fibers originate from a cricoesophageal tendon arising from the dorsal upper edge of the anteriorly located cricoid cartilage. The two bundles of muscle diverge and meet in the midline on the posterior wall of the esophagus about 3 cm below the cricoid (see Fig. 23-5). From this point on, the entire circumference of the esophagus is covered by a layer of longitudinal muscle fibers. This configuration of the longitudinal muscle fibers around the most proximal part of the esophagus leaves a V-shaped area in the posterior wall covered only with circular muscle fibers. Contraction of the longitudinal muscle fibers shortens the esophagus. The circular muscle layer of the esophagus is thicker than the outer longitudinal layer. In situ, the geometry of the circular muscle is helical and makes the peristalsis of the esophagus assume a wormlike drive as opposed to segmental and sequential squeezing. As a consequence, severe motor abnormalities of the esophagus assume a corkscrewlike pattern on the barium swallow radiogram.

The cervical portion of the esophagus receives its main blood supply from the inferior thyroid artery. The thoracic portion receives its blood supply from the bronchial arteries, with 75 percent of individuals having one right-sided and two left-sided branches. Two esophageal branches arise directly from the aorta. The abdominal portion of the esophagus receives its blood supply from the ascending branch of the left gastric artery and from inferior phrenic arteries (Fig. 23-9). On entering the wall of the esophagus, the arteries assume a T-shaped division to form a longitudinal plexus giving rise to an intramural vascular network in the muscular and submucosal layers. As a consequence the esophagus can be mobilized from the stomach to the level of the aortic arch without fear of devascularization and ischemic necrosis. Caution should be exercised as to the extent of esophageal mobilization in patients who have had a previous thyroidectomy with ligation of the inferior thyroid arteries proximal to the origin of the esophageal branches.

Blood from the capillaries of the esophagus flows into a submucosal venous plexus and then into a periesophageal venous plexus from which the esophageal veins originate. In the cervical region, the esophageal veins empty into the inferior thyroid vein; in the thoracic region into the bronchial, azygos, or hemiazygos veins; and in the abdominal region into the coronary vein (Fig. 23-10). The submucosal venous networks of the esophagus and stomach are in continuity with each other, and in patients with portal venous obstruction, this communication functions as a collateral pathway for portal blood to enter the superior vena cava via the azygos vein.

The parasympathetic innervation of the pharynx and esophagus is provided mainly by the vagus nerves. The constrictor muscles of the pharynx receive branches from the pharyngeal plexus, which is on the posterior lateral surface of the middle constrictor muscle and is formed by pharyngeal branches of the vagus nerves with a small contribution from the IXth and XIth cranial nerves (Fig. 23-11). The cricopharyngeal sphincter and the cervical portion of the esophagus receive branches from both recurrent laryngeal nerves, which originate from the vagus nerves—the right recurrent nerve at the lower margin of the subclavian artery, the left at the lower margin of the aortic arch. They are slung dorsally around these vessels and ascend in the groove between the esophagus and trachea, giving branches to each. Damage to these nerves interferes not only with the function of the vocal cords but also with the function of the cricopharyngeal sphincter and the motility of the cervical esophagus, predisposing the individual to pulmonary aspiration on swallowing.

Afferent visceral sensory pain fibers from the esophagus end without synapse in the first four segments of the thoracic spinal cord, using a combination of sympathetic and vagal pathways. These pathways are also occupied by afferent visceral sensory fibers from the heart; hence, both organs have similar symptomatology.

The lymphatics located in the submucosa of the esophagus are so dense and interconnected that they constitute a single plexus (Fig. 23-12). There are more lymph vessels than blood capillaries in the submucosa. Lymph flow in the submucosal plexus runs in a longitudinal direction, and on injection of a contrast medium the longitudinal spread is seen to be about six times that of the transverse spread. In the upper two-thirds of the esophagus the lymphatic flow is mostly cephalad, and in the lower third caudad. In the thoracic portion of the esophagus, the submucosal lymph plexus extends over a long distance in a longitudinal direction before penetrating the muscle layer to enter lymph vessels in the adventitia. As a consequence of this nonsegmental lymph drainage, a primary tumor can extend for a considerable length superiorly or inferiorly in the submucosal plexus. Consequently, free tumor cells can follow the submucosal lymphatic plexus in either direction for a long distance before they pass through the muscularis and on into the regional lymph nodes. The cervical esophagus has a more direct segmental lymph drainage into the regional nodes, and as a result, lesions in this portion of the esophagus have less submucosal extension and a more regionalized lymphatic spread.

The efferent lymphatics from the cervical esophagus drain into the paratracheal and deep cervical lymph nodes, and those from the upper thoracic esophagus empty mainly into the paratracheal lymph nodes. Efferent lymphatics from the lower thoracic esophagus drain into the subcarinal nodes and nodes in the inferior pulmonary ligaments. The superior gastric nodes receive lymph not only from the abdominal portion of the esophagus but also from the adjacent lower thoracic segment.

PHYSIOLOGY

Swallowing Mechanism

The act of alimentation requires the passage of food and drink from the mouth into the stomach. One-third of this distance consists of the mouth and hypopharynx, and two-thirds is made up by the esophagus. To comprehend the mechanics of alimentation, it is useful to visualize the gullet as a mechanical model in which the tongue and pharynx function as a piston pump with three valves, and the body of the esophagus and cardia function as a worm drive pump with a single valve. The three valves in the pharyngeal cylinder are the soft

pallet, the epiglottis, and the cricopharyngeus. The valve of the esophageal pump is the LES. Failure of the valves or the pumps leads to abnormalities in swallowing—that is, difficulty in food propulsion from mouth to stomach—or regurgitation of gastric contents into the esophagus or pharynx.

Food is taken into the mouth in a variety of bite sizes, where it is broken up, mixed with saliva, and lubricated. Swallowing, once initiated, is entirely a reflex. When food is ready for swallowing, the tongue, acting like a piston, moves the bolus into the posterior oropharynx and forces it into the hypopharynx (Fig. 23-13). Concomitantly with the posterior movement of the tongue, the soft palate is elevated, thereby closing the passage between the oropharynx and nasopharynx. This partitioning prevents pressure generated in the oropharynx from being dissipated through the nose. When the soft palate is paralyzed, for example, after a cerebral vascular accident, food is commonly regurgitated into the nasopharynx. During swallowing, the hyoid bone moves upward and anteriorly, elevating the larynx and opening the retrolaryngeal space, bringing the epiglottis under the tongue (see Fig. 23-13). The backward tilt of the epiglottis covers the opening of the larynx to prevent aspiration. The whole pharyngeal part of swallowing occurs within 1.5 s.

During swallowing, the pressure in the hypopharynx rises abruptly to at least 60 mmHg due to the backward movement of the tongue and contraction of the posterior pharyngeal constrictors. A sizable pressure difference develops between the hypopharyngeal pressure and the less- than-atmospheric midesophageal or intrathoracic pressure (Fig. 23-14). This pressure gradient speeds the movement of food from the hypopharynx into the esophagus when the cricopharyngeus or upper esophageal sphincter relaxes. The bolus is both propelled by peristaltic contraction of the posterior pharyngeal constrictors and sucked into the thoracic esophagus. Critical to receiving the bolus is the compliance of the cervical esophagus; when compliance is lost due to muscle pathology, dysphagia can result. The upper esophageal sphincter closes within 0.5 s of the initiation of the swallow, with the immediate closing pressure reaching approximately twice the resting level of 30 mmHg. The postrelaxation contraction continues down the esophagus as a peristaltic wave (Fig. 23-15). The high closing pressure and the initiation of the peristaltic wave prevents reflux of the bolus from the esophagus back into the pharynx. After the peristaltic wave has passed farther down the esophagus, the pressure in the upper esophageal sphincter returns to its resting level.

Swallowing can be started at will, or it can be reflexly elicited by the stimulation of areas in the mouth and pharynx, among them the anterior and posterior tonsillar pillars or the posterior lateral walls of the hypopharynx. The afferent sensory nerves of the pharynx are the glossopharyngeal nerves and the superior laryngeal branches of the vagus nerves. Once aroused by stimuli entering via these nerves, the swallowing center in the medulla coordinates the complete act of swallowing by discharging impulses through the Vth, VIIth, Xth, XIth, and XIIth cranial nerves, as well as the motor neurons of C1 to C3. Discharges through these nerves occur in a rather specific pattern and last for approximately 0.5 s. Little is known about the organization of the swallowing center except that it can trigger swallowing after a variety of different inputs, but the response is always a rigidly ordered pattern of outflow. Following a cerebral vascular accident, this coordinated outflow may be altered, causing mild-to-severe abnormalities of swallowing. In more severe injury, swallowing can be grossly disrupted, leading to repetitive aspiration.

The striated muscles of the cricopharyngeus and the upper third of the esophagus are activated by efferent motor fibers distributed through the vagus nerve and its recurrent laryngeal branches. The integrity of innervation is required for the cricopharyngeus to relax in coordination with the pharyngeal contraction and resume its resting tone once a bolus has entered the upper esophagus. Operative damage to the innervation can interfere with laryngeal, cricopharyngeal, and upper esophageal function and predispose the patient to aspiration.

The pharyngeal activity in swallowing initiates the esophageal phase. The body of the esophagus functions as a worm drive propulsive pump, due to the helical arrangement of its circular muscles, and is responsible for transmitting a bolus of food into the stomach. The esophageal phase of swallowing represents esophageal work done during alimentation in that food is moved into the stomach from a negative-pressure environment of - 6mmHg intrathoracic pressure to a positive-pressure environment of 6 mmHg intraabdominal pressure or over a gradient of 12mmHg (see Fig. 23- 14). Effective and coordinated smooth muscle function in the lower third of the esophagus is, therefore, important in pumping the food across this gradient.

The peristaltic wave generates an occlusive pressure varying from 30 to 120 mmHg (see Fig. 23-15). The wave rises to a peak in 1 s, lasts at the peak for about 0.5 s, and then subsides in about 1.5 s. The whole course of the rise and fall of occlusive pressure may occupy one point in the esophagus for 3 to 5 s. The peak of a primary peristaltic contraction initiated by a swallow (primary peristalsis) moves down the esophagus at 2 to 4cm/s and reaches the distal esophagus about 9 s after swallowing starts (see Fig. 23-15). Consecutive swallows produce similar primary peristaltic waves, but when the act of swallowing is rapidly repeated, the esophagus remains relaxed and the peristaltic wave occurs only after the last movement of the pharynx. Progress of the wave in the esophagus is caused by sequential activation of its muscles initiated by efferent vagal nerve fibers arising in the swallowing center.

Continuity of the esophageal muscle is not necessary for sequential activation if the nerves are intact. If the muscles, but not the nerves, are cut across, the pressure wave begins distally below the cut as it dies out at the proximal end above the cut. This allows a sleeve resection of the esophagus to be done without destroying its normal function. Afferent impulses from receptors within the esophageal wall are not essential for progress of the coordinated wave. Afferent nerves, however, do go to the swallowing center from the esophagus, because if the esophagus is distended at any point, a contractual wave begins with a forceful closure of the upper esophageal sphincter and sweeps down the esophagus. This secondary contraction occurs without any movements of the mouth or pharynx. Secondary peristalsis can occur as an independent local reflex to clear the esophagus of ingested material left behind after the passage of the primary wave. Current studies suggest that secondary peristalsis is not as common as once thought.

Despite the powerful occlusive pressure, the propulsive force of the esophagus is relatively feeble. If a subject attempts to swallow a bolus attached by a string to a counterweight, the maximum weight that can be overcome is 5 to 10 g. Orderly contractions of the muscular wall and anchoring of the esophagus at its inferior end are necessary for efficient aboral propulsion to occur. Loss of the inferior anchor, as occurs with a large hiatal hernia, can lead to inefficient propulsion.

The LES provides a pressure barrier between the esophagus and stomach and acts as the valve on the worm drive pump of the esophageal body. Although an anatomically distinct LES has been difficult to identify, microdissection studies show that, in humans, the sphincterlike function is related to the architecture of the muscle fibers at the junction of the esophageal tube with the gastric pouch (Fig. 23-16). The sphincter actively remains closed to prevent reflux of gastric contents into the esophagus and opens by a relaxation that coincides with a pharyngeal swallow (see Fig. 23- 15). The LES pressure returns to its resting level after the peristaltic wave has passed through the esophagus. Consequently, reflux of gastric juice that may occur through the open valve during a swallow is cleared back into the stomach.

If the pharyngeal swallow does not initiate a peristaltic contraction, then the coincident relaxation of the LES is unguarded and reflux of gastric juice can occur. This may be an explanation for the observation of spontaneous lower esophageal relaxation, thought by some to be a causative factor in gastroesophageal reflux disease. The power of the worm drive pump of the esophageal body is insufficient to force open a valve that does not relax. In dogs, a bilateral cervical parasympathetic blockade abolishes the relaxation of the LES that occurs with pharyngeal swallowing or distention of the esophagus. Consequently, vagal function appears to be important in coordinating the relaxation of the LES with esophageal contraction.

The antireflux mechanism in human beings is composed of three components: a mechanically effective LES, efficient esophageal clearance, and an adequately functioning gastric reservoir. A defect of any one of these three components can lead to increased esophageal exposure to gastric juice and the development of mucosal injury.

Physiologic Reflux

On 24-h esophageal pH monitoring, healthy individuals have occasional episodes of gastroesophageal reflux. This physiologic reflux is more common when awake and in the upright position than during sleep in the supine position. When reflux of gastric juice occurs, normal subjects rapidly clear the acid gastric juice from the esophagus regardless of their position.

There are several explanations for the observation that physiologic reflux in normal subjects is more common when they are awake and in the upright position than during sleep in the supine position. First, reflux episodes occur in healthy volunteers primarily during transient losses of the gastroesophageal barrier, which may be due to a relaxation of the LES or intragastric pressure overcoming sphincter pressure. Gastric juice can also reflux when a swallow-induced relaxation of the LES is not protected by an oncoming peristaltic wave. The average frequency of these “unguarded moments” or of transient losses of the gastroesophageal barrier is far less while asleep and in the supine position than while awake and in the upright position. Consequently, there are fewer opportunities for reflux to occur in the supine position. Second, in the upright position there is a 12-mmHg pressure gradient between the resting, positive intraabdominal pressure measured in the stomach and the most negative intrathoracic pressure measured in the esophagus at midthoracic level. This gradient favors the flow of gastric juice up into the thoracic esophagus when upright. The gradient diminishes in the supine position. Third, the LES pressure in normal subjects is significantly higher in the supine position than in the upright position. This is due to the apposition of the hydrostatic pressure of the abdomen to the abdominal portion of the sphincter when supine. In the upright position, the abdominal pressure surrounding the sphincter is negative compared with atmospheric pressure, and, as expected, the abdominal pressure gradually increases the more caudally it is measured. This pressure gradient tends to move the gastric contents toward the cardia and encourages the occurrence of reflux into the esophagus when the individual is upright. By contrast, in the supine position the gastroesophageal pressure gradient diminishes, and the abdominal hydrostatic pressure under the diaphragm increases, causing an increase in sphincter pressure and a more competent cardia.

The LES has intrinsic myogenic tone, which is modulated by neural and hormonal mechanisms. Alpha-adrenergic neurotransmitters or beta blockers stimulate the LES, and alpha blockers and beta stimulants decrease its pressure. It is not clear to what extent cholinergic nerve activity controls LES pressure. The vagus nerve carries both excitatory and inhibitory fibers to the esophagus and sphincter. The hormones gastrin and motilin have been shown to increase LES pressure; and cholecystokinin, estrogen, glucagon, progesterone, somatostatin, and secretin decrease LES pressure. The peptides bombesin, l-enkephalin, and substance P increase LES pressure; and calcitonin gene-related peptide, gastric inhibitory peptide, neuropeptide Y, and vasoactive intestinal polypeptide decrease LES pressure. Some pharmacologic agents such as antacids, cholinergics, agonists, domperidone, metoclopramide, and prostaglandin F₂ are known to increase LES pressure; and anticholinergics, barbiturates, calcium channel blockers, caffeine, diazepam, dopamine, meperidine, prostaglandin E₁ and E₂, and theophylline decrease LES pressure. Peppermint, chocolate, coffee, ethanol, and fat are all associated with decreased LES pressure and may be responsible for esophageal symptoms after a sumptuous meal.

ASSESSMENT OF ESOPHAGEAL FUNCTION

A thorough understanding of the patient's underlying anatomic and functional deficits, prior to therapeutic decisions, is fundamental to the successful treatment of esophageal disease. The diagnostic tests as presently employed may be divided into five broad groups: (1) tests to detect structural abnormalities of the esophagus; (2) tests to detect functional abnormalities of the esophagus; (3) tests to detect increased esophageal exposure to gastric juice; (4) tests to provoke esophageal symptoms; and (5) tests of duodenogastric function as they relate to esophageal disease.

Tests to Detect Structural Abnormalities

Radiographic Evaluation

The first diagnostic test in patients with suspected esophageal disease should be a barium swallow including a full assessment of the stomach and duodenum. Esophageal motility is optimally assessed by observing several individual swallows of barium traversing the entire length of the organ, with the patient in the horizontal position. Hiatal hernias are best

demonstrated with the patient prone because the increased intraabdominal pressure produced in this position promotes displacement of the esophagogastric junction above the diaphragm. To detect lower esophageal narrowing, such as rings and strictures, fully distended views of the esophagogastric region are crucial. The density of the barium used to study the esophagus can potentially affect the accuracy of the examination. Esophageal disorders shown well by a full-column technique include circumferential carcinomas, peptic strictures, large esophageal ulcers, and hiatal hernias. A small hiatal hernia is usually not associated with significant symptoms or illness; and its presence is an irrelevant finding unless the hiatal hernia is large (Fig. 23-17), the hiatal opening is narrow and interrupts the flow of barium into the stomach (Fig. 23-18), or the hernia is of the paraesophageal variety. Lesions extrinsic but adjacent to the esophagus can be reliably detected by the full-column technique if they contact the distended esophageal wall. Conversely, a number of important disorders may go undetected if this is the sole technique used to examine the esophagus. These include small esophageal neoplasms, mild esophagitis, and esophageal varices. Thus, the full-column technique should be supplemented with mucosal relief or double-contrast films to enhance detection of these smaller or more subtle lesions.

Motion-recording techniques greatly aid in evaluating functional disorders of the pharyngoesophageal and esophageal phase of swallowing. The technique and indications for cine- and videoradiography will be discussed later, as it is more useful to evaluate function and seldom used to detect structural abnormalities.

The radiographic assessment of the esophagus is not complete unless the entire stomach and duodenum have been examined. A gastric or duodenal ulcer, partially obstructing gastric neoplasm, or scarred duodenum and pylorus may contribute significantly to symptoms otherwise attributable to an esophageal abnormality.

When a patient's complaints include dysphagia and no obstructing lesion is seen on the barium swallow, it is useful to have the patient swallow a barium-impregnated marshmallow, a barium-soaked piece of bread, or a hamburger mixed with barium. This test may bring out a functional disturbance in esophageal transport that can be missed when liquid barium is used.

Endoscopic Evaluation

In any patient complaining of dysphagia, esophagoscopy is indicated, even in the face of a normal radiographic study. A barium study obtained prior to esophagoscopy is helpful to the endoscopist by directing attention to locations of subtle change and alerting the examiner to such potential danger spots as a cervical vertebral osteophyte, esophageal diverticulum, a deeply penetrating ulcer, or a carcinoma. Regardless of the radiologist's interpretation of an abnormal finding, each structural abnormality of the esophagus should be confirmed visually.

For the initial endoscopic assessment, the flexible fiberoptic esophagoscope is the instrument of choice because of its technical ease, patient acceptance, and the ability to simultaneously assess the stomach and duodenum. Rigid endoscopy may be required in specific instances and should be part of the armamentarium of the endoscopist. The rigid esophagoscope may be an essential instrument when deeper biopsies are required or the cricopharyngeus and cervical esophagus need closer assessment.

When gastroesophageal reflux disease is the suspected diagnosis, particular attention should be paid to detecting the presence of esophagitis and Barrett's columnar-lined esophagus. When endoscopic esophagitis is seen, severity and the length of esophagus involved are recorded. Grade I esophagitis is defined as reddening of the mucosa without ulceration, and its identification varies depending on the observer. Usually its presence may be confirmed by biopsy. Grade I esophagitis is an unreliable indicator of esophagitis and reflects mainly esophageal mucosal injury secondary to a variety of insults. Grade II esophagitis is defined by the presence of linear erosions lined with granulation tissue that bleeds easily when touched. Grade III esophagitis represents a more advanced stage, where the linear ulcerations coalesce, leaving islands of epithelium which on endoscopy appear as a "cobblestone" esophagus. Grade IV esophagitis is the presence of a stricture. Its severity can be assessed by the ease of passing a 36 French endoscope. When a stricture is observed, the severity of the esophagitis above it should be recorded. The absence of esophagitis above a stricture suggests a chemical-induced injury or a neoplasm as a cause. The latter should always be considered and is ruled out only by evaluation of a tissue biopsy of adequate size.

Barrett's esophagus is a condition where the tubular esophagus is lined with columnar epithelium as opposed to the normal squamous epithelium. Histologically it appears as intestinal metaplasia. It is suspected at endoscopy when there is difficulty in visualizing the squamocolumnar junction at its normal location and by the appearance of a redder, more luxuriant mucosa than is normally seen in the lower esophagus. Its presence is confirmed by biopsy. Multiple biopsies should be taken in a cephalad direction to determine the level at which the junction of Barrett's epithelium with normal squamous mucosa occurs. Barrett's esophagus is susceptible to ulceration, bleeding, stricture formation, and, most important, malignant degeneration. The earliest sign of the latter is severe dysplasia or intramucosal adenocarcinoma (Fig. 23-19). These dysplastic changes have a patchy distribution, so a minimum of four biopsy samples every 2 cm should be taken from the Barrett's-lined portion of the esophagus. Changes seen in one biopsy are significant. Nishimaki has determined that the tumors occur in an area of specialized columnar epithelium near the squamocolumnar junction in 85 percent of the patients and within 2 cm of the squamocolumnar junction in virtually all patients. Particular attention should be focused in this area in patients suspected of harboring a carcinoma.

Abnormalities of the gastroesophageal flap valve can be visualized by retroflexion of the endoscope. Hill has graded the appearance of the gastroesophageal valve from I to IV according to the degree of unfolding or deterioration of the normal valve architecture (Figure 23-20). The appearance of the valve correlates with the presence of increased esophageal acid exposure, occurring predominantly in patients with grade III and IV valves.

A hiatal hernia is endoscopically confirmed by finding a pouch lined with gastric rugal folds lying 2 cm or more above the margins of the diaphragmatic crura, identified by having the patient sniff. A prominent sliding hiatal hernia frequently is associated with increased esophageal exposure to gastric juice. When a paraesophageal hernia is observed, particular attention is taken to exclude a gastric ulcer or gastritis within the pouch. The intragastric retroflex or J maneuver is important in evaluating the full circumference of the mucosal lining of the herniated stomach.

When an esophageal diverticulum is seen, it should be carefully explored with the flexible endoscope to exclude ulceration or neoplasia. When a submucosal mass is identified, biopsies are usually not performed. Normally a submucosal leiomyoma or reduplication cyst can be easily dissected away from the intact mucosa, but if a biopsy sample is taken, the mucosa may become fixed to the underlying abnormality. This complicates the surgical dissection by increasing the risk of mucosal perforation.

Tests to Detect Functional Abnormalities

In many patients with symptoms of an esophageal disorder, standard radiographic and endoscopic evaluation fails to demonstrate a structural abnormality. In these situations, esophageal function tests are necessary to identify a functional disorder.

Stationary Manometry

Esophageal manometry is a widely used technique to examine the motor function of the esophagus and its sphincters. Manometry is indicated whenever a motor abnormality of the esophagus is suspected on the basis of complaints of dysphagia, odynophagia, or noncardiac chest pain and the barium swallow or endoscopy does not show a clear structural abnormality. Esophageal manometry is particularly necessary to confirm the diagnosis of specific primary esophageal motility disorders, i.e., achalasia, diffuse esophageal spasm, nutcracker esophagus, and hypertensive LES. It also identifies nonspecific esophageal motility abnormalities and motility disorders secondary to systemic disease such as scleroderma, dermatomyositis, polymyositis, or mixed connective tissue disease. In patients with symptomatic gastroesophageal reflux disease, manometry of the esophageal body can identify a mechanically defective LES and evaluate the adequacy of esophageal peristalsis and contraction amplitude. Manometry has become an essential tool in the preoperative evaluation of patients prior to antireflux surgery, allowing selection of the appropriate procedure based upon the patient's underlying esophageal function.

Esophageal manometry is performed using electronic pressure-sensitive transducers located within the catheter or water-perfused catheters with lateral side holes attached to transducers outside the body. The catheter usually consists of a train of five pressure transducers or five or more water-perfused tubes bound together. The transducers or lateral openings are placed at 5-cm intervals from the tip and oriented radially at 72° from each other around the circumference of the catheter. A special catheter assembly consisting of four lateral openings at the same level, oriented at 90° to each other, is of special use in measuring the three-dimensional vector volume of the LES. Other specially designed catheters can be used to assess the upper sphincter.

As the pressure-sensitive station is brought across the gastroesophageal junction, a rise in pressure above the gastric baseline signals the beginning of the LES. The respiratory inversion point is identified when the positive excursions that occur in the abdominal cavity with breathing change to negative deflections in the thorax. The respiratory inversion point serves as a reference point at which the amplitude of LES pressure and the length of the sphincter exposed to abdominal pressure are measured. As the pressure-sensitive station is withdrawn into the body of the esophagus, the upper border of the LES is identified by the drop in pressure to the esophageal baseline. From these measurements the pressure, abdominal length, and overall length of the sphincter are determined (Fig. 23-21). To account for the asymmetry of the sphincter (Fig. 23-22), the pressure profile is repeated with each of the five radially oriented transducers, and the average values for sphincter pressure above gastric baseline, overall sphincter length, and abdominal length of the sphincter are calculated.

Table 23-1 shows the values for these parameters in 50 normal volunteers without subjective or objective evidence of a foregut disorder. The level at which a deficiency in the mechanics of the LES occurs was defined by comparing the frequency distribution of these values in the 50 healthy volunteers with a population of similarly studied patients with symptoms of gastroesophageal reflux disease. The presence of increased esophageal exposure to gastric juice was documented by 24-h esophageal pH monitoring. Based on these studies, a mechanically defective sphincter is identified by having one or more of the following characteristics: an average LES pressure of less than 6 mmHg, an average length exposed to the positive-pressure environment in the abdomen of 1 cm or less, and an average overall sphincter length of 2 cm or less. Compared with the normal volunteers, these values are below the 2.5 percentile for sphincter pressure and overall length and for abdominal length. It has been shown that the resistance of the sphincter to reflux of gastric juice is determined by the integrated effects of radial pressures extended over the entire length, resulting in three-dimensional computerized imaging of sphincter pressures. Calculating the volume of this image reflects the sphincter's resistance and is called the *sphincter pressure vector volume (SPVV)* (Fig. 23-23). A calculated SPVV less than the 5th percentile is an indication of a mechanically defective sphincter.

In a study of 50 normal volunteers and 150 patients with increased esophageal exposure to gastric juice and various degrees of esophageal mucosal injury, the calculation of the sphincter pressure vector volume increased the ability of manometry to identify a mechanically defective sphincter compared with standard techniques (Fig. 23-24). This was particularly so in patients without mucosal injury and borderline sphincter abnormalities. Three-dimensional LES manometry and calculation of the vector volume should, therefore, become the standard technique to assess the barrier function of the LES in patients with gastroesophageal reflux disease. Patients with gastroesophageal reflux disease and an SPVV below the 5th percentile of normal or a deficiency of one, two, or all three mechanical components of a LES on standard manometry have a mechanical defect of their antireflux barrier that a surgical antireflux procedure is designed to correct.

To assess the relaxation and postrelaxation contraction of the LES, a pressure transducer is positioned within the high-pressure zone, with a distal transducer located in the stomach and the proximal transducer within the esophageal body. Ten wet swallows (5 mL water each) are performed. The normal pressure of the LES should drop to the level of gastric pressure during each wet swallow.

The function of the esophageal body is assessed with the five pressure transducers located in the esophagus. The standard procedure is to locate the most proximal pressure transducer 1 cm below the well-defined cricopharyngeal sphincter, allowing a pressure response throughout the whole esophagus to be obtained on one swallow. Ten wet swallows are recorded. Amplitude, duration, and morphology of contractions following each swallow are calculated at all recorded levels of the esophageal body. The delay between the onset or peak of esophageal contractions at the various levels of the esophagus is used to calculate the speed of wave propagation. The relationship of the esophageal contractions following a swallow is classified as peristaltic or simultaneous. The data are used to identify motor disorders of the esophagus.

The position, length, and pressure of the cricopharyngeal sphincter are assessed with a stationary pull-through technique similar to that used for the LES. The manometric catheter is withdrawn in 0.5-cm intervals from the upper esophagus through the upper esophageal sphincter region into the pharynx. The relaxation of the upper esophageal sphincter is studied by straddling the eight pressure transducers across the sphincter so that some are in the pharynx and some in the upper esophagus. High-speed graphic recordings (50 mm/s) are necessary to obtain an assessment of the coordination of cricopharyngeal relaxation with hypopharyngeal contraction. It has been difficult to consistently demonstrate a motility abnormality in patients with pharyngoesophageal disorders.

24-Hour Ambulatory Manometry

The development of miniaturized electronic pressure transducers and portable digital data recorders with large storage capacity has made ambulatory monitoring of esophageal motor function over an entire circadian cycle possible. The broad clinical application of this new technology in a large number of asymptomatic normal volunteers and patients with primary esophageal motor disorders or gastroesophageal reflux disease provides new insights into esophageal motor function in health and disease under a variety of physiologic conditions. In both normal volunteers and symptomatic patients, esophageal motor activity increases with the state of consciousness and focus on eating activity, i.e., from the supine, to the upright, to meal periods. In the normal situation, there is a higher prevalence of nonperistaltic contractions then appreciated on stationary manometry.

Compared with standard manometry, ambulatory esophageal manometry provides a more than 100 times larger database for the classification and quantification of abnormal esophageal motor function and leads to a change in the diagnosis in a substantial portion of patients with symptoms suggestive of a primary esophageal motor disorder (Fig. 23-25). In patients with nonobstructive dysphagia, the circadian esophageal motor pattern is characterized by an inability to organize the motor activity into peristaltic contraction during a meal period (Fig. 23-26). This finding can be used to provide a new classification of motility disorders on the basis of when they may give rise to dysphagia (Fig. 23-27). In patients with noncardiac chest pain, ambulatory motility monitoring can document a direct correlation of abnormal esophageal motor activity with the symptom and shows that the abnormal motor activity immediately preceding the pain episodes is characterized by an increased frequency of simultaneous double- and triple-peaked high-amplitude long contractions. Ambulatory motility monitoring of patients with gastroesophageal reflux disease shows that the contractility of the esophageal body in these patients deteriorates with increasing severity of esophageal mucosal injury, compromising the clearance function of the esophageal body. Ambulatory esophageal manometry will replace standard manometry in the assessment of esophageal body function and has the potential to improve the diagnosis and management of patients with esophageal motor abnormalities. The combination of ambulatory 24-h esophageal manometry with esophageal and gastric pH monitoring is currently the most physiologic way to assess patients with foregut motility disorders.

Esophageal Transit Scintigraphy

The esophageal transit of a 10-mL water bolus containing technetium 99m sulfur colloid can be recorded with a gamma camera. Using this technique, delayed bolus transit has been shown in patients with a variety of esophageal motor disorders, including achalasia, scleroderma, diffuse esophageal spasm, and nutcracker esophagus.

Video- and Cineradiography

High-speed cinematic or video recording of radiographic studies allows reevaluation by reviewing the studies at various speeds. This technique is more useful than manometry in the evaluation of the pharyngeal phase of swallowing. Observations suggesting oropharyngeal or cricopharyngeal dysfunction include misdirection of barium into the trachea or nasopharynx, prominence of the cricopharyngeal muscle (Fig. 23-28), a Zenker's diverticulum, a narrow pharyngoesophageal segment, and stasis of the contrast medium in the valleculae or hypopharyngeal recesses (Fig. 23-29). These findings are usually not specific, but rather common manifestations of neuromuscular disorders affecting the pharyngoesophageal area. Studies using liquid barium, barium-impregnated solids, or radiopaque pills aid the evaluation of normal and abnormal motility in the esophageal body. Loss of the normal stripping wave or segmentation of the barium column with the patient in the recumbent position correlates with abnormal motility of the esophageal body. In addition, structural abnormalities such as small diverticula, webs, and minimal extrinsic impressions of the esophagus may be recognized only with motion- recording techniques. The simultaneous computerized capture of videofluoroscopic images and manometric tracings is now available, and is referred to as manofluorography. Manofluorographic studies allow precise correlation of the anatomic events, such as opening of the upper esophageal sphincter, with manometric observations, such as sphincter relaxation. Manofluorography, while not widely available, is presently the best means available to evaluate complex functional abnormalities.

Tests to Detect Increased Exposure to Gastric Juice

24-Hour Ambulatory pH Monitoring

The most direct method of measuring increased esophageal exposure to gastric juice is by an indwelling pH electrode. Prolonged monitoring of esophageal pH is performed by placing a pH probe 5 cm above the manometrically measured upper border of the distal sphincter for 24 h. It quantifies the actual time the esophageal mucosa is exposed to gastric juice, measures the ability of the esophagus to clear refluxed acid, and correlates esophageal acid exposure with the patient's symptoms. A 24-h period is necessary so that measurements are made over one complete circadian cycle. This allows measuring the effect of physiologic activity, such as eating or sleeping, on the reflux of gastric juice into the esophagus (Fig. 23-30).

The 24-h esophageal pH monitoring should not be considered a test for reflux, but rather a measurement of the esophageal exposure to gastric juice. The measurement is expressed by the time the esophageal pH was below a given threshold during the 24-h period. This single assessment, although concise, does not reflect how the exposure has occurred: that is, did it occur in a few long episodes or several short episodes? Consequently, two other assessments are necessary: the frequency of the reflux episodes and their duration.

The units used to express esophageal exposure to gastric juice are (1) cumulative time the esophageal pH is below a chosen threshold, expressed as the percent of the total, upright, and supine monitored time; (2) frequency of reflux episodes below a chosen threshold, expressed as number of episodes per 24 h; and (3) duration of the episodes, expressed as the number of episodes greater than 5 min/24 h and the time in minutes of the longest episode recorded. Table 23-2 shows the normal values for these components of the 24-h record at the whole-number pH threshold derived from 50 normal asymptomatic subjects. The upper limits of normal were established at the 95th percentile. Most centers use pH 4 as the threshold.

To combine the result of the six components into one expression of the overall esophageal acid exposure below a pH threshold, a pH score was calculated by using the standard deviation of the mean of each of the six components measured in the 50 normal subjects as a weighting factor. By accepting an abstract zero level 2 SD below the mean, the data measured in normal subjects could be treated as though they had a normal distribution. Thus, any measured patient value could be referenced to this zero point and, in turn, be awarded points based on whether it was below or above the normal mean value for that component, according to this formula:

Component score = Point value - mean / SD + 1

The upper limits of normal for the composite score for each whole-number pH threshold are shown in Table 23-3.

The detection of increased esophageal exposure to acid gastric juice is more dependable than the detection of increased exposure to alkaline gastric juice. The latter is suggested by an increased alkaline exposure time above pH 7 or 8. Increased exposure in this pH range can be caused by abnormal calibration of the pH recorder; dental infection, which increases salivary pH; esophageal obstruction, which results in static pools of saliva with an increase in pH secondary to bacterial overgrowth; or regurgitation of alkaline gastric juice into the esophagus. Using a properly calibrated probe, in the absence of dental infections or esophageal obstruction, the percentage of time the pH is measured above 7 correlates with the concentration of bile acids continuously aspirated over a 24-h period.

When done in a test population with an equal distribution of normal healthy subjects and patients with the classical reflux symptoms and a defective sphincter, 24-h esophageal pH monitoring had a sensitivity and specificity of 96 percent. (Sensitivity is the ability to detect a disease when known to be present; specificity is the ability to exclude the disease when known to be absent.) This gave a predictive value of a positive and a negative test of 96 percent and an overall accuracy of 96 percent. Based on these studies and extensive clinical experience, 24-h esophageal pH monitoring has emerged as a gold standard for the diagnosis of gastroesophageal reflux disease.

24-Hour Ambulatory Bile Monitoring

The potentially injurious components that reflux into the esophagus include gastric secretions, such as acid and pepsin, as well as biliary and pancreatic secretions that regurgitate from the duodenum into the stomach. The presence of duodenal content within the esophagus can now be determined via an indwelling spectrophotometric probe capable of detecting bilirubin (Fig. 23-31). Bilirubin serves as a marker for the presence of duodenal juice. Its absorbance is measured and recorded by a portable optoelectronic data logger capable of directly measuring bilirubin by spectrophotometry, based on its specific light absorption at a wavelength of 453 nm. Figure 23-32 shows the results of 24-h ambulatory bilirubin monitoring in normal subjects compared to those with esophagitis and Barrett's esophagus. Ambulatory bilirubin monitoring can be used to identify patients who are at risk for esophageal mucosal injury and are thus candidates for surgical antireflux treatment.

Standard Acid Reflux Test (SART)

The development of powerful acid reduction agents such as omeprazole has created difficulties in the measurement of esophageal acid exposure. Many patients are placed on the medications prior to study, altering normal physiology and thus complicating interpretation of ambulatory pH monitoring. The acid-reducing effects of omeprazole have been noted to be present as long as 40 days following cessation of the drug, making a prolonged period without the medication necessary before study. Standard acid reflux testing can be used to provide additional information when it is suspected that the results of ambulatory monitoring may have been altered by medications.

The test is performed following manometry by placing a pH electrode 5 cm above the upper border of the LES. The manometry catheter is then advanced temporarily into the stomach, and 300 mL of 0.1 N HCl is infused. In children the gastric acid load is reduced accordingly. The manometry catheter is flushed and then pulled back into the body of the esophagus. The pH of the esophagus is monitored while the patient rests quietly in the supine position and then while he or she performs four maneuvers: deep breathing, Valsalva, Mueller (inspiration against a closed glottis), and cough. These maneuvers are repeated in the right and left lateral decubitus position and with the head down 20°, giving 16 possibilities for acid reflux to occur. A decrease in esophageal pH to less than 4 is considered evidence of reflux. At the beginning of the test, before the patient is placed in the supine position, the distal esophagus must have a pH greater than 4. In some patients this necessitates their standing erect and swallowing repeatedly in order to clear the esophagus of acid.

Patients who fail to clear the esophagus in the erect position after 20 effective swallows monitored on a motility tracing are considered to be abnormal in all positions and maneuvers and are scored as 16. Among 90 healthy volunteers, only 2 individuals had more than two reflux episodes. Accordingly, one or two drops in pH during these challenges to the cardia are considered normal, and three or more drops in pH are taken as evidence of a mechanical incompetence of the cardia. Patients with severe reflux may be unable to clear acid from the esophagus after reflux has been documented.

When used in a test population with an equal distribution of normal healthy subjects and patients with classic symptoms of gastroesophageal reflux disease, the SART had a sensitivity of 59 percent and a specificity of 98 percent. This gave a predictive value of a positive test of 96 percent and a negative test of 75 percent, with an overall accuracy of 81 percent.

Radiographic Detection of Gastroesophageal Reflux

The definition of radiographic gastroesophageal reflux varies depending on whether reflux is spontaneous or induced by various maneuvers. In only about 40 percent of patients with classic symptoms of gastroesophageal reflux disease is spontaneous reflux—i.e., reflux of barium from the stomach into the esophagus with the patient in the upright position—observed by the radiologist. In most patients who show spontaneous reflux on radiography, the diagnosis of increased esophageal acid exposure is confirmed by 24-h esophageal pH monitoring. Therefore, the radiographic demonstration of spontaneous regurgitation of barium into the esophagus in the upright position is a reliable indicator that reflux is present. Failure to see this does not indicate the absence of disease.

Tests of Duodenogastric Function

Esophageal disorders are frequently associated with abnormalities of duodenogastric function. Abnormalities of the gastric reservoir or increased gastric acid secretion can be responsible for increased esophageal exposure to gastric juice. Reflux of alkaline duodenal juice, including bile salts, pancreatic enzymes, and bicarbonate, is thought to have a role in the pathogenesis of esophagitis and complicated Barrett's esophagus. Furthermore, functional disorders of the esophagus are often not confined to the esophagus alone, but are associated with functional disorders of the rest of the foregut, i.e., stomach and duodenum. Tests of duodenogastric function that are helpful to investigate esophageal symptoms include gastric emptying studies, gastric acid analysis, and cholescintigraphy (for the diagnosis of pathologic duodenogastric reflux). The single test of 24-h gastric pH monitoring can be used to identify gastric hypersecretion and imply the presence of duodenogastric reflux and delayed gastric emptying.

Gastric Emptying

Gastric emptying studies are performed with radionuclide-labeled meals. Emptying of solids and liquids can be assessed simultaneously when both phases are marked with different tracers. After ingestion of a labeled standard meal, gamma camera images of the stomach are obtained at 5- to 15-min intervals for 1.5 to 2 h. After correction for decay, the counts in the gastric area are plotted as percentage of total counts at the start of the imaging. The resulting emptying curve can be compared with data obtained in normal volunteers. In general, normal subjects will empty 59 percent of a meal within 90 min.

24-Hour Ambulatory Gastric Emptying

Although a variety of different gamma camera scanning techniques exist, none permits the continuous measurement of gastric emptying under physiologic conditions. Ambulatory assessment of gastric emptying using an intraluminal cadmium-telluride gamma detector has become available (Fig. 23-33). This technique allows continuous measurements without the need for supervision by skilled personnel, and allows the subject to engage in normal activities with assessment of gastric emptying in different postures.

Barostat

Because the stomach is a hollow viscus, filled with air and fluid, direct measurement of the contractions of the gastric wall is difficult, particularly in the proximal stomach. Malagelada and others have introduced an electronic “barostat” to record the volume changes of the proximal stomach in response to a meal, thus providing a direct measure of proximal gastric motility. The barostat consists of a highly compliant low- pressure bag that is introduced into the proximal stomach and filled with air. The system is maintained at a constant pressure via an electronic circuit. Changes in gastric tone are reflected in the increase or decrease in volume of the air-filled bag necessary to maintain a constant pressure (Fig. 23-34). Although its precise role in clinical decision making remains to be determined, it has been shown to be a sensitive and useful technique to measure gastric relaxation following meals in patients.

Gastric Acid Analysis

The gastric secretory state is usually evaluated by determination of the titratable gastric acid in aspirated gastric juice. Interdigestive or basal gastric acid secretion is measured in the fasting state and varies between 0 and 5 mmol/h in normal volunteers. The maximal acid secretory capacity of the stomach, which reflects the available parietal cell mass, is calculated following stimulation of gastric acid secretion with pentagastrin or histamine. Acid hypersecretors have a basal gastric acid secretory capacity of greater than 5 mmol/h and a maximal acid secretory capacity of over 30 mmol/h.

Cholescintigraphy

Scintigraphic hepatobiliary imaging is performed after intravenous injection of 5 mCi of technetium 99m iminodiacetic acid derivatives such as disofenin (99mTc-DISIDA). Gamma camera images of the upper abdomen including the gallbladder and stomach are obtained at 5-min intervals for 60 min. Imaging is continued for an additional 30 min after stimulation of gallbladder contraction with 20 mg/kg of synthetic C-terminal octapeptide of *cholecystokinin* (*CCK*). Duodenogastric reflux is demonstrated as an increase of radioactivity in the stomach in the sequential images. The clinical value of this test is limited due to its short duration and a relatively high false-positive rate in normal volunteers.

24-Hour Gastric pH Monitoring

Monitoring is performed over a complete circadian cycle with a pH electrode placed 5cm below the manometrically located LES. The patient is fully ambulatory during the test and is encouraged to perform normal daily activity. The gastric pH profile is assessed separately for the meal, postprandial period, and fasting period. The latter is divided into the time spent upright and supine.

The interpretation of continuous gastric pH recordings is more difficult than that of esophageal pH recordings. This is because the gastric pH environment is determined by a complex interplay of acid secretion, mucous secretion, ingested food, swallowed saliva, regurgitated duodenal, pancreatic, and biliary secretions, and the effectiveness of the mixing and evacuation of the chyme. Using 24-h gastric pH monitoring to evaluate the gastric secretory state is based on studies that have shown that a good correlation exists between increased basal acid output on standard gastric acid analysis and a left shift on the frequency distribution graph of gastric pH recordings during the supine fasting period. The evaluation of gastric emptying by 24-h gastric pH monitoring is based on studies demonstrating a good correlation between the emptying of a solid meal and the duration of the postprandial plateau and decline phase of the gastric pH record.

Using 24-h gastric pH monitoring to evaluate duodenogastric reflux is based on the observation that reflux of alkaline duodenal juice into the stomach can alkalize the gastric pH environment. The measurement is not straightforward because of the effect of meals, and reduction in acid secretion can result in changes in gastric pH that mimic alkaline reflux episodes. To overcome this problem, computerized measurements of the number and height of alkalizing peaks, the baseline pH, the postprandial pH plateau, and the pattern of pH decline from the plateau can be used to identify the probability of duodenogastric reflux. The results are presented as an overall score that indicates the likelihood of pathologic duodenogastric reflux. Initial data indicate that this approach has a higher sensitivity and specificity for the diagnosis of pathologic duodenogastric reflux than scintigraphic methods do.

Combined 24-h esophageal and gastric pH monitoring can identify excessive alkaline duodenogastric and alkaline gastroesophageal reflux in symptomatic patients. The combined tracings can often identify simultaneous gastric and esophageal alkalization, suggesting a duodenal origin for the esophageal alkaline exposure (Fig. 23-35).

24-Hour Gastric Bilirubin Monitoring

Ambulatory gastric bilirubin monitoring using the Bilitech 2000 spectrophotometric probe is now available, for the detection of duodenogastric reflux, and the test may be superior to gastric pH monitoring as it allows direct measurement of duodenal content, rather than its inference via an alkaline pH. The probe is placed 5 cm below the LES and connected to a portable data logger, and a 24-h ambulatory study is performed on an outpatient basis. The patient is allowed a solid diet consisting of food that does not interfere with the absorbance spectrum of bilirubin. Studies have suggested that duodenal juice is detected in the stomach of both normal subjects and patients most commonly during the supine period and that duodenogastric reflux may be aggravated by previous cholecystectomy.

GASTROESOPHAGEAL REFLUX DISEASE

Definition

Gastroesophageal reflux is a common disease that accounts for approximately 75 percent of esophageal pathology. Despite its high prevalence, it can be one of the most challenging diagnostic and therapeutic problems in benign esophageal disease. A contributing factor to this is the lack of a universally accepted definition of the disease.

The simplest approach is to define the disease by its symptoms. However, symptoms thought to be indicative of gastroesophageal reflux disease, such as heartburn or acid regurgitation, are very common in the general population, and many individuals consider them to be normal and do not seek medical attention. Even when excessive, these symptoms are not specific for gastroesophageal reflux and can be caused by other diseases such as achalasia, diffuse spasm, esophageal carcinoma, pyloric stenosis, cholelithiasis, gastritis, gastric or duodenal ulcer, and coronary artery disease. In addition, patients with gastroesophageal reflux disease can present with atypical symptoms, such as nausea, vomiting, postprandial fullness, chest pain, choking, chronic cough, wheezing, and hoarseness. Further, bronchiolitis, recurrent pneumonia, idiopathic pulmonary fibrosis, and asthma can be primarily due to gastroesophageal reflux disease. To confuse the issue more, gastroesophageal reflux disease can coexist with cardiac and pulmonary disease. Thus using clinical symptoms to define gastroesophageal reflux disease lacks sensitivity and specificity.

An alternative definition for gastroesophageal reflux disease is the presence of endoscopic esophagitis. Using this criterion for diagnosis assumes that all patients who have esophagitis have excessive regurgitation of gastric juice into their esophagus. This is true in 90 percent of patients, but in 10 percent the esophagitis has other causes, the most common being unrecognized chemical injury from drug ingestion. In addition, the definition leaves undiagnosed those patients who have symptoms of gastroesophageal reflux but do not have endoscopic esophagitis.

A third approach to define gastroesophageal reflux disease is to measure the basic pathophysiologic abnormality of the disease, that is, increased exposure of the esophagus to gastric juice. In the past this was inferred by the presence of a hiatal hernia, later by endoscopic esophagitis, and more recently by a hypotensive LES pressure. The development of miniaturized pH electrodes and data recorders allowed measurement of esophageal exposure to gastric juice by calculating the percentage of time the pH was less than 4 over a 24-h period. This provided an opportunity to objectively identify the presence of the disease and stimulated a rational stepwise approach to determining the cause for the abnormal

esophageal exposure to gastric juice.

The Human Antireflux Mechanism and the Pathophysiology of Gastroesophageal Reflux

The human antireflux mechanism consists of a pump, the esophageal body, and a valve, the LES. The common denominator for virtually all episodes of gastroesophageal reflux in both patients and normal subjects is the loss of the normal gastroesophageal barrier to reflux. This is usually secondary to low or reduced LES resistance. The loss of this resistance may be either permanent or transient. A structurally defective sphincter results in a permanent loss of LES resistance and permits unhampered reflux of gastric contents into the esophagus, throughout the circadian cycle. Transient loss of the gastroesophageal barrier may occur secondary to gastric abnormalities including gastric distention with air or food, increased intragastric or intraabdominal pressure, and delayed gastric emptying. These transient losses of sphincter resistance occur in the early stages of gastroesophageal reflux disease and are likely the mechanism for both physiologic and pathophysiologic postprandial reflux. Thus, gastroesophageal reflux disease may begin in the stomach.

Data have shown that transient loss of sphincter resistance is due to gastric distention. This results in shortening of the LES, upright reflux, and inflammatory changes at the gastroesophageal junction secondary to prolapse of esophageal squamous mucosa into the gastric environment (Fig. 23-36). Over time, persistent inflammation and the development of a hiatal hernia result in the permanent loss of LES function. Reflux in the supine position, erosive esophagitis, and Barrett's esophagus follow.

Several studies support the biomechanical effects of a distended stomach in the pathogenesis of gastroesophageal reflux disease and provide a mechanical explanation for why patients with a structurally normal LES may have increased esophageal acid exposure. In vitro bench studies have shown that increasing the diameter of the cardia, as occurs with gastric distention, results in incompetency of the gastroesophageal sphincter by reducing its overall length (Fig. 23-37). Further, as the sphincter length decreases, its resting pressure, as measured by a perfused catheter, also decreases. The decrease usually occurs suddenly when an inefficient length of sphincter is reached. This phenomenon may explain how transient LES relaxations (TLESRs) are caused by gastric distention. When the stomach is distended, gastric wall tension vectors play on the gastroesophageal junction; the resulting forces pull the sphincter open, thereby reducing sphincter length. When a critical length is reached, usually between 1 to 2 cm, the sphincter pressure drops precipitously. As gastric volume or distention increases, sphincter length decreases. If mechanical forces, set in play by gastric distention, are important in pulling open the sphincter and shortening its length, then the anatomic structure of the cardia, i.e., the presence of an acute angle of His or the dome architecture of a hiatal hernia, should influence the ease at which the sphincter was pulled open. Reflux episodes in this situation are more common after meals, when the stomach is distended.

The mechanism by which gastric distention contributes to shortening of sphincter length so that its resistance drops and reflux occurs provides a mechanical explanation for "transient relaxations" of the LES without invoking a neuromuscular reflex. Rather than a "spontaneous" muscular relaxation, there is a mechanical shortening of the sphincter length, as a consequence of gastric distention, to the point where it becomes incompetent. After gastric venting, sphincter length is restored and competence returns until distention again shortens the sphincter and encourages further venting and reflux. This sequence results in the common complaints of repetitive belching and bloating heard from patients with gastroesophageal reflux disease. Gastric distention may initially occur due to overeating, stress aerophagia, or delayed gastric emptying, secondary to fatty diet or a systemic disorder. The distention is subsequently augmented by an increased swallowing frequency that occurs in patients as they repetitively swallow their saliva in an effort to neutralize the acid refluxed into their esophagus.

The consequence of fundic distention, with the LES being "taken up" into the stretched fundus, is that the squamous epithelium of the sphincter is exposed to gastric juice and mucosal injury. This initial step in the pathogenesis of gastroesophageal reflux disease explains why mild esophagitis is usually limited to the very distal esophagus. Erosions in the terminal squamous epithelium caused by this mechanism may also explain the complaint of epigastric pain so often registered by patients with early disease. It may also be the stimulus to increase the swallowing of saliva to bathe the erosions in order to alleviate the discomfort induced by exposure to gastric acid. With increased swallowing come aerophagia, gastric distention, bloating, and repetitive belching. During this process there is repeated exposure of the squamous epithelium to gastric juice, due to the sphincter being "taken up" into the stretched fundus, which may cause erosion, ulceration, fibrosis (ring formation), and metaplasia of the terminal squamous mucosa.

This mechanism is supported by the finding that as the severity of gastroesophageal reflux disease progresses, the length of columnar lining above the anatomic gastroesophageal junction is increased. This suggests that the presence and extent of columnar epithelium lining the distal esophageal sphincter result from a metaplastic process associated with a loss of sphincter function and increased esophageal acid exposure (Fig. 23-38). To interrupt the progression from early to late disease requires preventing gastric distention and the resultant unfolding of the sphincter. Currently, this can only be achieved with an antireflux operative procedure.

In summary, gastroesophageal reflux diseases starts in the stomach (Fig. 23-39). It is caused by gastric distention due to overeating or ingestion of fried foods, typical of the Western diet, which delays gastric emptying. Gastric distention causes unfolding of the sphincter as it is taken up by the distended fundus and exposure of the terminal squamous epithelium within the sphincter to noxious gastric juice. Signs of injury to the exposed squamous epithelium are erosions, ulceration, fibrosis, and columnar metaplasia with an inflammatory infiltrate or foveolar hyperplasia. Intestinal metaplasia within the sphincter may result, as in Barrett's metaplasia of the esophageal body. This process results in the loss of muscle function and the sphincter becomes mechanically defective, allowing free reflux with progressively higher degrees of mucosal injury.

Structural damage to the components of the LES, such as loss of pressure, inadequate overall length, or the loss of abdominal length—i.e., the portion exposed to the positive-pressure environment of the abdomen as measured by manometry—results in permanent failure of sphincter function. The probability of increased exposure to gastric juice is 73 percent if one component of the sphincter is abnormal, 74 percent if two components are abnormal, and 92 percent if all three are abnormal. This indicates that the failure of one or two of the components of the sphincter may be compensated for by the clearance of the esophageal body. Failure of all three sphincter components inevitably leads to increased esophageal exposure to gastric juice.

The most common cause of a structurally defective LES is inadequate sphincter pressure. The reduced pressure is most likely due to an abnormality of myogenic function. This is supported by two observations. First, the location of the LES, in either the abdomen or the chest, is not a major factor in the genesis of the sphincter pressure, since it can still be measured when the chest and abdomen are surgically opened and the distal esophagus is held free in the surgeon's hand. Second, Biancani and co-workers have shown that the distal esophageal sphincter's muscle response to stretch is reduced in patients with an incompetent cardia. This suggests that sphincter pressure depends on the length and tension properties of the sphincter's smooth muscle. Surgical fundoplication has been shown to restore the mechanical efficiency of the sphincter to normal by correcting the abnormal length-tension characteristics.

Although an inadequate pressure is the most common cause of a structurally defective sphincter, the efficiency of a sphincter with normal pressure can be nullified by an inadequate abdominal length or an abnormally short overall resting length. An adequate abdominal length is important in preventing reflux caused by increases in intraabdominal pressure, and an adequate overall length is important in providing the resistance to reflux caused by gastric distention independent of intraabdominal pressure. Therefore, patients with a low sphincter pressure or those with a normal pressure but a short abdominal length are unable to protect against reflux caused by fluctuations of intraabdominal pressure that occur with daily activities or changes in body position. Patients with a low sphincter pressure or those with a normal pressure but short overall length are unable to protect against reflux related to gastric distention caused by outlet obstruction, aerophagia, gluttony, delayed gastric emptying associated with a fatty diet, or various gastropathies. Persons who have a short overall length on a resting motility study are at a disadvantage in protecting against excessive gastric distention secondary to eating, and suffer postprandial reflux. This is because with normal dilatation of the stomach, sphincter length becomes shorter, and if already shortened in the resting state, there is very little tolerance for further shortening before incompetency occurs.

The combined effects of pressure, overall length, and abdominal length on the resistance of the sphincter to the reflux of gastric juice can be determined by integrating the effects of radial pressures extended over the entire length of the sphincter. This can be quantified by three-dimensional computerized imaging of pressures throughout the sphincter length and calculating the volume of this image. This is referred to as the sphincter pressure vector volume. Figure 23-40 shows the three-dimensional sphincter representations for a normal volunteer and a patient with Barrett's esophagus before and after Nissen fundoplication. The marked improvement in the sphincter is apparent.

The second portion of the human antireflux mechanism is an effective esophageal pump which clears the esophagus of physiologic reflux episodes. Ineffective esophageal clearance can result in an abnormal esophageal exposure to gastric juice in individuals who have a normal LES and gastric function but fail to clear physiologic reflux episodes. This situation is relatively rare, and ineffectual clearance is more apt to be seen in association with a structurally defective sphincter, which augments the esophageal exposure to gastric juice by prolonging the duration of each reflux episode.

Four factors important in esophageal clearance are gravity, esophageal motor activity, salivation, and anchoring of the distal esophagus in the abdomen. The loss of any one can augment esophageal exposure to gastric juice by contributing to ineffective clearance. This explains why, in the absence of peristalsis, reflux episodes are prolonged in the supine position. The bulk of refluxed gastric juice is cleared from the esophagus by a primary peristaltic wave initiated by a pharyngeal swallow. Secondary peristaltic waves are initiated by either distention of the lower esophagus or a drop in the intraesophageal pH. Ambulatory motility studies indicate that secondary waves are less common and play less of a role in clearance than previously thought. The esophageal contractions initiated by a drop in esophageal pH rarely have a normal peristaltic pattern; they commonly have a broad-based, powerful synchronous pattern, which reduces the efficiency of esophageal clearance and encourages the regurgitation of refluxed material into the pharynx, predisposing the patient to aspiration.

Manometry of the esophageal body can detect failure of esophageal clearance by analysis of the pressure amplitude and speed of wave progression through the esophagus. The work of Kahrilas and Dodds has shown that the amplitude of an esophageal contraction required to clear the esophagus of liquid barium varies according to the level. Lower segments require a greater amplitude than upper segments. Inadequate amplitude results in ineffective clearance.

Salivation contributes to esophageal clearance by neutralizing the minute amount of acid that is left following a peristaltic wave. Return of esophageal pH to normal takes significantly longer if salivary flow is reduced, such as after radiotherapy, and is shorter if saliva is stimulated by sucking lozenges. Saliva production may also be increased by the presence of acid in the lower esophagus. The patient experiences excessive mucus in the throat. Clinically, this is referred to as "water brash."

A hiatal hernia can also contribute to an esophageal propulsion defect due to loss of anchorage of the esophagus in the abdomen. This results in a reduction in the efficiency of acid clearance (Fig. 23-41). Kahrilas has shown that complete esophageal emptying without retrograde flow was achieved in 86 percent of test swallows in control subjects without a hiatal hernia, 66 percent in patients with a reducing hiatal hernia, and only 32 percent of patients with a nonreducing hiatal hernia. Impaired clearance in patients with nonreducing hiatal hernias suggests that the presence of a hiatal hernia contributes to the pathogenesis of gastroesophageal reflux disease.

Gastric Reservoir

The third component of the human antireflux mechanism is the gastric reservoir. Abnormalities of the gastric reservoir that increase esophageal exposure to gastric juice include gastric dilatation, increased intragastric pressure, persistent gastric reservoir, and increased gastric acid secretion. The effects of gastric dilatation on reflux are discussed above. Increased intragastric pressure may be the result of outlet obstruction due to a scarred pylorus or duodenum, or the result of a vagotomy; it can also be found in the diabetic patient

with gastroparesis. The latter two conditions are secondary to abnormalities of the normal adaptive relaxation of the stomach. The increase in intragastric pressure due to alteration in the pressure-volume relationship in these abnormalities can overcome the sphincter resistance and results in reflux.

A persistent gastric reservoir results from delayed gastric emptying and increases the exposure of the esophagus to gastric juice by accentuating physiologic reflux. It is caused by myogenic abnormalities such as gastric atony in advanced diabetes, diffuse neuromuscular disorders, anticholinergic medications, and postviral infections. Nonmyogenic causes are vagotomy, antropyloric dysfunction, and duodenal dysmotility. Delayed gastric emptying can result in increased exposure of the gastric mucosa to bile and pancreatic juice refluxed from the duodenum into the stomach, with the development of gastritis.

Gastric hypersecretion can increase esophageal exposure to gastric acid juice by the physiologic reflux of concentrated gastric acid. Barlow has shown that 28 percent of patients with increased esophageal exposure to gastric juice measured by 24-h pH monitoring have gastric hypersecretion. A mechanically defective sphincter seems to be more important than gastric hypersecretion in the development of complications of reflux disease. In this respect, gastroesophageal reflux disease differs from duodenal ulcer disease, as the latter is specifically related to gastric hypersecretion.

Complications of Gastroesophageal Reflux

The complications of gastroesophageal reflux result from the damage inflicted by gastric juice on the esophageal mucosa or respiratory epithelium and changes caused by their subsequent repair and fibrosis. Complications due to repetitive reflux are esophagitis, stricture, and Barrett's esophagus; repetitive aspiration may lead to progressive pulmonary fibrosis. The severity of the complications is directly related to the prevalence of a structurally defective sphincter (Table 23-4). The observation that a structurally defective sphincter occurs in 42 percent of patients without complications (most of whom have one or two components failed) suggests that disease may be confined to the sphincter due to compensation by a vigorously contracting esophageal body. Eventually all three components of the sphincter fail, allowing unrestricted reflux of gastric juice into the esophagus and overwhelming its normal clearance mechanisms. This leads to esophageal mucosal injury with progressive deterioration of esophageal contractility, as is commonly seen in patients with strictures and Barrett's esophagus. The loss of esophageal clearance increases the potential for regurgitation into the pharynx with aspiration.

The potential injurious components that reflux into the esophagus include gastric secretions such as acid and pepsin, as well as biliary and pancreatic secretions that regurgitate from the duodenum into the stomach. There is a considerable body of experimental evidence to indicate that maximal epithelial injury occurs during exposure to bile salts combined with acid and pepsin. These studies have shown that acid alone does minimal damage to the esophageal mucosa, but the combination of acid and pepsin is highly deleterious. Similarly, the reflux of duodenal juice alone does little damage to the mucosa, while the combination of duodenal juice and gastric acid is particularly noxious (Table 23-5).

Experimental animal studies have shown that the reflux of duodenal contents into the esophagus enhances inflammation, increases the prevalence of Barrett's esophagus, and results in the development of esophageal adenocarcinoma. The component of duodenal juice thought to be most damaging are the bile acids. In order for bile acids to injure mucosal cells, it is necessary that they be both soluble and unionized, so that the unionized nonpolar form may enter mucosal cells. Before the entry of bile into the gastrointestinal tract, 98 percent of bile acids are conjugated with either taurine or glycine in a ratio of about 3:1. Conjugation increases the solubility and ionization of bile acids by lowering their pKa. At the normal duodenal pH of approximately 7, over 90 percent of bile salts are in solution and completely ionized. At pH ranges from 2 to 7, there is a mixture of the ionized salt and the lipophilic, nonionized acid. Acidification of bile to below pH 2 results in an irreversible bile acid precipitation. Consequently, under normal physiologic conditions, bile acids precipitate and are of minimal consequence when an acid gastric environment exists. On the other hand, in a more alkaline gastric environment, such as occurs with excessive duodenogastric reflux and after acid suppression therapy or vagotomy and partial or total gastrectomy, bile salts remain in solution, are partially dissociated, and when refluxed into the esophagus can cause severe mucosal injury by crossing the cell membrane and damaging the mitochondria.

Complications of gastroesophageal reflux such as esophagitis, stricture, and Barrett's metaplasia occur in the presence of two predisposing factors: a mechanically defective LES and an increased esophageal exposure to fluid with a pH of <4 and >7 (Fig. 23-42). The duodenal origin of esophageal contents in patients with an increased exposure to a pH >7 has been confirmed by esophageal aspiration studies (Fig. 23-43). Studies have clarified and expanded these observations by measuring esophageal bilirubin exposure over a 24-h period as a marker for the presence of duodenal juice. Direct measurement of esophageal bilirubin exposure as a marker for duodenal juice has shown that 58 percent of patients with gastroesophageal reflux disease have increased esophageal exposure to duodenal juice and that this exposure occurs most commonly when the esophageal pH is between 4 and 7 (Fig. 23-44). Further, it is associated with a more severe mucosal injury (Fig. 23-45).

The fact that the combination of refluxed gastric and duodenal juice is more noxious to the esophageal mucosa than gastric juice alone may explain the repeated observation that 25 percent of patients with reflux esophagitis develop recurrent and progressive mucosal damage, often despite medical therapy. A potential reason is that acid suppression therapy is unable to consistently maintain the pH of refluxed gastric and duodenal juice above 6. Lapses into pH ranges from 2 to 6 encourage the formation of undissociated, nonpolarized, soluble bile acids, which are capable of penetrating the cell wall and injuring mucosal cells. To assure that bile acids remain completely ionized in their polarized form, and thus unable to penetrate the cell, requires that the pH of the refluxed material be maintained above 7, 24 h a day, 7 days a week, for the patient's lifetime. In practice this would not only be impractical but likely impossible, unless very high doses of medications were used. The use of lesser doses would allow esophageal mucosal damage to occur while the patient was relatively asymptomatic. Antireflux operative procedures reestablish the barrier between stomach and esophagus, protecting the esophagus from damage in patients with mixed gastroesophageal reflux. If reflux of gastric juice is allowed to persist and sustained or repetitive esophageal injury occurs, two sequelae can result. First, a luminal stricture can develop from submucosal and eventually intramural fibrosis. Second, the tubular esophagus may become replaced with columnar epithelium. The columnar epithelium is resistant to acid and is associated with the alleviation of the complaint of heartburn. This columnar epithelium often becomes intestinalized, identified histologically by the presence of goblet cells. This specialized intestinal metaplasia is currently required for the diagnosis of Barrett's esophagus. Endoscopically, Barrett's esophagus can be quiescent or associated with complications of esophagitis, stricture, Barrett's ulceration, and dysplasia. The complications associated with Barrett's esophagus may be due to the continuous irritation from refluxed duodenogastric juice. This continued injury is pH-dependent and may be modified by medical therapy. The incidence of metaplastic Barrett's epithelium becoming dysplastic and progressing to adenocarcinoma is approximately 1 percent per year.

An esophageal stricture can be associated with severe esophagitis or Barrett's esophagus. In the latter situation, it occurs at the site of maximal inflammatory injury, i.e., the columnar-squamous epithelial interface. As the columnar epithelium advances into the area of inflammation, the inflammation extends higher into the proximal esophagus and the site of the stricture moves progressively up the esophagus. Patients who have a stricture in the absence of Barrett's esophagus should have the presence of gastroesophageal reflux documented before the presence of the stricture is ascribed to reflux esophagitis. In patients with normal acid exposure, the stricture may be due to cancer or a drug-induced chemical injury, the latter resulting from the lodgment of a capsule or tablet in the distal esophagus. In such patients, dilation usually corrects the problem of dysphagia. Heartburn, which may have occurred only because of the chemical injury, need not be treated. It is also possible for drug-induced injuries to occur in patients who have underlying esophagitis and a distal esophageal stricture secondary to gastroesophageal reflux. In this situation, a long stringlike stricture progressively develops as a result of repetitive caustic injury from capsule or tablet lodgment on top of an initial reflux stricture. These strictures are often resistant to dilation.

When the refluxed gastric juice is of sufficient quantity, it can reach the pharynx, with the potential for pharyngeal tracheal aspiration, causing symptoms of repetitive cough, choking, hoarseness, and recurrent pneumonia. This is often an unrecognized complication of gastroesophageal reflux disease, since either the pulmonary or the gastrointestinal symptoms may predominate in the clinical situation and focus the physician's attention on one to the exclusion of the other. Three factors are important in these patients. First, it may take up to 7 days for the loss of respiratory epithelium secondary to the aspiration of gastric contents to be recovered, and a chronic cough may develop, between episodes of aspiration, that is not related to a reflux episode. Second, the presence of an esophageal motility disorder is observed in 75 percent of patients with reflux-induced aspiration and is believed to promote the aboral movement of the refluxate toward the pharynx. Third, if the pH in the cervical esophagus in patients with increased esophageal acid exposure is below 4 for less than 1 percent of the time, there is a high probability that the respiratory symptoms have been caused by aspiration. Increasingly, benign pulmonary pathology is recognized as secondary to gastroesophageal reflux disease including asthma, idiopathic pulmonary fibrosis, and bronchiectases.

Medical Therapy

Gastroesophageal reflux disease is such a common condition that most patients with mild symptoms carry out self-medication. Patients when first seen with symptoms of heartburn without obvious complications can reasonably be placed on 8 to 12 weeks of simple antacids before extensive investigations are carried out. In many situations, this successfully aborts the attacks. Patients should be advised to elevate the head of the bed, avoid tight clothing, eat small, frequent meals, avoid eating their nighttime meal shortly before retiring, lose weight, and avoid alcohol, coffee, chocolate, and peppermints, which may aggravate the symptoms.

Alginic acid, used in combination with simple antacids, may augment the relief of symptoms by creating a physical barrier to reflux as well as by acid reduction. Alginic acid reacts with sodium bicarbonate in the presence of saliva to form a highly viscous solution that floats like a raft on the surface of the gastric contents. When reflux occurs, this protective layer is refluxed into the esophagus and acts as a protective barrier against the noxious gastric contents. Medications to promote gastric emptying such as metoclopramide, domperidone, or cisapride are beneficial in early disease but of little value in more severe disease.

In patients with persistent symptoms, the mainstay of medical therapy is acid suppression. High dosage regimens of hydrogen potassium proton pump inhibitors, such as omeprazole (up to 40mg/day) can reduce gastric acidity by as much as 80 to 90 percent. This usually heals mild esophagitis. In severe esophagitis, healing may occur in only half of the patients. In patients who reflux a combination of gastric and duodenal juice, acid suppression therapy may give relieve symptoms while still allowing mixed reflux to occur. This can allow persistent mucosal damage in an asymptomatic patient. Unfortunately, within 6 months of discontinuation of any form of medical therapy for gastroesophageal reflux disease, 80 percent of patients have a recurrence of symptoms.

Suggested Therapeutic Approach

The traditional stepwise approach to the therapy of gastroesophageal reflux disease should be reexamined in view of a more complete understanding of the pathophysiology of gastroesophageal reflux, the rising incidence of Barrett's esophagus, and the increasing mortality rates associated with end-stage reflux disease. The approach should be to identify risk factors for persistent and progressive disease early in the course of the disease and encourage surgical treatment when these factors are present. The following approach is suggested.

Patients presenting for the first time with symptoms suggestive of gastroesophageal reflux may be given initial therapy with H₂ blockers. In view of the availability of these as over-the-counter medication, many patients will have already self-medicated their symptoms. Failure of H₂ blockers to control the symptoms, or immediate return of symptoms after stopping treatment, suggests either that the diagnosis is incorrect or that the patient has relatively severe disease. Endoscopic examination at this stage of the patient's evaluation provides the opportunity for assessing the severity of mucosal damage and the presence of Barrett's esophagus. Both of these findings on initial endoscopy are associated with a high probability that medical treatment will fail. A measurement of the degree and pattern of esophageal exposure to gastric and duodenal juice, via 24-h pH and bilirubin monitoring, should be obtained at this point. The status of the LES and the function of the esophageal body should also be measured. These studies identify features such as the following, which are predictive of a poor response to medical therapy, frequent relapses, and the development of complications: supine reflux, poor esophageal contractility, erosive esophagitis (or a columnar-lined esophagus at initial presentation), bile in the refluxate, and a structurally defective sphincter. Patients who have these risk factors should be given

the option of surgery as a primary therapy with the expectation of long-term control of symptoms and complications.

Selection of Patients for Surgery

Studies on the natural history of gastroesophageal reflux disease indicate that most patients have a relatively benign form of the disease that is responsive to lifestyle changes and dietary and medical therapy, and do not need surgical treatment. Approximately 25 to 50 percent of the patients with gastroesophageal reflux disease have persistent or progressive disease, and it is this patient population that is best suited to surgical therapy. These patients are identified by the same risk factors that predict a poor response to medical therapy. In the past, the presence of esophagitis and a structurally defective LES were the primary indications for surgical treatment, and many internists and surgeons were reluctant to recommend operative procedures in their absence. However, one should not be deterred from considering antireflux surgery in a symptomatic patient with or without esophagitis or a defective sphincter, provided the disease process has been objectively documented by 24-h pH monitoring. This is particularly true in patients who have become dependent upon therapy with proton pump inhibitors, or require increasing doses to control their symptoms. It is important to note that a good response to medical therapy in this group of patients predicts an excellent outcome following antireflux surgery.

A structurally defective LES is the most important factor predicting failure of medical therapy. While patients with normal sphincter pressures tend to remain well controlled with medical therapy, patients with a structurally defective LES do not respond well to medical therapy, usually developing recurrent symptoms within 1 to 2 years of the onset of therapy, and these patients should be considered for an antireflux operation, regardless of the presence or absence of endoscopic esophagitis.

Young patients with documented reflux disease and a defective LES are also excellent candidates for antireflux surgery. They will invariably require long-term medical therapy for control of their symptoms and many will go on to develop complications of the disease. An analysis of the cost of therapy based upon data from the Veterans Administration Cooperative trial indicates that surgery has a cost advantage over medical therapy in patients less than 49 years of age.

Severe endoscopic esophagitis in a symptomatic patient with a structurally defective LES is also an indication for early surgical therapy. These patients are prone to breakthrough of their symptoms while receiving medical therapy. Symptoms and mucosal injury can be controlled in such patients, but careful monitoring is required, and increasing dosages of proton pump inhibitors are necessary. In everyday clinical practice, however, such treatment can be both difficult and impractical, and in such cases, antireflux surgery should be considered early as a therapeutic option.

The development of a stricture in a patient represents a failure of medical therapy, and is also an indication for a surgical antireflux procedure. In addition, strictures are often associated with a structurally defective sphincter and loss of esophageal contractility. Before proceeding with surgical treatment, malignancy and a drug etiology of the stricture should be excluded, and the stricture progressively dilated up to a 60 French bougie. When the stricture is fully dilated, the relief of dysphagia is evaluated and esophageal manometry is performed to determine the adequacy of peristalsis in the distal esophagus. If dysphagia is relieved and the amplitude of esophageal contractions is adequate, an antireflux procedure should be performed; if there is a global loss of esophageal contractility, caution should be exercised in performing an antireflux procedure with a complete fundoplication and a partial fundoplication should be considered.

Barrett's columnar-lined esophagus is commonly associated with a severe structural defect of the LES and often poor contractility of the esophageal body. Patients with Barrett's esophagus are at risk of progression of the mucosal abnormality up the esophagus, formation of a stricture, hemorrhage from a Barrett's ulcer, and the development of an adenocarcinoma. An antireflux procedure may arrest the progression of the disease, heal ulceration, and resolve strictures. Evidence is accumulating that surgical treatment also reduces the risk of progression to cancer. If severe dysplasia or intramucosal carcinoma is found on mucosal biopsy specimens, an esophageal resection should be done.

The ideal therapy of gastroesophageal reflux disease can be viewed conceptually along the continuum depicted in Fig. 23-46. The majority of patients requiring treatment have a relatively mild form of disease and will respond to antisecretory medications. Patients with more severe forms of disease, particularly those with risk factors predictive of medical failure and those who develop persistent or progressive disease, should be considered for early definitive therapy. Laparoscopic Nissen fundoplication will provide a long-term cure in the majority of these patients, with minimal discomfort, and an early return to normal activity. Patients who present with long-standing disease associated with poor esophageal function or a short esophagus should undergo an open antireflux procedure tailored to their underlying anatomic and physiologic abnormalities. Finally, if the disease has resulted in global failure of esophageal contractility, Barrett's metaplasia with high-grade dysplasia, or esophageal adenocarcinoma, an esophagectomy is required.

Surgical Therapy

Preoperative Evaluation

Before proceeding with an antireflux operation, several factors should be evaluated. First, the propulsive force of the body of the esophagus should be evaluated by esophageal manometry to determine if it has sufficient power to propel a bolus of food through a newly reconstructed valve. Patients with normal peristaltic contractions do well with a 360° Nissen fundoplication. When peristalsis is absent or severely disordered or the amplitude of the contraction is below 20 mmHg throughout the lower esophagus, the Belsey two-thirds partial fundoplication is the procedure of choice.

Second, anatomic shortening of the esophagus can compromise the ability to do an adequate repair without tension and lead to an increased incidence of breakdown or thoracic displacement of the repair. Esophageal shortening is identified on a barium swallow roentgenogram by a sliding hiatal hernia that will not reduce in the upright position or that measures larger than 5 cm between the diaphragmatic crura and gastroesophageal junction on endoscopy. When esophageal shortening is present, the motility of the esophageal body must be carefully evaluated and, if inadequate, a gastroplasty should be performed. In patients who have a global absence of contractility, more than 50 percent interrupted or dropped contractions, or a history of several failed previous antireflux procedures, esophageal resection should be considered as an alternative.

Third, the surgeon should specifically query the patient for complaints of nausea, vomiting, and loss of appetite. In the past, these symptoms were accepted as part of the reflux syndrome, but we now realize that they can be due to excessive duodenogastric reflux or gastric pathology. This problem is most pronounced in patients who have had previous upper gastrointestinal surgery, particularly cholecystectomy, although this is not always the case. In such patients, these symptoms may persist after an antireflux procedure, and patients should be given this information before the operation. In these patients, 24-h bilirubin monitoring and gastric emptying studies can be performed to detect and quantify duodenogastric abnormalities. Antireflux surgery alone may influence these symptoms by improving the efficiency of gastric emptying.

Fourth, approximately 30 percent of patients with proven gastroesophageal reflux on 24-h pH monitoring will have hypersecretion on gastric analysis, and 2 to 3 percent of patients who have an antireflux operation will develop a gastric or duodenal ulcer. The presence of *Helicobacter pylori* should be assessed in these patients and treated if present.

Principles of Surgical Therapy

The primary goal of antireflux surgery is to safely restore the structure of the sphincter or to prevent its shortening with gastric distention, while preserving the patient's ability to swallow normally, to belch to relieve gaseous distention, and to vomit when necessary. Regardless of the choice of the procedure, this goal can be achieved if attention is paid to five principles in reconstructing the cardia. First, the operation should restore the pressure of the distal esophageal sphincter to a level twice resting gastric pressure—i.e., 12 mmHg for a gastric pressure of 6 mmHg—and its length to at least 3 cm. This not only augments sphincter characteristics in patients in which they are reduced prior to surgery but prevents unfolding of a normal sphincter in response to gastric distention (Fig. 23-47). Preoperative and postoperative esophageal manometry measurements have shown that the resting sphincter pressure and the overall sphincter length can be surgically augmented over preoperative values, and that the change in the former is a function of the degree of gastric wrap around the esophagus (Fig. 23-48).

Second, the operation should place an adequate length of the distal esophageal sphincter in the positive-pressure environment of the abdomen by a method that ensures its response to changes in intraabdominal pressure. The permanent restoration of 1.5 to 2 cm of abdominal esophagus in a patient whose sphincter pressure has been augmented to twice resting gastric pressure will maintain the competency of the cardia over various challenges of intraabdominal pressure. All three of the popular antireflux procedures increase the length of the sphincter exposed to abdominal pressure by an average of 1 cm. When poorly performed, however, an operation may result in a reduction of the length of abdominal sphincter. Increasing the length of sphincter exposed to abdominal pressure will improve competency only if it is acted on by challenges of intraabdominal pressure. The creation of a conduit that will ensure the transmission of intraabdominal pressure changes around the abdominal portion of the sphincter is a necessary aspect of surgical repair. The fundoplication in the Nissen and Belsey repairs serves this purpose.

Third, the operation should allow the reconstructed cardia to relax on deglutition. In normal swallowing a vagally mediated relaxation of the distal esophageal sphincter and the gastric fundus occurs. The relaxation lasts for approximately 10 s and is followed by a rapid recovery to the former tonicity. To ensure relaxation of the sphincter, three factors are important: (1) only the fundus of the stomach should be used to buttress the sphincter, since it is known to relax in concert with the sphincter; (2) the gastric wrap should be properly placed around the sphincter and not incorporate a portion of the stomach or be placed around the stomach itself, since the body of the stomach does not relax with swallowing; and (3) damage to the vagal nerves during dissection of the thoracic esophagus should be avoided because it may result in failure of the sphincter to relax.

Fourth, the fundoplication should not increase the resistance of the relaxed sphincter to a level that exceeds the peristaltic power of the body of the esophagus. The resistance of the relaxed sphincter depends on the degree, length, and diameter of the gastric fundic wrap and on the variation in intraabdominal pressure. A 360° gastric wrap should be no longer than 2 cm and constructed over a 60 French bougie. This will ensure that the relaxed sphincter will have an adequate diameter with minimal resistance. This is not necessary when constructing a partial wrap.

Fifth, the operation should ensure that the fundoplication can be placed in the abdomen without undue tension and maintained there by approximating the crura of the diaphragm above the repair. Leaving the fundoplication in the thorax converts a sliding hernia into a paraesophageal hernia with all the complications associated with that condition. Maintaining the repair in the abdomen under tension predisposes to an increased incidence of recurrence. This can occur in patients who have a stricture or Barrett's esophagus and is due to shortening of the esophagus from the inflammatory process. This problem can be resolved by lengthening the esophagus by gastroplasty and constructing a partial fundoplication.

Procedure Selection

Selection of the surgical procedure and approach is based upon an assessment of esophageal contractility and length (Fig. 23-49). A transabdominal approach is used in patients with normal esophageal contractility and length. Patients with poor contractility or questionable esophageal length are approached transthoracically. Those with weak esophageal contractions and/or abnormal wave progression are treated with a partial fundoplication in order to avoid the increased outflow resistance associated with a complete fundoplication. If the esophagus is short after it is mobilized from diaphragm to aortic arch, a Collis gastroplasty is done to provide additional length and avoid placing the repair under tension. In the majority of patients who have good esophageal contractility and normal esophageal length the laparoscopic Nissen fundoplication is the procedure of choice for a primary antireflux repair. Experience and randomized studies have shown that the Nissen fundoplication is an effective and durable antireflux repair with minimal side effects that provides long-lasting relief of reflux symptoms in over 90 percent of patients.

Primary Antireflux Repairs Nissen Fundoplication

The most common antireflux procedure is the Nissen fundoplication. The procedure can be performed through an abdominal or a chest incision, and more recently through a laparoscope. Rudolph Nissen described the procedure as a 360° fundoplication around the lower esophagus for a distance of 4 to 5 cm. Although this provided good control of reflux, it was associated with a number of side effects which have encouraged modifications of the procedure as originally described. These include using only the gastric fundus to envelope the esophagus in fashion analogous to a Witzel jejunostomy, sizing the fundoplication with a 60 French bougie, and limiting the length of the fundoplication to 1 to 2 cm. The essential elements necessary for the performance of a transabdominal fundoplication are common to both the laparoscopic and open procedures and include the following:

- (1) Crural dissection, identification, and preservation of both vagi, and the anterior hepatic branch
- (2) Circumferential dissection of the esophagus
- (3) Crural closure
- (4) Fundic mobilization by division of short gastric vessels
- (5) Creation of a short, loose fundoplication by placing the posterior fundic wall posterior, and the anterior fundus anterior, to the esophagus, meeting at the right lateral position.

The Laparoscopic Approach

Laparoscopic fundoplication has become commonplace and has replaced the open abdominal Nissen fundoplication as the procedure of choice. Five 10-mm ports are utilized (Fig. 23-50). Dissection is begun by an incision of the portion of the gastrohepatic omentum above the hepatic branch of the anterior vagal nerve. The circumference of the diaphragmatic crura is dissected and the esophagus is mobilized by careful dissection of the anterior and posterior soft tissues within the hiatus. The esophagus is held anterior and to the left and the crura approximated with three to four interrupted 0 silk sutures, starting just above the aortic decussation and working anterior. Complete fundic mobilization allows construction of a tension-free fundoplication. Short gastric vessels along the upper third of the greater curvature are sequentially dissected and divided. Following complete mobilization, the posterior wall of the fundus is gently brought behind the esophagus to the right side. The anterior wall of the fundus is brought anterior to the esophagus and the fundic lips are manipulated to allow the fundus to envelope the esophagus without twisting (Fig. 23-51). A 60 French bougie is passed to properly size the fundoplication, and it is sutured utilizing a single U stitch of 2-0 Prolene buttressed with felt pledgets.

Transthoracic Nissen Fundoplication

The indications for performing an antireflux procedure by a transthoracic approach are as follows:

- (1) A patient who has had a previous hiatal hernia repair. In this situation, a peripheral circumferential incision in the diaphragm is made to provide simultaneous exposure of the upper abdomen. This allows safe dissection of the previous repair from both the abdominal and thoracic sides of the diaphragm.
- (2) A patient who requires a concomitant esophageal myotomy for achalasia or diffuse spasm.
- (3) A patient who has a short esophagus. This is usually associated with a stricture or Barrett's esophagus. In this situation, the thoracic approach is preferred to allow maximum mobilization of the esophagus and to perform a Collis gastroplasty in order to place the repair without tension below the diaphragm.
- (4) A patient with a sliding hiatal hernia that does not reduce below the diaphragm during a roentgenographic barium study in the upright position. This can indicate esophageal shortening, and, again, a thoracic approach is preferred for maximum mobilization of the esophagus and, if necessary, the performance of a Collis gastroplasty.
- (5) A patient who has associated pulmonary pathology. In this situation, the nature of the pulmonary pathology can be evaluated and the proper pulmonary surgery, in addition to the antireflux repair, can be performed.
- (6) An obese patient. In this situation, the abdominal repair is difficult because of poor exposure, particularly in men in whom the intraabdominal fat is more abundant.

In the thoracic approach the hiatus is exposed through a left posterior lateral thoracotomy incision in the sixth intercostal space, i.e., over the upper border of the seventh rib. When necessary, the diaphragm is incised circumferentially 2 to 3 cm from the lateral chest wall for a distance of approximately 10 to 15cm. The esophagus is mobilized from the level of the diaphragm to underneath the aortic arch. Mobilization up to the aortic arch is usually necessary to place the repair in a patient with a shortened esophagus into the abdomen without undue tension. Failure to do this is one of the major causes for subsequent breakdown of a repair and return of symptoms. The cardia is then freed from the diaphragm. When all the attachments between the cardia and diaphragmatic hiatus are divided, the fundus and part of the body of the stomach are drawn up through the hiatus into the chest. The vascular fat pad that lies at the gastroesophageal junction is excised. Crural sutures are then placed to close the hiatus and the fundoplication constructed by enveloping the fundus around the distal esophagus in a manner similar to that described for the abdominal approach. When complete, the fundoplication is placed into the abdomen by compressing the fundic ball with the hand and manually maneuvering it through the hiatus.

Belsey Mark IV Partial Fundoplication

In the presence of altered esophageal motility, where the propulsive force of the esophagus is not sufficient to overcome the outflow obstruction of a complete fundoplication, a partial fundoplication is indicated. Although a partial fundoplication may be performed laparoscopically (Toupet fundoplication), the Belsey Mark IV repair is the prototype of partial fundoplications. It consists of a 270° gastric fundoplication around the distal 4 cm of esophagus performed through a left chest incision (Fig. 23-52). The dissection of the Belsey Mark IV and the transthoracic Nissen operations are the same, differing only in the technique of constructing the fundoplication.

To perform the Belsey Mark IV partial fundoplication, the esophagus is mobilized up to the aortic arch, the cardia is dissected free of the hiatus, and the fundus of the stomach is brought up through the hiatus as described for the transthoracic Nissen procedure. The partial fundoplication is held in place by two rows of three horizontal mattress sutures placed equidistantly between the seromuscular layers of the stomach and the muscular layers of the esophagus. A second row of sutures is placed 1.5 to 2.0 cm above the first row, using the position of the previously placed sutures in the first row as a guide. The diaphragmatic sutures are placed at the 4, 8, and 12 o'clock positions on a clock face, oriented with the 6 o'clock position placed posteriorly between the right and left crus just anterior to the aorta. The reconstructed cardia is gently pushed through the hiatus and placed in the abdomen. Once in the abdomen, the cardia should remain there without tension. The diaphragmatic and crural sutures are then tied, anchoring the wrap in the abdomen.

In patients with a short esophagus secondary to a stricture, Barrett's esophagus, or a large hiatal hernia, the esophagus is lengthened with a Collis gastroplasty (Fig. 23-53). The esophagus is lengthened by constructing a gastric tube along the lesser curvature. This allows a tension-free constriction of a Belsey Mark IV or Nissen fundoplication around the newly formed gastric tube, with placement of the repair in the abdomen. Because a short esophagus is commonly associated with a reduction in esophageal contraction amplitude and the gastric tube is inert, most surgeons prefer to combine the gastroplasty procedure with a 280° Belsey Mark IV fundoplication rather than a 360° Nissen fundoplication.

Outcome after Fundoplication

Nearly all published reports of laparoscopic fundoplication show that this procedure relieves the typical symptoms of gastroesophageal reflux—heartburn, regurgitation, and dysphagia—in greater than 90 percent of the patients. Overall, there is a 4.2 percent conversion rate to open surgery, and a 0.5 percent rate of early reoperation. Persistent postoperative dysphagia occurred in approximately 9 percent of patients in the early series, a rate 2 to 3 times higher than what is accepted for open fundoplication. The incidence of dysphagia has decreased to the 3 to 5 percent range with increasing experience and attention to the technical details in constructing the fundoplication. Resting LES characteristics and esophageal acid exposure return to normal in nearly all patients (Fig. 23-54). Morbidity after laparoscopic fundoplication is similar to that after open fundoplication, averaging 10 to 15 percent. A pitfall unique to the laparoscopic approach is that 1 to 2 percent of patients develop a pneumothorax and surgical emphysema. This is related to excessive hiatal dissection and has decreased as surgeons' experience has increased. Unrecognized perforation of the esophagus or stomach is the most life-threatening complication. Perforations occur most often during hiatal and circumferential dissection of the esophagus, and their incidence is also related to the surgeon's experience. Intraoperative recognition and repair are the keys to preventing a life-threatening complication.

It is recommended that the surgical approach to patients with gastroesophageal reflux disease be selective, that is, that the specific antireflux procedure for any patient be based upon the patient's existing esophageal function. The benefit of a selective approach is shown in our experience with 85 consecutive patients with different spectra of the disease (Table 23-6). In approximately 75 to 80 percent of the patients a transabdominal Nissen fundoplication was the most suitable treatment. The remaining 20 to 25 percent were best treated by tailoring the antireflux procedure to their existing amplitude of esophageal contractility and esophageal length. Of interest, patients selected for a Belsey partial fundoplication because of poor motility but normal esophageal length benefited the least. This suggests that in these patients the motility disorder may be a primary abnormality rather than secondary to reflux-induced injury.

BARRETT'S ESOPHAGUS

The condition whereby the tubular esophagus is lined with columnar epithelium rather than squamous epithelium was first described by Norman Barrett in 1950. He incorrectly believed it to be congenital in origin. It is now realized that it is an acquired abnormality, occurs in 7 to 10 percent of patients with gastroesophageal reflux disease, and represents the end stage of the natural history of this disease. It is also distinctly different from the congenital condition in which islands of gastric fundic epithelium are found in the upper half of the esophagus.

The definition of Barrett's esophagus has evolved considerably over the past decade. Traditionally, Barrett's esophagus was identified by the presence of columnar mucosa extending at least 3 cm into the esophagus. It is now recognized that the specialized intestinal type epithelium found in the Barrett's mucosa is the only tissue predisposed to malignant degeneration. Consequently, the diagnosis of Barrett's esophagus is presently made given any length of endoscopically identifiable columnar mucosa that proves on biopsy to show intestinal metaplasia. While long segments of columnar mucosa without intestinal metaplasia do occur, they are uncommon and are probably congenital in origin.

The hallmark of intestinal metaplasia is the presence of intestinal goblet cells. There is a high prevalence of biopsy-demonstrated intestinal metaplasia at the cardia, on the gastric side of the squamocolumnar junction, in the absence of endoscopic evidence of a columnar-lined esophagus. Evidence is accumulating that these patches of what appears to be

Barrett's in the cardia have a similar malignant potential as the longer segments and are precursors for carcinoma of the cardia.

The long-term relief of symptoms remains the primary reason for performing antireflux surgery in patients with Barrett's esophagus. Healing of esophageal mucosal injury and the prevention of disease progression are important secondary goals. In this regard, patients with Barrett's esophagus are no different than the broader population of patients with gastroesophageal reflux. They should be considered for antireflux surgery when patient data suggest severe disease or predict the need for long-term medical management. Most patients with Barrett's esophagus are symptomatic. Although it has been argued that some patients with Barrett's esophagus may not have symptoms, careful history taking will reveal the presence of symptoms in most, if not all, patients.

Patients with Barrett's esophagus have a spectrum of disease ranging from visually identifiable but short segments to long segments of classic Barrett's esophagus. In general, however, they represent a relatively severe stage of gastroesophageal reflux, usually with markedly increased esophageal acid exposure, deficient LES characteristics, poor esophageal body function, and a high prevalence of duodenogastroesophageal reflux. Gastric hypersecretion occurs in 44 percent of patients. Most will require long-term proton pump inhibitor therapy for relief of symptoms and control of coexistent esophageal mucosal injury. Given such profound deficits in esophageal physiology, antireflux surgery is an excellent means of long-term control for most patients with Barrett's esophagus. In years past, referral for antireflux surgery was reserved for patients with associated complications such as stricture, ulceration, or progression of the metaplastic segment. The advent of laparoscopic fundoplication and its successful control of gastroesophageal reflux in over 90 percent of patients has lowered the threshold for referral. Patients with quiescent, uncomplicated Barrett's esophagus, particularly young patients, are now considered by many to be excellent candidates for antireflux surgery.

The typical complications in Barrett's esophagus include ulceration in the columnar-lined segment, stricture formation, and a dysplasia-cancer sequence. Barrett's ulceration is unlike the erosive ulceration of reflux esophagitis in that it more closely resembles peptic ulceration in the stomach or duodenum and has the same propensity to bleed, penetrate, or perforate. The strictures found in Barrett's esophagus occur at the squamocolumnar junction and are typically higher than peptic strictures in the absence of Barrett's esophagus. Ulceration and stricture in association with Barrett's esophagus were commonly reported prior to 1975, but with the advent of potent acid suppression medication they have become less common. In contrast, the complication of adenocarcinoma developing in Barrett's mucosa has become more common. Adenocarcinoma developing in Barrett's mucosa was considered a rare tumor prior to 1975. Today it occurs in approximately 1 in every 100 patient years of follow-up, which represents a risk 40 times that of the general population. Most, if not all, cases of adenocarcinoma of the esophagus arise in Barrett's epithelium (Fig. 23-55). About one-third of all patients with Barrett's esophagus present with malignancy.

The long-term risk of progression to dysplasia and adenocarcinoma, while not the driving force behind the decision to perform antireflux surgery, is a significant concern for both patient and physician. Although to date there have been no prospective randomized studies documenting that antireflux surgery has an effect on the risk of progression to dysplasia and carcinoma, complete control of reflux of gastric juice into the esophagus is clearly a desirable goal. As data accumulate regarding the relative impact of medical and surgical therapy on the natural history of Barrett's metaplasia, the risk of progression may play a larger role in therapeutic decisions.

Outcome of Antireflux Surgery in Patients with Barrett's Esophagus

Few studies have focused on the alleviation of symptoms after antireflux surgery in patients with Barrett's esophagus (Table 23-7). Those that are available document excellent to good results in 72 to 95 percent of patients at 5 years following surgery.

Several studies have compared medical and surgical therapy. Attwood et al., in a prospective but nonrandomized study, reported upon 45 patients undergoing either medical (26) or surgical (19) treatment of Barrett's esophagus. The groups were similar in age, length of Barrett's segment, the percentage of time during which pH was less than 4, and length of follow-up. Improvement of symptoms was dramatic after antireflux surgery. Symptoms of heartburn or dysphagia recurred in 88 percent of patients treated with medical therapy alone, and 21 percent after antireflux surgery. Complications, largely the development of an esophageal stricture, occurred in 38 percent of medically treated and 16 percent of surgically treated patients ($p < 0.05$) over the 3-year follow-up period. One patient in each group developed esophageal adenocarcinoma. They concluded that antireflux surgery was superior to acid suppression for both the control of symptoms and the prevention of complications in patients with Barrett's esophagus. Other nonrandomized comparisons of medical and surgical therapy have reported similar results.

Ortiz et al. have published a prospective randomized comparison of medical and surgical therapy in 59 patients with Barrett's esophagus. Twenty-seven patients were treated medically and thirty-two with antireflux surgery. Medical therapy consisted of either H2 blockade or proton pump inhibitors. All patients in the medical group were treated with omeprazole beginning in 1992. Symptomatic improvement occurred in the majority of patients in both groups, 85 percent of those receiving medical therapy and 89 percent of those who underwent antireflux surgery. There was a marked difference in the prevalence of persistent esophagitis and strictures between the medical and surgical groups. In the group receiving medical therapy 53 and 45 percent of patients had persistent esophagitis or stricture, compared to 5 and 15 percent of patients having antireflux surgery. The authors concluded that the "systematic" nonsurgical approach to Barrett's esophagus should be questioned. Taken together these studies document the ability of antireflux surgery to provide long-term symptomatic relief in patients with Barrett's esophagus.

Long-Term Results

Three relevant questions arise concerning the fate, over time, of the metaplastic tissue found in Barrett's esophagus: (1) Does antireflux surgery cause regression of Barrett's epithelium? (2) Does it prevent progression? (3) Can the development of Barrett's metaplasia be prevented by early antireflux surgery in patients with reflux disease?

Evidence suggests that neither medical nor surgical therapy reliably results in complete regression of Barrett's epithelium. Considerable enthusiasm followed the initial report of Brand and Pope in which Barrett's epithelium in 4 of 10 patients appeared to regress following antireflux surgery. Most subsequent studies have shown no consistent regression. In one of the most recent studies, Sagar et al. evaluated the effect of antireflux surgery on 56 patients with Barrett's esophagus. The majority of patients (31) had a partial fundoplication in which a higher incidence of persistent reflux was well documented. Of the 56 patients, 24 had partial or complete regression of the Barrett's segment at a median follow-up of 5.5 years. Complete regression occurred in only five patients. Nine patients showed progression of the length of Barrett's mucosa, and adenocarcinoma developed in one patient with persistent symptoms 9 years after surgery. All patients in which the Barrett's segment regressed were asymptomatic. The authors concluded that both regression and progression of columnar-lined esophagus occur, although the relief of symptoms and effective prevention of reflux favored regression.

Current data indicate that patients with Barrett's esophagus should remain in an endoscopic surveillance program following antireflux surgery since there is no reliable evidence to indicate that the Barrett's mucosa will regress. Biopsy specimens should be reviewed by a pathologist with expertise in the field. If low-grade dysplasia is confirmed, biopsies should be repeated after 12 weeks of high-dose acid suppression therapy. If high-grade dysplasia is evident on more than one biopsy specimen, esophageal resection is advisable because of the greater than 50 percent probability that an invasive cancer is already present. Early detection and resection have been shown to decrease the mortality rate from esophageal cancer in these patients.

Since Barrett's esophagus results from chronic uncontrolled gastroesophageal reflux, and esophageal adenocarcinoma is virtually always associated with intestinal metaplasia, there are strong theoretical grounds for halting the progression toward malignancy by permanently and effectively stopping reflux of gastric contents. Thus prevention of progression, not regression, becomes the central issue. Although some cancers have developed after antireflux surgery, the absence of preexistent dysplasia prior to surgery or the efficacy of the operative procedure in reducing 24-h esophageal acid exposure to normal has not been documented. If the dysplasia is reported as lower grade or indeterminant, then inflammatory change often confused with dysplasia should be suppressed by a course of acid suppression therapy in high doses for 2 to 3 weeks followed by rebiopsy of the Barrett's segment.

There is a growing body of evidence to attest to the ability of fundoplication to protect against dysplasia and invasive malignancy. Three studies suggest that an effective antireflux procedure can impact the natural history of Barrett's esophagus in this regard. McCallum and co-workers analyzed the longitudinal follow-up of patients with Barrett's esophagus in the registry of the American College of Gastroenterologists. All patients had nondysplastic quiescent Barrett's esophagus at initial endoscopy; 152 received medical treatment and 29 underwent antireflux surgery. Surveillance endoscopy was performed annually. Of patients in the medically treated group 19 percent developed dysplasia while on medical therapy while only 3.4 percent did after antireflux surgery. A single adenocarcinoma was identified in a patient undergoing medical treatment.

In the prospective randomized study of Ortiz et al., dysplasia developed in 6 of 27 patients (22 percent) while on medical treatment. It was low-grade in five patients and high-grade in one. A single patient developed dysplasia following antireflux surgery; 24-h pH monitoring showed that the fundoplication procedure in this patient had been ineffective. The small number of patients developing dysplasia prevented these data from being statistically significant, but the trend is clear and is similar to that identified by McCallum.

Data from the Mayo clinic strongly suggest that antireflux surgery impacts the development of adenocarcinoma in patients with Barrett's esophagus. The authors reviewed the outcome of 118 patients with Barrett's esophagus undergoing antireflux surgery between 1960 and 1990. Three cancers occurred over an 18.5 year follow-up period, all within the first 3 years after surgery (Fig. 23-56). The fact that the development of adenocarcinoma was clustered in the early years after antireflux surgery and not randomly dispersed throughout the follow-up period suggests that antireflux surgery altered the natural history of the disease. Hammeetman has shown that once dysplasia has developed, carcinoma ensues in an average of 3 years. The occurrence of all observed cancers in the first few years suggests that the point of no return in the dysplasia-cancer sequence had already occurred prior to the time of antireflux surgery.

A more fundamental issue, but one that is less often addressed, is that of preventing Barrett's esophagus. It is likely that an aggressive diagnostic approach coupled with early and complete elimination of gastroesophageal reflux in patients free of Barrett's esophagus would dramatically reduce its incidence. While it is theoretically possible to eliminate gastroesophageal reflux with high doses of proton pump inhibitors, the requirement to do so 24 h a day, 7 days a week, 365 days a year for the patient's lifetime makes it impractical. Rather, carefully performed antireflux surgery in carefully selected patients is likely to be a more expedient choice. The effect of reliable and complete control of gastroesophageal reflux on the natural history of Barrett's metaplasia once it has developed, and on its prevention in symptomatic patients free of complications, is one of the most important areas of study over the next decade.

Atypical Reflux Symptoms (Fig. 23-57)

Chronic respiratory symptoms, such as chronic cough, recurrent pneumonias, episodes of nocturnal choking, waking up with gastric contents in the mouth, or soilage of the bed pillow, may also indicate the need for surgical therapy. The chest radiogram in patients suffering from repetitive pulmonary aspiration secondary to gastroesophageal reflux often shows signs of pleural thickening, bronchiectasis, and chronic interstitial pulmonary fibrosis. If 24-h pH monitoring confirms the presence of increased esophageal acid exposure and manometry shows normal esophageal body motility, an antireflux procedure can be done with an expected good result. Usually these patients have, however, a nonspecific motor abnormality of the esophageal body which tends to propel the refluxed material toward the pharynx. In some of these patients, the motor abnormality will disappear after a surgical antireflux procedure. In others, the motor disorder will persist and contribute to postoperative aspiration of swallowed saliva and food. Consequently, the results of an antireflux procedure in patients with a motor disorder of the esophageal body are variable.

Chest pain may be an atypical symptom of gastroesophageal reflux and is often confused with coronary artery disease. Fifty percent of patients in whom a cardiac cause of the

chest pain has been excluded will have increased esophageal acid exposure as a cause of the episode of pain. An antireflux procedure provides relief of the chest pain with greater constancy than will occur with medical therapy.

Dysphagia, regurgitation, or chest pain on eating in a patient with normal endoscopy and esophageal function studies can be an indication for an antireflux procedure. These symptoms are usually related to the presence of a large paraesophageal hernia, intrathoracic stomach, or a small hiatal hernia with a narrow diaphragmatic hiatus. A Schatzki ring may be present with the latter. All these conditions are easily identified with an upper gastrointestinal radiographic barium examination done by a knowledgeable radiologist. These patients may have no heartburn, since the LES is usually normal and reflux of gastric acid into the esophagus does not occur. The surgical repair of the hernia usually includes an antireflux procedure because of the potential of destroying the competency of the cardia during the surgical dissection.

Reoperation for Failed Antireflux Repairs

Failure of an antireflux procedure occurs when the patient, after the repair, is unable to swallow normally, experiences upper abdominal discomfort during and after meals, or has recurrence or persistence of reflux symptoms. The assessment of these symptoms and the selection of patients who need further surgery are challenging problems. Functional assessment of patients who have recurrent, persistent, or emergent new symptoms following a primary antireflux repair is critical to identify the cause of failure. Analysis of patients requiring reoperation after a previous antireflux procedures shows that placement of the wrap around the stomach is the most frequent cause for failure after open procedures, while herniation of the repair into the chest is the most frequent cause of failure after a laparoscopic procedure. This is probably due, in both instances, to an unrecognized short esophagus. In the laparoscopic approach the elevated diaphragm allows the repair to be properly placed but the tension from the short esophagus pulls the repair above the diaphragm when the crura are relaxed. Partial or complete breakdown of the fundoplication and construction of a too-tight or too-long wrap of a fundoplication occurs with both open and closed procedures. The fact that 10 percent of these patients had an undiagnosed underlying esophageal motor disorder underlines the critical role of preoperative esophageal function tests before the initial procedure.

The preferred surgical approach to a patient who has had a previously failed antireflux procedure is through a left thoracotomy with a peripheral circumferential incision in the diaphragm to provide for simultaneous exposure of the upper abdomen and safe dissection of the previous repair from both abdominal and thoracic sides of the diaphragm. Patients who have recurrence of heartburn and regurgitation without dysphagia and have good esophageal motility are most amenable to reoperation and can be expected to have an excellent outcome. When dysphagia is the cause of failure, the situation is more difficult to manage. If the dysphagia occurred immediately following the repair, it is usually due to a technical failure, most commonly a misplaced fundoplication around the upper stomach, and rerepair is usually satisfactory. When dysphagia is associated with poor motility and multiple previous repairs, serious consideration should be given to esophageal resection and replacement. With each reoperation the esophagus is damaged further and the chances of preserving function become less. Also, blood supply is reduced and ischemic necrosis of the esophagus can occur after several previous mobilizations.

MOTILITY DISORDERS OF THE PHARYNX AND ESOPHAGUS

Clinical Manifestations

Dysphagia, i.e., difficulty in swallowing, is the primary symptom of esophageal motor disorders. Its perception by the patient is a balance between the severity of the underlying abnormality causing the dysphagia and the adjustment made by the patient in altering eating habits. Consequently, any complaint of dysphagia must include an assessment of the patient's dietary history. It must be known whether the patient experiences pain, chokes, or vomits with eating; whether the patient requires liquids with the meal, is the last to finish, or is forced to interrupt a social meal; and whether he or she has been admitted to the hospital for food impaction. These assessments, plus an evaluation of the patient's nutritional status, help to determine how severe the dysphagia is and evaluate the indications for surgical therapy.

A surgical myotomy is designed to improve the symptoms of dysphagia caused by a motility disorder. The results can profoundly improve the patient's ability to ingest food, but rarely return the function of the foregut to normal. The principle of the procedure is to destroy esophageal contractility in order to correct a defect in esophageal motility, resulting in improvement but never a return to normal function. To use a surgical myotomy to treat the problem of dysphagia, the surgeon needs to know the precise functional abnormality causing the symptom. This usually entails a complete esophageal motility evaluation. A clear understanding of the physiologic mechanism of swallowing and determination of the motility abnormality giving rise to the dysphagia are essential for determining if surgery is indicated and the extent of the myotomy to be performed. Endoscopy is necessary only to exclude the presence of tumor or inflammatory changes as the cause of dysphagia.

Motility Disorders of the Pharyngoesophageal Segment

Disorders of the pharyngoesophageal phase of swallowing result from a discoordination of the neuromuscular events involved in chewing, initiation of swallowing, and propulsion of the material from the oropharynx into the cervical esophagus. They can be categorized into one or a combination of the following abnormalities: (1) inadequate oropharyngeal bolus transport; (2)inability to pressurize the pharynx; (3) inability to elevate the larynx; (4) discoordination of pharyngeal contraction and cricopharyngeal relaxation; and (5) decreased compliance of the pharyngoesophageal segment secondary to muscle pathology. The latter results in incomplete anatomic relaxation of the cricopharyngeus and cervical esophagus.

Pharyngoesophageal swallowing disorders are usually congenital or due to acquired disease involving the central and peripheral nervous system. This includes cerebrovascular accidents, brainstem tumors, poliomyelitis, multiple sclerosis, Parkinson's disease, pseudobulbar palsy, peripheral neuropathy, and operative damage to the cranial nerves involved in swallowing. Muscular diseases such as radiation-induced myopathy, dermatomyositis, myotonic dystrophy, and myasthenia gravis are less common causes. Rarely, extrinsic compression by thyromegaly, cervical lymphadenopathy, or hyperostosis of the cervical spine can cause pharyngoesophageal dysphagia.

Diagnostic Assessment of the Cricopharyngeal Segment

Abnormalities of pharyngoesophageal swallowing are difficult to assess with standard manometric techniques because of the rapidity of the oropharyngeal phase of swallowing, the movement of the gullet, and the asymmetry of the cricopharyngeus. Video- or cineradiography is currently the most objective test to evaluate oropharyngeal bolus transport, pharyngeal compression, relaxation of the pharyngoesophageal segment, and the dynamics of airway protection during swallowing. It readily identifies a diverticulum (Fig. 23-58), stasis of the contrast medium in the valleculae, a cricopharyngeal bar, and/or narrowing of the pharyngoesophageal segment. These are anatomic manifestations of neuromuscular disease, and result from the loss of muscle compliance in portions of the pharynx and esophagus composed of skeletal muscle.

Careful analysis of video- or cineradiographic studies combined with manometry using specially designed catheters can identify the cause of a pharyngoesophageal dysfunction in most situations (Fig. 23-59). Motility studies may demonstrate inadequate pharyngeal pressurization, insufficient or lack of cricopharyngeal relaxation, marked discoordination of pharyngeal pressurization, cricopharyngeal relaxation and cervical esophageal contraction, or a hypopharyngeal bolus pressure suggesting decreased compliance of the skeletal portion of the cervical esophagus (Fig. 23-60).

In many patients with cricopharyngeal dysfunction, including those with Zenker's diverticulum, it has been difficult to consistently demonstrate a motility abnormality or discoordination of pharyngoesophageal events. The abnormality most apt to be present is a loss of compliance in the pharyngoesophageal segment manifested by an increased bolus pressure. Cook and colleagues have demonstrated an increased resistance to the movement of a bolus through what appears on manometry to be a completely relaxed cricopharyngeal sphincter. Using simultaneous manometry and videofluoroscopy they showed that in these patients the cricopharyngeus is only partially relaxed; that is, the sphincter is relaxed enough to allow a drop of its pressure to esophageal baseline on manometry, but insufficiently relaxed to allow unimpaired passage of the bolus into the esophagus (Fig. 23-61). This incomplete relaxation is due to a loss of compliance of the muscle in the pharyngoesophageal segment and may be associated with a cricopharyngeal bar or Zenker's diverticulum. This decreased compliance of the cricopharyngeal sphincter can be recognized on esophageal manometry by a "shoulder" on the pharyngeal pressure wave, the amplitude of which correlates directly with the degree of outflow obstruction (Fig. 23-62). Increasing the diameter of this noncompliant segment reduces the resistance imposed to the passage of a bolus. Consequently, patients with low pharyngeal pressure, that is, poor piston function of the pharynx, or patients with increased resistance of the pharyngocervical esophageal segment from loss of skeletal muscle compliance are improved by a pharyngocricocervical esophageal myotomy. This enlarges the pharyngoesophageal segment and reduces outflow resistance. Esophageal muscle biopsy specimens from patients with Zenker's diverticulum have shown histologic evidence of the restrictive myopathy in the pharyngoesophageal segment. These findings correlate well with the observation of a decreased compliance of the upper esophagus demonstrated by videoradiography and the findings on detailed manometric studies of the pharynx and cervical esophagus. They suggest that the diverticulum develops as a consequence of the outflow resistance to bolus transport through the noncompliant muscle of the pharyngoesophageal segment.

The requirements for a successful pharyngoesophageal myotomy are (1) adequate oropharyngeal bolus transport; (2) the presence of an intact swallowing reflux; (3) reasonable coordination of pharyngeal pressurization with cricopharyngeal relaxation; (4) a cricopharyngeal bar, Zenker's diverticulum, or a narrowed pharyngoesophageal segment on videosophagogram and/or the presence of excessive pharyngoesophageal shoulder pressure on motility study.

Zenker's Diverticulum

In the past, the most common recognized sign of pharyngoesophageal dysfunction was the presence of a Zenker's diverticulum, originally described by Ludlow in 1769. The eponym resulted from Zenker's classic clinicopathologic descriptions of 34 cases published in 1878. Pharyngoesophageal diverticula have been reported to occur in 0.1 percent of 20,000 routine barium examinations and classically occur in elderly white males. Zenker's diverticula tend to enlarge progressively with time due to the decreased compliance of the skeletal portion of the cervical esophagus that occurs with aging.

Presenting symptoms include dysphagia associated with the spontaneous regurgitation of undigested, bland material, often interrupting eating or drinking. The symptom of dysphagia is due initially to the loss of muscle compliance in the pharyngoesophageal segment, later augmented by the presence of an enlarging diverticulum. On occasion, the dysphagia can be severe enough to cause debilitation and significant weight loss. Chronic aspiration and repetitive respiratory infection are common associated complaints. The diagnosis, once suspected, is established by a barium swallow. Endoscopy is usually difficult in the presence of a cricopharyngeal diverticulum, and potentially dangerous, owing to obstruction of the true esophageal lumen by the diverticulum and the attendant risk of diverticular perforation.

Pharyngocricoesophageal Myotomy

The low morbidity and mortality associated with cricopharyngeal and upper esophageal myotomy have encouraged a liberal approach toward its use for almost any problem in the oropharyngeal phase of swallowing. This attitude has resulted in an overall success rate in the relief of symptoms of only 64 percent. When patients are selected using radiographic or motility markers of disease as outlined above, it is unusual for patients not to be benefited.

The myotomy can be performed under local or general anesthesia through an incision along the anterior border of the left sternocleidomastoid muscle. The pharynx and cervical esophagus are exposed by retracting the sternocleidomastoid muscle and carotid sheath laterally and the thyroid, trachea, and larynx medially (Fig. 23-63). When a pharyngoesophageal diverticulum is present, localization of the pharyngoesophageal segment is easy. The diverticulum is carefully freed from the overlying areolar tissue to expose its neck, just below the inferior pharyngeal constrictor and above the cricopharyngeus muscle. It can be difficult to identify the cricopharyngeus muscle in the absence of a

diverticulum. A benefit of local anesthesia is that the patient can swallow and demonstrate an area of persistent narrowing at the pharyngoesophageal junction. Further, before closing the incision, gelatin can be fed to the patient to ascertain whether the symptoms have been relieved and to inspect the opening of the previously narrowed pharyngoesophageal segment. Under general anesthesia, and in the absence of a diverticulum, the placement of a nasogastric tube to the level of the manometrically determined cricopharyngeal sphincter helps in localization of the structures. The myotomy is extended cephalad by dividing 1 to 2 cm of inferior constrictor muscle of the pharynx and caudad by dividing the cricopharyngeal muscle and the cervical esophagus for a length of 4 to 5 cm. The cervical wound is closed only when all oozing of blood has ceased since a hematoma after this procedure is common and is often associated with temporary dysphagia while the hematoma absorbs. Oral alimentation is started the day after surgery. The patient is usually discharged on the first or second postoperative day.

If a diverticulum is present and is large enough to persist after a myotomy, it may be sutured in the inverted position to the prevertebral fascia using a permanent suture, i.e., diverticulopexy (Fig. 23-64). If the diverticulum is excessively large so that it would be redundant if suspended, or its walls are thickened, a diverticulectomy should be performed.

Postoperative complications include fistula formation, abscess, hematoma, recurrent nerve paralysis, difficulties in phonation, and Horner's syndrome. The incidence of the first two can be reduced by performing a diverticulopexy. Recurrence of a Zenker's diverticulum occurs late, and is more common after diverticulectomy without myotomy, presumably due to persistence of the underlying loss of compliance of the cervical esophagus when a myotomy is not performed.

Postoperative motility studies have shown that the peak pharyngeal pressure generated on swallowing is not affected, the resting cricopharyngeal pressure is reduced but not eliminated and the cricopharyngeal sphincter length is shortened. Consequently, after myotomy there is protection against esophagopharyngeal regurgitation.

Motility Disorders of the Esophageal Body and LES

Disorders of the esophageal phase of swallowing result from abnormalities in the propulsive pump action of the esophageal body or the relaxation of the LES. These disorders result from either primary esophageal abnormalities or from generalized neural, muscular, or collagen vascular disease (Table 23-8). The use of standard esophageal manometry techniques has allowed specific primary esophageal motility disorders to be identified out of a pool of nonspecific motility abnormalities. These include achalasia, diffuse esophageal spasm, the so-called nutcracker esophagus, and the hypertensive LES. The manometric characteristics of these disorders are shown in Table 23-9.

The boundaries between the primary esophageal motor disorders are, however, vague, and intermediate types exist. This is because their diagnosis usually is based on the analysis of 10 wet swallows performed in a laboratory setting. The technique of ambulatory 24-h monitoring of esophageal motor activity allows the classification of esophageal motor disorders to be performed on the basis of more than 1000 contractions recorded during different physiologic states, that is, normal daily activity, eating, and sleeping. There are significant differences in the classification of esophageal motor disorders based on standard manometry and classification based on ambulatory monitoring (see Fig. 23-25). The degree of reclassification that occurs when analysis of esophageal motor function is done on the basis of ambulatory manometry indicates that the classic categories of esophageal motor disorders are inappropriate. These findings indicate that esophageal motility disorders should be looked at as a spectrum of abnormalities which reflects various stages of destruction of esophageal motor function.

Achalasia

The best-known and best-understood primary motility disorder of the esophagus is achalasia, with an incidence of 6 per 100,000 population per year. Although complete absence of peristalsis in the esophageal body has been proposed as the major abnormality, present evidence indicates achalasia is a primary disorder of the LES. This is based on 24-h outpatient esophageal motility monitoring, which shows that even in advanced disease up to 5 percent of contractions can be peristaltic. Simultaneous esophageal waves develop as a result of the increased resistance to esophageal emptying caused by the nonrelaxing LES. This is supported by experimental studies in which a Gore-Tex band placed loosely around the gastroesophageal junction in cats did not change sphincter pressures, but resulted in impaired relaxation of the LES and outflow resistance. This led to a marked increased frequency of simultaneous waveforms and a decrease in contraction amplitude. The changes were associated with radiographic dilation of the esophagus and were reversible after removal of the band. Observations in patients with pseudoachalasia due to tumor infiltration, a tight stricture in the distal esophagus, or an antireflux procedure that is too tight also provide evidence that dysfunction of the esophageal body can be caused by the increased outflow obstruction of a nonrelaxing LES. The observation that esophageal peristalsis can return in patients with classic achalasia following dilation or myotomy provides further support that achalasia is a primary disease of the LES.

The pathogenesis of achalasia is presumed to be a neurogenic degeneration, which is either idiopathic or due to infection. In experimental animals, the disease has been reproduced by destruction of the nucleus ambiguus and the dorsal motor nucleus of the vagus nerve. In patients with the disease, degenerative changes have been shown in the vagus nerve and in the ganglia in the Auerbach plexus of the esophagus itself. This degeneration results in hypertension of the LES, a failure of the sphincter to relax on deglutition, elevation of intraluminal esophageal pressure, esophageal dilatation, and a subsequent loss of progressive peristalsis in the body of the esophagus. The esophageal dilatation results from the combination of a nonrelaxing sphincter, which causes a functional holdup of ingested material in the esophagus, and elevation of intraluminal pressure from repetitive pharyngeal air swallowing (Fig. 23-65). With time, the functional disorder results in anatomic alterations seen on radiographic studies, such as a dilated esophagus with a tapering, beaklike narrowing of the distal end (Fig. 23-66). There is usually an air-fluid level in the esophagus from the retained food and saliva, the height of which reflects the degree of resistance imposed by the nonrelaxing sphincter. As the disease progresses, the esophagus becomes massively dilated and tortuous.

A subgroup of patients with otherwise typical features of classic achalasia have simultaneous contractions of their esophageal body which can be of high amplitude. This manometric pattern has been termed "vigorous achalasia," and chest pain episodes are a common finding in these patients. Differentiation of vigorous achalasia from diffuse esophageal spasm can be difficult. In both diseases videoradiographic examination can show a corkscrew deformity of the esophagus and diverticulum formation.

Diffuse and Segmental Esophageal Spasm

Diffuse esophageal spasm is characterized by substernal chest pain and/or dysphagia. Diffuse esophageal spasm differs from classic achalasia in that it is primarily a disease of the esophageal body, produces a lesser degree of dysphagia, causes more chest pain, and has less effect on the patient's general condition. True symptomatic diffuse esophageal spasm is a rare condition, occurring about five times less frequently than achalasia.

The causation and neuromuscular pathophysiology of diffuse esophageal spasm are unclear. The basic motor abnormality is rapid wave progression down the esophagus secondary to an abnormality in the latency gradient. Hypertrophy of the muscular layer of the esophageal wall and degeneration of the esophageal branches of the vagus nerve have been observed in this disease, although these are not constant findings. Manometric abnormalities in diffuse esophageal spasm may be present over the total length of the esophageal body, but usually are confined to the distal two-thirds. In *segmental esophageal spasm* the manometric abnormalities are confined to a short segment of the esophagus.

The classic manometric findings in these patients are characterized by the frequent occurrence of simultaneous wave forms and multi peaked esophageal contractions, which may be of abnormally high amplitude or long duration. Key to the diagnosis of diffuse esophageal spasm is that there remain some peristaltic waveforms in excess of those seen in achalasia. A criterion of 20 percent or more simultaneous waveforms out of 10 wet swallows has been used to diagnose diffuse esophageal spasm. This figure is, however, arbitrary and often debated. Discriminate analysis has identified a series of abnormalities on the ambulatory motility record of patients with classic diffuse esophageal spasm. A composite score based on these parameters of the ambulatory motility record has allowed diagnosis of the disease with a sensitivity of 90 percent and a specificity of 100 percent. When applied prospectively, this scoring system identified severely deteriorated esophageal motor function in symptomatic patients despite the absence of the classic motility abnormalities of diffuse spasm on standard manometry.

The LES in patients with diffuse esophageal spasm usually shows a normal resting pressure and relaxation on deglutition. A hypertensive sphincter with poor relaxation may also be present. In patients with advanced disease, the radiographic appearance of tertiary contractions appears helical and has been termed *corkscrew esophagus* or *pseudodiverticulosis* (Fig. 23-67). Patients with segmental or diffuse esophageal spasm can compartmentalize the esophagus and develop an epiphrenic or midesophageal diverticulum (Fig. 23-68).

Nutcracker Esophagus

The disorder termed "nutcracker" or "supersqueezer" esophagus was recognized in the late 1970s. Other terms used to describe this entity are "hypertensive peristalsis" or "high-amplitude peristaltic contractions." It is the most frequent of the primary esophageal motility disorders. By definition the so-called nutcracker esophagus is a manometric abnormality in patients with chest pain characterized by peristaltic esophageal contractions with peak amplitudes greater than 2 SDs above the normal values in individual laboratories. Contraction amplitudes in these patients can easily be above 400 mmHg. Ambulatory 24-h monitoring of esophageal motor function in patients diagnosed as having nutcracker esophagus has identified a subgroup of patients with a motor pattern characteristic of diffuse esophageal spasm. These patients usually complain of dysphagia in addition to chest pain and probably are misclassified on the basis of standard manometric findings. The identification of these patients is important, since esophageal myotomy is a therapeutic option for patients with dysphagia and diffuse esophageal spasm but is of questionable value in patients with chest pain secondary to nutcracker esophagus.

Hypertensive Lower Esophageal Sphincter

Hypertensive LES in patients with chest pain or dysphagia was first described as a separate entity by Code et al. This disorder is characterized by an elevated basal pressure of the LES with normal relaxation and normal propulsion in the esophageal body. About half of these patients, however, have associated motility disorders of the esophageal body, particularly hypertensive peristalsis and simultaneous waveforms. In the remainder, the disorder exists as an isolated abnormality. Dysphagia in these patients may be caused by a lack of compliance of the sphincter even in its relaxed state. Myotomy of the LES may be indicated in patients not responding to medical therapy or dilation.

Nonspecific Esophageal Motor Disorders

Many patients complaining of dysphagia or chest pain of noncardiac origin demonstrate a variety of wave patterns and contraction amplitudes on esophageal manometry that are clearly out of the normal range but do not meet the criteria of a primary esophageal motility disorder. Esophageal manometry in these patients frequently shows an increased number of multi peaked or repetitive contractions, contractions of prolonged duration, nontransmitted contractions, an interruption of a peristaltic wave at various levels of the esophagus, or contractions of low amplitude. These motility abnormalities have been termed nonspecific esophageal motility disorders. Their significance in the causation of chest pain or dysphagia is still unclear. Surgery plays no role in the treatment of these disorders unless there is an associated diverticulum.

A clear distinction between primary esophageal motility disorders and nonspecific esophageal motility disorders is often not possible. Patients diagnosed as having nonspecific esophageal motility abnormalities on repeated studies will occasionally show abnormalities consistent with nutcracker esophagus. Similarly, progression from a nonspecific esophageal motility disorder to classic diffuse esophageal spasm has been demonstrated. The finding of a nonspecific esophageal motility disorder, therefore, may represent only a manometric marker of an intermittent more severe esophageal motor abnormality. Combined ambulatory 24-h esophageal pH and motility monitoring has shown that an increased esophageal exposure to gastric juice is common in patients diagnosed as having a nonspecific esophageal motility disorder. In some situations, the motor abnormalities may be induced by the irritation of refluxed gastric juice; in other situations it may be a primary event unrelated to the presence of reflux.

Diverticula of the Esophageal Body

Radiographic abnormalities such as segmental spasm, corkscrewing, compartmentalization, and diverticulum are the anatomic results of disordered motility function. Of these, the most persistent and easiest to demonstrate is an esophageal diverticulum. Diverticula occur most commonly with nonspecific motility disorders but can occur with all of the primary motility disorders. In the latter situation, the motility disorder is usually diagnosed before the development of the diverticulum. When present, a diverticulum may temporarily alleviate the symptom of dysphagia by becoming a receptacle for ingested food, and substitute the symptoms of postprandial pain and the regurgitation of undigested food. If a motility abnormality of the esophageal body or LES cannot be identified, a traction or congenital cause for the diverticulum should be considered. Because development in radiology preceded development in motility monitoring, diverticula of the esophagus were considered historically to be a primary abnormality, the cause, rather than the consequence, of motility disorders. Consequently, earlier texts focused on them as specific entities based upon their location.

Epiphrenic diverticula arise from the terminal third of the thoracic esophagus and are usually found adjacent to the diaphragm. They have been associated with distal esophageal muscular hypertrophy, esophageal motility abnormalities, and increased luminal pressure. They are “pulsion” diverticula, and have been associated with diffuse spasm, achalasia, or nonspecific motor abnormalities in the body of the esophagus.

Whether the diverticulum should be surgically resected or suspended depends upon its size and proximity to the vertebral body. When diverticula are associated with esophageal motility disorders, esophageal myotomy from the distal extent of the diverticulum to the stomach is indicated; otherwise, one can expect a high incidence of suture line rupture due to the same intraluminal pressure that initially gave rise to the diverticulum. If the diverticulum is suspended to the prevertebral fascia of the thoracic vertebra, a myotomy is begun at the neck of the diverticulum and extended across the LES. If the diverticulum is excised by dividing the neck, the muscle is closed over the excision site and a myotomy is performed on the opposite esophageal wall, starting at the level of diverticulum. When a large diverticulum is associated with a hiatal hernia, the diverticulum is excised, a myotomy is performed if there is an associated esophageal motility abnormality, and the hernia is repaired because of the high incidence of postoperative reflux when it is omitted.

Midesophageal or traction diverticula were first described in the nineteenth century (Fig. 23-69). At that time they were frequently noted in patients who had mediastinal lymph node involvement with tuberculosis. It was theorized that adhesions form between the inflamed mediastinal nodes and the esophagus. By contraction, the adhesions exerted “traction” on the esophageal wall and led to a localized diverticulum (Fig. 23-70). This theory was based on the findings of early dissections, where adhesions between diverticula and lymph nodes were commonly found. It is now believed that some diverticula in the midesophagus may also be caused by motility abnormalities.

Most midesophageal diverticula are asymptomatic and incidentally discovered during investigation for nonesophageal complaints. In such patients, the radiologic abnormality may be ignored. Patients with symptoms of dysphagia, regurgitation, chest pain, or aspiration in whom a diverticulum is discovered should be thoroughly investigated for an esophageal motor abnormality and treated appropriately. Occasionally, a patient will present with a bronchoesophageal fistula manifested by a chronic cough on ingestion of meals. The diverticulum in such patients is most likely to have an inflammatory etiology.

The indication for surgical intervention is the degree of symptomatic disability. Usually midesophageal diverticula can be suspended due to their proximity to the spine. If motor abnormality is documented, a myotomy should be performed similar to that described for an epiphrenic diverticulum.

Operations Long Esophageal Myotomy for Motor Disorders of the Esophageal Body

A long esophageal myotomy is indicated for dysphagia caused by any motor disorder characterized by segmental or generalized simultaneous waveforms in a patient whose symptoms are not relieved by medical therapy. Such disorders include diffuse and segmental esophageal spasm, vigorous achalasia, and nonspecific motility disorders associated with a mid or epiphrenic esophageal diverticulum. However, the decision to operate must be made by a balanced evaluation of the patient's symptoms, diet, lifestyle adjustments, and nutritional status, with the most important factor being the possibility of improving the patient's swallowing disability. The symptom of chest pain alone is not an indication for a surgical procedure.

Twenty-four-hour ambulatory motility monitoring has greatly aided in the identification of patients with symptoms of dysphagia and chest pain who might benefit from a surgical myotomy. Ambulatory motility studies have shown that when the prevalence of “effective contractions,” that is, peristaltic waveforms consisting of contractions with an amplitude above 30 mmHg, drops below 50 percent during meals, the patient is likely to experience dysphagia (Fig. 23-71). This would suggest that relief from the symptom can be expected with an improvement of esophageal contraction amplitude or amelioration of nonperistaltic waveforms. Prokinetic agents may increase esophageal contraction amplitude but do not alter the prevalence of simultaneous waveforms. Patients in whom the efficacy of esophageal propulsion is severely compromised because of a high prevalence of simultaneous waveforms usually receive little benefit from medical therapy. In these patients, a surgical myotomy of the esophageal body can improve the patients' dysphagia provided the loss of contraction amplitude in the remaining peristaltic waveforms, caused by the myotomy, has less effect on swallowing function than the presence of the excessive simultaneous contractions. This situation is reached when the prevalence of effective waveforms during meals drops below 30 percent.

In patients selected for surgery, preoperative manometry is essential to determine the proximal extent of the esophageal myotomy. Most surgeons extend the myotomy distally across the LES to reduce outflow resistance. Consequently, some form of antireflux protection is needed to avoid gastroesophageal reflux if there has been extensive dissection of the cardia. In this situation, most authors prefer a partial, rather than a full, fundoplication in order not to add back-resistance that will further interfere with the ability of the myotomized esophagus to empty (Fig. 23-72). If the symptoms of reflux are present preoperatively, 24-h pH monitoring is required to confirm its presence.

The procedure may be performed either open or via thoracoscopy. The open technique is performed through a left thoracotomy in the sixth intercostal space (Fig. 23-73). An incision is made in the posterior mediastinal pleura over the esophagus, and the left lateral wall of the esophagus is exposed. The esophagus is not circumferentially dissected unless necessary. A 2-cm incision is made into the abdomen through the parietal peritoneum at the midportion of the left crus. A tongue of gastric fundus is pulled into the chest. This exposes the gastroesophageal junction and its associated fat pad. The latter is excised to give a clear view of the junction. A myotomy is performed through all muscle layers, extending distally over the stomach 1 to 2 cm below the gastroesophageal junction and proximally on the esophagus over the distance of the manometric abnormality. The muscle layer is dissected from the mucosa laterally for a distance of 1 cm. Care is taken to divide all minute muscle bands, particularly in the area of the junction. The gastric fundic tongue is sutured to the margins of the myotomy over a distance of 3 to 4 cm and replaced into the abdomen. This maintains separation of the muscle and acts as a partial fundoplication to prevent reflux.

If an epiphrenic diverticulum is present, it is excised by dividing the neck and closing the muscle. The myotomy is then performed on the opposite esophageal wall. If a midesophageal diverticulum is present, the myotomy is made so that it includes the muscle around the neck, and the diverticulum is inverted and suspended by attaching it to the paravertebral fascia of the thoracic vertebra.

The results of myotomy for motor disorders of the esophageal body have improved in parallel with the improved preoperative diagnosis afforded by manometry. Previous published series report between 40 and 92 percent improvement of symptoms, but interpretation is difficult due to the small number of patients involved and the varying criteria for diagnosis of the primary motor abnormality. When myotomy is accurately done, 93 percent of the patients have effective palliation of dysphagia after a mean follow-up of 5 years, and 89 percent would have the procedure again if it was necessary. Most patients gain or maintain rather than lose weight after the operation. Postoperative motility studies show that the myotomy reduces the amplitude of esophageal contractions to near zero and eliminates simultaneous and peristaltic waves. If the benefit of obliterating the simultaneous waves exceeds the adverse effect on bolus propulsion caused by the loss of peristaltic waveforms, the patient's dysphagia is likely to be improved by the procedure. If not, the patient is likely to continue to complain of dysphagia and to have little improvement as a result of the operation. Preoperative motility studies are thus crucial in deciding which patients are most likely to benefit from a long esophageal myotomy.

The thoracoscopic technique is complicated by the fact that it requires complete retraction of the lung anteriorly to expose the esophagus. Proper positioning of the patient is critical to achieving this exposure. A prone position is ideal, allowing the left lung to fall forward away from the esophagus. Because of the possibility of open thoracotomy, however, it is best to place the patient in the right lateral decubitus position with the left thorax up, and then roll the patient anteriorly 45° toward prone. A beanbag secured to the table is used to hold the patient. The table can be rotated a further 30 to 40° so that the patient ends up nearly prone. Should thoracotomy be necessary, the table can be rotated back to the horizontal position and a thoracotomy performed without difficulty. Prone positioning is the key element in providing exposure for long myotomy. Four thoracoscopic ports in the left chest are utilized. With suitable lung retraction the myotomy is performed through all esophageal muscle layers extending distally to the endoscopic gastroesophageal junction and proximally over the distance of the manometric abnormality.

Few reports exist concerning the minimally invasive technique for performing long esophageal myotomy. Cuschieri has reported a preliminary experience with a thoracoscopically performed long esophageal myotomy for the treatment of nutcracker esophagus. Three patients with symptoms of chest pain and high-amplitude esophageal contractions and peristaltic waves forms were operated upon. No major morbidity was encountered. Nasogastric tubes were removed the first postoperative day and oral feeding was started on the second postoperative day. Two patients were discharged on postoperative day 4 and one patient on day 5. All had symptomatic relief on short follow-up.

Myotomy of the Lower Esophageal Sphincter

Secondary to reflux disease, achalasia is the most common functional disorder of the esophagus to require surgical intervention. The goal of treatment is to relieve the functional outflow obstruction secondary to the loss of relaxation and compliance of the LES. This requires disrupting the LES muscle. When performed adequately, that is, reducing sphincter pressure below 10 mmHg, and done early in the disease, LES myotomy results in symptomatic improvement with the occasional return of esophageal peristalsis. Reduction in LES resistance can be accomplished intraluminally by hydrostatic balloon dilation, which ruptures the sphincter muscle, or by a surgical myotomy that cuts the sphincter. The difference between these two methods appears to be the greater likelihood of reducing sphincter pressure to less than 10 mmHg by surgical myotomy as compared with hydrostatic balloon dilation. However, patients whose sphincter pressure has been reduced by hydrostatic balloon dilation to less than 10 mmHg have an outcome similar to those after surgical myotomy (Fig. 23-74). In performing a surgical myotomy of the LES, there are four important principles: (1) minimal dissection of the cardia, (2) adequate distal myotomy to reduce outflow resistance, (3) prevention of postoperative reflux, and (4) preventing rehealing of the myotomy site. In the past, the drawback of a surgical myotomy was the need for an open procedure. With the advent of limited-access technology, the myotomy can now be performed without a thoracotomy, either thoracoscopically or laparoscopically. The benefit of a thoracoscopic approach is that a myotomy can be done with no more disruption of the cardia than occurs with a balloon rupture of the sphincter muscle. A laparoscopic approach requires more extensive dissection of the cardia and the addition of an antireflux procedure to prevent postmyotomy gastroesophageal reflux.

The therapeutic decisions regarding the treatment of patients with achalasia center around four issues. The first issue is, should newly diagnosed patients be treated with pneumatic dilation or a surgical myotomy? Long-term follow up studies have shown that pneumatic dilation achieves adequate relief of dysphagia and pharyngeal regurgitation in 50 to 60 percent of patients (Fig. 23-75). Close follow-up is required and, if dilation fails, myotomy is indicated. For those patients who have a dilated and tortuous esophagus or an associated hiatal hernia, balloon dilation is dangerous and surgery is the better option. Whether it is better to treat a newly diagnosed esophageal achalasia patient by forceful dilation or by operative cardiomyotomy remains undecided. The outcome of the one controlled randomized study (38 patients) comparing the two modes of therapy suggests that surgical myotomy as a primary treatment gives better long-term results. There are several large retrospective series which report the outcome obtained with the two modes of

treatment (Table [23- 10](#)). Despite objections regarding variations in surgical and dilation techniques and the number of physicians performing the procedures, these collective data would appear to support operative myotomy as the initial treatment of choice when performed by a surgeon of average skill and experience. This view is confirmed by the large series of 899 patients reported by the Mayo Clinic spanning a 27-year period, and by the series of Csendes et al. on 100 patients followed for 5 to 7 years after surgery. Although it has been reported that a myotomy after previous balloon dilation is more difficult, this has not been our experience unless the cardia has been ruptured in a sawtooth manner. In this situation, operative intervention either immediately or after healing has occurred can be difficult.

The second issue is, should a surgical myotomy be performed through the abdomen or the chest? Myotomy of the LES can be accomplished via either an abdominal or thoracic approach. Advantages of the thoracic approach are (1) the myotomy can be achieved with minimal dissection of the esophageal hiatus, thereby preserving normal antireflux mechanisms and avoiding postoperative gastroesophageal reflux, (2) it allows easy extension of the myotomy cephalad, which may be necessary to encompass the full extent of abnormal motor function in patients with vigorous achalasia, and (3) associated pathology such as esophageal diverticula can be readily managed. The abdominal approach has the advantages of (1) better access to the gastroesophageal junction and (2) the ease of then performing an antireflux procedure, if so desired.

The third issue—one that has been long debated—is, should an antireflux procedure be added to a surgical myotomy? Excellent results have been reported following meticulously performed myotomy without an antireflux component. Complicating the controversy is the virtual absence of studies including objective documentation of the presence or absence of pathologic reflux following myotomy. The results of published studies are mixed, although the majority support the need for antireflux protection, particularly if there is extensive dissection of the hiatus as occurs when a transabdominal myotomy is performed. Further support for an antireflux procedure is that the development of a reflux-induced stricture after an esophageal myotomy is a serious problem and usually necessitates esophagectomy for relief of symptoms. On the other hand, the complications of gastroesophageal reflux are paradoxically more common in patients who had a myotomy plus an antireflux procedure than those who had only a transthoracic myotomy. This indicates that the addition of an antireflux procedure does not protect against the complications of reflux. Consequently, there is little reason to accept the degree of dissection required for the performance of an antireflux procedure if less dissection is beneficial in maintaining the competency of the cardia; similarly, there is little reason to accept the resistance an antireflux procedure imposes to esophageal emptying when the elimination of this resistance is the purpose for performing a myotomy in the first place. If an antireflux procedure is used as an adjunct to esophageal myotomy, a complete 360° fundoplication should be avoided. Rather, a 270° Belsey fundoplication or a Dor hemifundoplication should be used to avoid long-term esophageal dysfunction secondary to the outflow obstruction afforded by the fundoplication itself (see Fig. [23-72](#)).

The fourth issue centers upon whether or not a cure of this disease is achievable. Long-term follow-up studies after surgical myotomy have shown that late deterioration in results occurs after this procedure regardless of whether an antireflux procedure is done, and also after balloon dilation even when the sphincter pressure is reduced to below 10mmHg. It may be that even though a myotomy or balloon rupture of the LES muscle reduces the outflow obstruction at the cardia, the underlying motor disorder in the body of the esophagus persists and deteriorates further with the passage of time, leading to increased impairment of esophageal emptying. The earlier an effective reduction in outflow resistance can be accomplished the better the outcome will be and the more likely some esophageal body function can be restored.

Open Esophageal Myotomy

A modified Heller myotomy is performed through a left thoracotomy in the sixth intercostal space along the upper border of the seventh rib. The esophagus, and a tongue of gastric fundus, are exposed as described for a long myotomy. A myotomy through all muscle layers is performed, extending distally over the stomach to 1 to 2 cm below the junction and proximally on the esophagus for 4 to 5 cm. The cardia is reconstructed by suturing the tongue of gastric fundus to the margins of the myotomy to prevent rehealing of the myotomy site and to provide reflux protection in the area of the divided sphincter. If an extensive dissection of the cardia has been done, a more formal Belsey repair is performed. The tongue of gastric fundus is allowed to retract into the abdomen. Postoperatively, nasogastric drainage is maintained for 6 days to prevent distention of the stomach during healing. An oral diet is resumed on the seventh day, after a barium swallow study shows unobstructed passage of the bolus into the stomach without extravasation.

In a randomized long-term follow-up by Csendes et al. of 81 patients treated for achalasia either by forceful dilation or by surgical myotomy, myotomy was associated with a significant increase in the diameter at the gastroesophageal junction and a decrease in the diameter at the middle third of the esophagus on follow-up radiographic studies. There was a greater reduction in sphincter pressure and improvement in the amplitude of esophageal contractions after myotomy. Thirteen percent of patients regained some peristalsis after dilation compared with twenty-eight percent after surgery. These findings were shown to persist over a 5-year follow-up period, at which time 95 percent of those treated with surgical myotomy were doing well. Of those who were treated with dilation, only 54 percent were doing well, while 16 percent required redilation and 22 percent eventually required surgical myotomy to obtain relief.

If simultaneous esophageal contractions are associated with the sphincter abnormality, the so-called vigorous achalasia, then the myotomy should extend over the distance of the abnormal motility as mapped by the preoperative motility study. Failure to do this will result in continuing dysphagia and a dissatisfied patient. The best objective evaluation of improvement in the patient following either balloon dilation or myotomy is a scintigraphic measurement of esophageal emptying time. A good therapeutic response improves esophageal emptying toward normal. Some degree of dysphagia may, however, persist despite improved esophageal emptying due to disturbances in esophageal body function. When an antireflux procedure is added to the myotomy it should be a partial fundoplication. A 360° fundoplication is associated with progressive retention of swallowed food, regurgitation, and aspiration to a degree that exceeds the patient's preoperative symptoms.

Endosurgical Esophageal Myotomy

Video-assisted endosurgical myotomy can be accomplished safely either laparoscopically or thoracoscopically. The thoracoscopic procedure is performed with the patient in the left lateral decubitus position. A double-lumen endotracheal tube is used to allow selective ventilation of the right lung. A seven-port technique is used: three are placed along the left costal margin for depressing the dome of the diaphragm and retracting the right and left crura, one is placed in the right axilla to retract the lung, one is placed behind the tip of the scapula for the camera, and two are operating ports. In contrast to laparoscopic surgery, the thoracoscopic ports are in close proximity to one another. With selective ventilation of the right lung, it is not necessary to insufflate the left hemithorax. Angled-viewing telescopes are preferable to the zero-degree scopes.

Identification and dissection of the esophagus are aided by the concomitant use of an endoscope within the esophageal lumen. The mediastinal pleura overlying the terminal esophagus is divided sharply with scissors. The myotomy is performed with a hook-type electrocautery probe (Fig. [23-76](#)). Having an endoscope within the esophagus helps prevent mucosal injury. Once the esophageal mucosa is clearly identified, the myotomy is carried distally with either the electrocautery probe or scissors. The inferior extent of the myotomy is identified by the presence of the junctional fat pad and the change from esophageal to gastric mucosa as observed externally and intraluminally with the endoscope. The myotomy is ended when the spasm of the valve commonly associated with achalasia is alleviated and a pull-through with an intraoperative motility probe shows loss of the gastroesophageal gradient.

The laparoscopic approach is similar to the Nissen fundoplication in terms of the trocar placement and exposure and dissection of the esophageal hiatus. Once the esophagus is mobilized, the myotomy is performed in a fashion analogous to the thoracoscopic approach. After the myotomy, the fundus of the stomach is brought posterior to the esophagus and sutured to both margins of the myotomy.

Outcome Assessment of the Therapy for Achalasia

Critical analysis of the results of therapy for motor disorders of the esophagus requires objective measurement. The use of symptoms alone as an endpoint to evaluate therapy for achalasia may be misleading. The propensity for patients to unconsciously modify their diet to avoid difficulty swallowing is underestimated, making an assessment of results based upon symptoms unreliable. Insufficient reduction in outflow resistance may allow progressive esophageal dilatation to develop slowly, giving the impression of improvement because the volume of food able to be ingested with comfort increases. A variety of objective measurements may be used to assess success, including LES pressure, esophageal baseline pressure, and scintigraphic assessment of esophageal emptying time. Esophageal baseline pressure is usually negative when compared to gastric pressure. Given that the goal of therapy is to eliminate the outflow resistance of a nonrelaxing sphincter, measurement of improvements in esophageal baseline pressure and scintigraphic transit time may be better indicators of success, but are rarely reported.

Eckardt et al. investigated whether the outcome of pneumatic dilation in patients with achalasia could be predicted on the basis of objective measurements. Postdilation LES pressure was the most valuable measurement for predicting long-term clinical response. A postdilatation sphincter pressure less than 10 mmHg predicted a good response. Fifty percent of the patients studied had postdilation sphincter pressures between 10 and 20 mmHg, with a 2-year remission rate of 71 percent. Importantly, 16 of 46 patients were left with a postdilation sphincter pressure of greater than 20 mmHg and had an unacceptable outcome. Overall, only 30 percent of patients dilated remain in symptomatic remission at 5 years.

Bonavina et al. reported good to excellent results with transabdominal myotomy and Dor fundoplication in 94 percent of patients after a mean follow-up of 5.4 years. No operative mortality occurred in either of these series, attesting to the safety of the procedure. Malthaner and Pearson reported the long-term clinical results in 35 patients with achalasia having a minimum follow up of 10 years (Table [23-11](#)). Twenty-two of these patients underwent primary esophageal myotomy and Belsey hemifundoplication at the Toronto General Hospital. Excellent to good results were noted in 95 percent of patients at 1 year, declining to 68, 69, and 67 percent at 10, 15, and 20 years, respectively. Two patients underwent early reoperation for an incomplete myotomy and three underwent an esophagectomy for progressive disease. They concluded that there was a deterioration of the initially good results after surgical myotomy and hiatal repair for achalasia, which is due to late complications of gastroesophageal reflux.

Ellis reported his lifetime experience with transthoracic short esophageal myotomy without an antireflux procedure. One hundred seventy-nine patients were analyzed at a mean follow-up of 9 years, ranging from 6 months to 20 years. Overall 89 percent of patients were improved at the 9-year mark. He also observed that the level of improvement deteriorated with time, with excellent results (patients continuing to be symptom-free) decreasing from 54 percent at 10 years to 32 percent at 20 years. He concluded that a short transthoracic myotomy without an antireflux procedure provides excellent long-term relief of dysphagia and, contrary to Malthaner and Pearson's experience, does not result in complications of gastroesophageal reflux. Both studies document nearly identical results 10 to 15 years following the procedure, and both report deterioration over time probably due to progression of the underlying disease. The addition of an antireflux procedure if the operation is performed transthoracically has no significant effect on the outcome.

Early experience with endosurgical esophageal myotomy is encouraging (Fig. [23-77](#)). Pellegrini reported on 17 patients, 15 of whom were treated thoracoscopically and 2 of whom were treated laparoscopically. Postoperatively, the mean LES pressure was 10 mmHG. The relief of dysphagia was graded as excellent in 12 patients, good in 2 patients, fair in 2 patients, and poor in one patient.

Esophageal Resection for End-Stage Motor Disorders of the Esophagus

Patients with dysphagia and long-standing benign disease whose esophageal function has been destroyed by the disease process or multiple previous surgical procedures are best managed by esophagectomy. Fibrosis of the esophagus and cardia can result in weak contractions and failure of the distal esophageal sphincter to relax. The loss of esophageal contractions can result in the stasis of food, esophageal dilatation, regurgitation, and aspiration. The presence of these abnormalities signals end-stage motor disease. In these situations esophageal replacement is usually required to establish normal alimentation. Before proceeding with esophageal resection for patients with end-stage benign disease,

the choice of the organ to substitute for the esophagus, that is,, stomach, jejunum, or colon, should be considered. The choice of replacement is affected by a number of factors, as described later in the section on techniques of esophageal reconstruction.

CARCINOMA OF THE ESOPHAGUS

Squamous carcinoma accounts for the majority of esophageal carcinomas. Its incidence is highly variable, ranging from approximately 20 per 100,000 in the United States and Britain to 160 per 100,000 in certain parts of South Africa and the Honan province of China, and even 540 per 100,000 in the Guriev district of Kazakhstan. The environmental factors responsible for these localized high-incidence areas have not been conclusively identified, though additives to local foodstuffs (nitroso compounds in pickled vegetables and smoked meats) and mineral deficiencies (zinc and molybdenum) have been suggested. In Western societies, smoking and alcohol consumption are strongly linked with squamous carcinoma. Other definite associations link squamous carcinoma with long-standing achalasia, lye strictures, tylosis (an autosomal dominant disorder characterized by hyperkeratosis of the palms and soles), and human papillomavirus.

Adenocarcinoma of the esophagus, once an unusual malignancy, is diagnosed with increasing frequency (Fig. [23-78](#)) and now accounts for over 50 percent of esophageal cancer in some Western countries. The gross appearance resembles that of squamous cell carcinoma. Microscopically, adenocarcinoma almost always originates in metaplastic Barrett's mucosa, and resembles gastric cancer. Rarely it arises in the submucosal glands and forms intramural growths which resemble the mucoepidermal and adenoid cystic carcinomas of the salivary glands. The most important etiologic factor in the development of primary adenocarcinoma of the esophagus is a metaplastic columnar-lined or Barrett's esophagus, which occurs as a complication in approximately 10 percent of patients with gastroesophageal reflux disease. The incidence of adenocarcinoma in a patient with Barrett's esophagus when studied prospectively is 1 in 100 patient years of follow-up; that is, for every 100 patients with Barrett's esophagus followed for 1 year, one will develop adenocarcinoma. Although this risk appears to be small, it is at least 30 to 40 times that expected for a similar population without Barrett's esophagus. This risk is similar to the risk for developing lung cancer in a person with a 20 pack per year history of smoking. Endoscopic surveillance for patients with Barrett's esophagus is recommended for two reasons: (1) at present there is no reliable evidence that medical therapy removes the risk of neoplastic transformation; and (2) malignancy in Barrett's esophagus is curable if detected at an early stage.

Clinical Manifestations

Esophageal cancer is a disease affecting patients of advancing age, with dysphagia and weight loss being by far the most common symptoms at the time of diagnosis. In a few patients, dysphagia does not occur and symptoms arise from invasion of the primary tumor into adjacent structures or from metastases. Extension of the primary tumor into the tracheobronchial tree can cause stridor, and if a tracheoesophageal fistula develops, coughing, choking, and aspiration pneumonia result. Rarely, severe bleeding from erosion into the aorta or pulmonary vessels occurs. Either vocal cord may be invaded, causing paralysis, but most commonly, paralysis is caused by invasion of the left recurrent laryngeal nerve by the primary tumor or lymph node metastasis. Systemic organ metastases are usually manifested by jaundice or bone pain. The situation is different in high-incidence areas where screening is practiced. In these communities, the most prominent early symptom is pain on swallowing rough or dry food.

Dysphagia usually presents late in the natural history of the disease because the lack of a serosal layer on the esophagus allows the smooth muscle to dilate with ease. As a result, the dysphagia becomes severe enough to for the patient to seek medical advice only when more than 60 percent of the esophageal circumference is infiltrated with cancer. Consequently, the disease is usually far advanced if symptoms herald its presence. Tracheoesophageal fistula may be present in some patients on their first visit to the hospital, and more than 40 percent will have evidence of distant metastases. With tumors of the cardia, anorexia and weight loss usually precede the onset of dysphagia. The physical signs of esophageal tumors are those associated with the presence of distant metastases.

Staging of Esophageal Carcinoma

At the initial encounter with a patient diagnosed as having carcinoma of the esophagus, a decision must be made as to whether he or she is a candidate for curative surgical therapy, palliative surgical therapy, or nonsurgical palliation. Making this decision is difficult, because evaluating the pretreatment disease stage of esophageal carcinoma is imprecise due to the difficulty of measuring the depth of tumor penetration of the esophageal wall and the inaccessibility of the organ's widespread lymphatic drainage.

The introduction of endoscopic ultrasound has made it possible to identify patients who are potentially curable prior to surgical therapy. Using an endoscope, the depth of the wall penetration by the tumor and the presence of five or more lymph node metastases can be determined with an 80 percent accuracy. A curative resection should be encouraged if endoscopic ultrasound indicates that the tumor has not penetrated the esophageal wall and/or fewer than five enlarged lymph nodes are imaged. Thoracoscopic and laparoscopic staging of esophageal cancer have been recommended; preliminary results indicate that correct staging of esophageal carcinomas, using these techniques, approaches 90 percent. If these results are confirmed and the cost is not prohibitive, then thoracoscopic and laparoscopic staging are likely to become valuable tools to determine the extent of disease prior to therapy. At the present time, despite the modern techniques of computed tomography, magnetic resonance imaging, endoscopic ultrasound, and laparoscopic and thoracoscopic technology, pretreatment staging still remains imprecise. This underscores the need for an intraoperative assessment of the potential for cure in each individual patient.

Experience with esophageal resections in patients with early disease has identified characteristics of esophageal cancer that are associated with improved survival. A number of studies suggest that only metastasis to lymph nodes and tumor penetration of the esophageal wall have a significant and independent influence on prognosis. The beneficial effects of the absence of one factor persists even when the other is present. Factors known to be important in the survival of patients with advanced disease, such as cell type, degree of cellular differentiation, or location of tumor in the esophagus, have no effect on survival of patients who have undergone resection for early disease. Studies also showed that patients having five or less lymph node metastases have a better outcome. Using these data, Skinner developed the wall penetration, lymph node, and distant organ metastases (WNM) system for staging.

The WNM system differed somewhat from the previous efforts to develop a satisfactory staging criteria for carcinoma of the esophagus. Most surgeons agreed that the 1983 TMN system left much to be desired. In the third edition of the manual for staging of cancer of the American Joint Committee on Cancer (1988), an effort was made to provide a finer discrimination between stages than had been contained in the previous edition, in 1983. Table [23-12](#) shows the definitions for the primary tumor, regional lymph nodes, and distant metastasis as listed in the 1988 manual. Recently, Ellis, in a study comparing different staging criteria, showed that the new staging criteria of the American Joint Committee provide no better discrimination of stages as they relate to survival than the earlier version had. The 5-year survival of stage IIA patients was similar to that of stage IIB patients, and the survival of stage IIB patients was similar to that of stage III patients. Similarly, there was no difference between the 5-year survival of patients with T1 and T2 disease, nor was there a survival difference between those with T3 and T4 disease. He did confirm the observation that the depth of wall penetration and extent of lymph node involvement were reliable independent predictors of survival.

Ellis proposed adoption of Skinner's WNM staging system with some modifications (Table [23-13](#)). In Ellis's proposal, tumors limited to above the muscularis mucosa would be equivalent to Skinner's W0 designation, T1 and T2 tumors would equate with the W1 classification, and T3 and T4 tumors to the W2 classification. These classifications are illustrated in Fig. [23-79](#). He further reported a clear distinction between the 5-year survival of patients with negative nodes and those with fewer than five nodes involved, and a highly significant difference between the latter group and those with five or more nodes involved. Table 23-13 shows the definitions for the primary tumor, regional lymph nodes, and distant metastasis for the Skinner WNM staging system as modified by Ellis.

Using an expanded base of 408 resected patients, Ellis compared the 1988 staging criteria with his modification of the Skinner WNM system and produced evidence that a modified WNM staging system was more useful from a prognostic standpoint (see Tables 23-12 and 23-13) than the 1988 criteria. Not only is the number of patients more evenly divided among the four stages in the modified Skinner system, but the comparison of the 5- year survival rates between stages is highly significant, with almost a 50 percent reduction in survival rates for each increasing stage. The major difference in the staging criteria of the proposed modification to the Skinner WNM system is the recognition that the number of nodes involved has a profound effect on prognosis. Ellis has been a proponent of a limited resection and lymph node dissection. The data used to validate the modified staging system represent what the outcome would be for simple tumor removal. Consequently, it serves as an excellent basis for comparison with the results for more extensive en bloc resections or preoperative chemotherapy programs.

Clinical Approach to Carcinoma of the Esophagus and Cardia

The selection of a curative versus a palliative operation for cancer of the esophagus is based on the location of the tumor, the patient's age and health, the extent of the disease, and intraoperative staging. Figure [23-80](#) shows an algorithm of the clinical decisions important in the selection for curative or palliative therapy.

Tumor Location

The selection of surgical therapy for patients with carcinoma of the esophagus depends not only on the anatomic stage of the disease and an assessment of the swallowing capacity of the patient, but also on the location of the primary tumor.

It is estimated that 8 percent of the primary malignant tumors of the esophagus occur in the cervical portion (Fig. [23-81](#)). They are almost always squamous cell lesions, with a rare adenocarcinoma arising from a congenital inlet patch of columnar lining. These tumors, particularly those in the postcricoid area, represent a separate pathologic entity for a number of reasons: (1) they are more common in females and appear to be a unique entity in this regard; and (2) the efferent lymphatics from the cervical esophagus drain completely differently from those of the thoracic esophagus. The latter drain directly into the paratracheal and deep cervical or internal jugular lymph nodes with minimal flow in a longitudinal direction. Except in advanced disease, it is unusual for intrathoracic lymph nodes to be involved.

Low cervical lesions that reach the level of the thoracic inlet are usually unresectable due to early invasion of the great vessels and trachea. The length of the esophagus below the cricopharyngeus is insufficient to allow intubation or construction of a proximal anastomosis for a bypass procedure. Consequently, palliation of these tumors is very difficult, and patients afflicted with disease at this site have a poor prognosis. Upper airway obstruction or the development of tracheoesophageal fistulas in such tumors may require surgical intervention for palliation. If possible, these tumors should be resected after a preoperative course of chemotherapy has reduced their size.

Tumors that arise within the middle or upper third of the thoracic esophagus lie too close to the trachea and aorta to allow an en bloc resection without removal of these vital structures. Consequently, in this location only tumors that have not penetrated through the esophageal wall and have not metastasized to the regional lymph nodes are potentially curable. The resection for a tumor at this level is done similarly whether for palliation or cure, and long-term survival is a chance phenomenon. This does not mean that when resecting such tumors, efforts to remove the adjacent lymph nodes should be abandoned. To do so may inadvertently leave unrecognized metastatic disease behind and compromise the patient's overall survival, because of recurrent local disease and compression of the trachea. It is recommended that a course of preoperative chemoradiotherapy should be given before resection to shrink the size of the tumor. It is recommended that the radiotherapy be limited to 3.5 Gy to allow for tissue healing.

Tumors of the lower esophagus and cardia are usually adenocarcinomas. Squamous cell carcinoma of the lower esophagus, however, does occur. Both types of tumor are amenable to en bloc resection. Unless preoperative and intraoperative staging clearly demonstrate an incurable lesion, an en bloc resection in continuity with a lymph node

dissection should be performed. Because of the propensity of gastrointestinal tumors to spread for long distances submucosally, long lengths of grossly normal gastrointestinal tract should be resected. The longitudinal lymph flow in the esophagus can result in skip areas, with small foci of tumor above the primary lesion, which underscores the importance of a wide resection of esophageal tumors. Wong has shown that local recurrence at the anastomosis can be prevented by obtaining a 10-cm margin of normal esophagus above the tumor. Anatomic studies have also shown that there is no submucosal lymphatic barrier between the esophagus and the stomach at the cardia, and Wong has shown that 50 percent of the local recurrences in patients with esophageal cancer who are resected for cure occur in the intrathoracic stomach along the line of the gastric resection. Considering that the length of the esophagus ranges from 17 to 25 cm, and the length of the lesser curve of the stomach is approximately 12 cm, a curative resection requires a cervical division of the esophagus and a greater than 50 percent proximal gastrectomy in most patients with carcinoma of the distal esophagus or cardia. This compromises the length of the stomach and esophagus remaining to reestablish gastrointestinal continuity and necessitates a colon interposition.

Age

An en bloc resection for cure of carcinoma of the esophagus in a patient older than 75 is rarely indicated, because of the additional operative risk and the shorter life expectancy. Regardless of how favorable the staging criteria, a palliative resection is performed in these patients. This approach provides relief of symptoms with less extensive surgical procedures, and cure is still a chance possibility.

Cardiopulmonary Reserve

Patients undergoing esophageal resection should have sufficient cardiopulmonary reserve to tolerate the proposed procedure. The respiratory function is best assessed with the FEV1, which ideally should be 2 L or more. Any patient with an FEV1 of less than 1.25 L is a poor candidate for surgery because he or she has a 40 percent risk of dying from respiratory insufficiency within 4 years. In such a patient, the chances of long-term survival, even if cured from the disease, do not justify an extensive en bloc resection. Clinical evaluation and electrocardiogram are not sufficient indicators of cardiac reserve. Echocardiography and dipyridamole thallium imaging provide accurate information on wall motion, ejection fraction, and myocardial blood flow. A defect on thallium imaging may require further evaluation with preoperative coronary angiography. A resting ejection fraction of less than 40 percent, particularly if there is no increase with exercise, is an ominous sign. We prefer to perform a palliative resection in such a patient, regardless of how favorable the other criteria are.

Clinical Stage

Clinical factors that indicate an advanced stage of carcinoma and exclude surgery with curative intent are recurrent nerve paralysis, Horner's syndrome, persistent spinal pain, paralysis of the diaphragm, fistula formation, and malignant pleural effusion. Factors that make surgical cure unlikely include a tumor greater than 8 cm in length, abnormal axis of the esophagus on a barium radiogram, enlarged lymph nodes on CT, a weight loss more than 20 percent, and loss of appetite. In patients where these findings are not present, staging depends primarily on the length of the tumor as measured with endoscopy and the degree of wall penetration and lymph node metastasis seen with endoscopic ultrasound. Studies indicate that there are a high number of favorable parameters associated with tumors less than 4 cm in length; there are fewer with tumors between 4 and 8 cm, and favorable criteria for tumors greater than 8 cm in length are absent. Consequently, the finding of a tumor over 8 cm in length should exclude curative resection; the finding of a smaller tumor should encourage an aggressive approach, and the smaller the tumor the more aggressive the approach should be. Endoscopic ultrasound imaging of esophageal tumors has recently become available, and provides further information regarding the size, wall penetration, and lymph node status of the lesion (Fig. [23-82](#)).

Intraoperative Staging

Intraoperative staging allows selection of favorable candidates for a curative en bloc resection. It is based on the fact that there is a low survival rate for patients with a tumor that penetrates through the esophageal wall, or has multiple or distant lymph node metastasis, and requires an approach that allows switching from an en bloc curative dissection to a palliative resection if during the course of an operation one of the following features is revealed: an unresectable primary tumor, cavitory spread of the tumor, distant organ metastasis, extension of the tumor through the mediastinal pleura, multiple gross lymph node metastases, or microscopic evidence of lymph node involvement at the margins of an en bloc resection, i.e., low paratracheal, portal triad, subpancreatic, or periaortic lymph nodes. Figure [23-83](#) is an algorithm for intraoperative decision making. For cancers of the distal esophagus and cardia, patients with a favorable stage of disease can be identified by a combination of preoperative and intraoperative assessment with an 86 percent accuracy. The overall 5-year survival of these selected patients after a curative en bloc resection is between 40 and 55 percent. If the tumor does not extend through the esophageal wall and there are less than five positive lymph nodes, the 5-year survival is 75 percent. These results support a clinical approach in which an en bloc resection of the esophagus and stomach is advocated only for patients most likely to benefit.

Management of Patients Excluded from Curative Resection

If the patient is considered incurable on preoperative or intraoperative evaluation, the severity of dysphagia or other incapacitating symptoms is assessed. Dysphagia of grade IV or higher (Table [23-14](#)) is an indication for a palliative resection. If the patient is physiologically fit, simple esophageal resection and reconstruction with a cervical esophagogastrostomy offer the best palliation. This procedure allows the patient to eat without dysphagia and prevents the local complications of perforation, hemorrhage, fistula formation, and incapacitating pain. Occasionally a patient will be cured by a palliative resection, but this should not be used as justification for a palliative resection in the absence of dysphagia. Malignant pleural effusion, obvious mediastinal spread, or distant organ metastases are usually contraindications for a palliative resection. In this setting, if dysphagia is not a problem, nothing more need be done.

If an obstructing tumor cannot be resected due to invasion of the trachea, aorta, or heart, or the patient's general condition precludes an operative procedure, relief of dysphagia by reestablishing the esophageal lumen is the focus of therapy. In this situation, the objective is to provide relief of dysphagia with the lowest mortality and the shortest hospital stay. A variety of techniques including bouginage, intubation, laser ablation, and electrical coagulation are available and can be used alone or in combination. Most centers prefer intubation of the esophagus.

Figure [23-84](#) shows the correlation between grade of dysphagia and a patient's score on the Linear Analogue Self Assessment test (LASA), which assesses the patient's physical and psychological well-being and symptom control, including dysphagia. There is a significant negative correlation between dysphagia grade and quality of life as measured by the LASA.

Surgical Treatment

A patient's nutritional status before surgery has a profound effect on the outcome of an esophageal resection. Low serum protein levels have a deleterious effect on the cardiovascular system, and a poor nutritional status affects the host resistance to infection and the rate of anastomotic and wound healing. A serum albumin level of less than 3.4 g/dL on admission indicates poor caloric intake and an increased risk of surgical complications, including anastomotic breakdown. A feeding jejunostomy tube provides the most reliable and safest method for nutritional support in patients who cannot consume an oral diet and have a functionally normal small bowel. In severely malnourished patients, the jejunostomy is performed as a separate procedure to allow for preoperative nutritional support. In these patients the abdomen is entered through a small supraumbilical midline incision. Otherwise, the jejunostomy tube is placed at the time of esophageal resection, and feeding is begun on the third postoperative day.

In an analysis of 43 patients undergoing curative en bloc resection for adenocarcinoma of the distal esophagus, lymph node metastasis were present in 76 percent. Tumor depth was a good indicator of nodal involvement (Fig. [23-85](#)). At the time of resection 89 percent of transmural tumors, 60 percent of intramural tumors, and 33 percent of intramucosal lesions had lymph node metastasis. This is in agreement with the findings of Japanese investigators who reported nodal involvement in 30 percent of 40 patients with intramucosal squamous cell carcinoma. This finding has implications in the management of patients with high-grade dysplasia, where the prevalence of unexpected adenocarcinoma is up to 50 percent.

In the same series, adenocarcinomas of the lower esophagus and cardia spread widely to regional nodes, most commonly to nodes along the lesser curvature, celiac axis, and parahiatal regions (Table [23-15](#)). Ten percent had subcarinal nodes and eighteen percent had involved nodes of the splenic hilum, along the splenic artery, or along the greater curvature of the stomach; subcarinal and splenic nodes would remain following transhiatal resection. The involved nodes along the greater curve of the stomach would, by necessity, be transposed into the chest if the stomach was used to reestablish gastrointestinal continuity. In addition, over 20 percent of patients had positive celiac axis nodes, an area not dissected by most during transhiatal esophagectomy.

Control of the local and regional disease forms the basis for the classical en bloc resection. In our experience, the site of nodal recurrence following en bloc resection was outside the limits of the resection in 80 percent of patients with recurrent disease (Fig. [23-86](#)). All patients with thoracic nodal recurrence had disease in their upper mediastinum or aortopulmonary window, suggesting that these recurrences arose from nodes lying along the recurrent laryngeal nerve chains, which are not routinely removed by the en bloc dissection. Performing a more extended mediastinal dissection combined with a radical neck dissection has been advocated by some authors in the treatment of esophageal squamous cell carcinoma. The increased morbidity and the possibility of permanent hoarseness associated with this approach have discouraged its widespread application. It has been reported that the finding of metastatic disease involving the celiac nodes precludes survival beyond 2 years. By contrast, we have reported prolonged survival in patients with celiac axis involvement.

Cervical and Upper Thoracic Esophageal Cancer

When considering resection, tumors of the esophagus are best divided into those above and below the carina. Czerny reported the first successful resection of a carcinoma of the cervical esophagus in 1877. It was hoped that the prognosis for patients with this disease might be better than for those with carcinoma of the thoracic esophagus; this has not proved to be the case. Early experience with resection of the cervical esophagus resulted in a high mortality rate, and reconstruction by neck flaps often required multiple operations. Because of these complexities and the generally disappointing results, radiotherapy frequently was elected. Immediate mortality decreased, but control of the tumor was not satisfactory.

The difference between the two forms of therapy is the manner in which the disease recurs. Tumors treated with radiation therapy initially tend to recur locally as well as systemically, and cause unmanageable local disease with eventual erosion into neck vessels and trachea, causing hemorrhage and dyspnea. Patients who undergo surgical therapy have few local recurrences of the tumor provided total excision was possible, but they succumb to metastatic disease. Colin has reported a local failure rate of 80 percent after definitive radiation therapy, and 20 percent of these patients required palliative surgery in order to control the disease locally. Improvements in the techniques of immediate esophageal reconstruction have reduced the complications of the surgical treatment of this disease and encouraged a more aggressive surgical approach. The data reported by Colin suggest that an initial aggressive surgical resection yields longer survival than radiation therapy (Fig. [23-87](#)). Positive surgical margins, tracheal invasion that cannot be removed, and vocal cord paralysis correlate with a significantly shorter survival following surgery. Palliation was better achieved in patients who underwent esophagectomy with immediate gastric pull-up than in those who underwent primary radiation therapy or chemotherapy.

Lesions that are not fixed to the spine, do not invade the vessels or trachea, and do not have fixed cervical lymph node metastases should be resected. If lymph node metastases are present or the tumor comes in close proximity to the cricopharyngeus muscle, a course of preoperative chemo- and radiotherapy should be given before surgical resection. This usually consist of 2 to 3 cycles of chemotherapy and no more than 3.5 Gy of radiation therapy. Neoadjuvant therapy is given in an attempt to salvage the larynx, since the larynx is

often invaded by microscopic tumors, and, in the past, a total laryngectomy in combination with esophagectomy was usually necessary. A simultaneous en bloc dissection of the superior mediastinum and cervical lymph nodes is done, sparing the jugular veins on both sides.

The thoracic esophagus is removed via a right posterolateral thoracotomy with a corresponding en bloc lymphadenectomy. The continuity of the gastrointestinal tract is reestablished by pulling the stomach up through the esophageal bed. If removing the larynx is necessary, a permanent tracheostomy stoma is constructed in the lower flap of the cervical incision. The division of the trachea in some patients may preclude the possibility of a permanent cervical standard tracheostomy, since the remaining tracheal stump distal to the tumor will not reach the suprasternal notch. Removal of the medial head of the clavicles and the manubrium down to the sternal angle of Louis provides excellent exposure and allows the construction of a mediastinal tracheostomy. A bipedicle skin flap over the pectoralis muscle can be advanced upward, or a single-pedicle musculocutaneous flap including the pectoralis muscle and its overlying skin can be rotated to cover the defect. A circular incision in the flap can be used as a port through which the tracheal remnant is brought out to the skin.

Tumors of the Thoracic Esophagus and Cardia

For tumors that arise below the carina, we prefer either an en bloc resection for cure or a transhiatal removal for palliation. A curative procedure is performed according to the principles of an en bloc resection in continuity with the regional lymph nodes (Fig. 23-88). It is attempted in a patient whose preresection physical condition and tumor characteristics have the potential for long-term survival. The en bloc resection is done through three incisions in the following order: right posterolateral thoracotomy, en bloc dissection of the distal esophagus, mobilization of the esophagus above the aortic arch, closure of the thoracotomy, repositioning of the patient in the recumbent position; upper midline abdominal incision, en bloc dissection of the stomach and associated lymph nodes; left neck incision and proximal division of the esophagus. The specimen is removed transhiatally and the stomach is divided at the angulus, preserving the antrum. Gastrointestinal continuity is reestablished with a left colon interposition. During the thoracic and abdominal dissection, intraoperative staging is done. If during the course of the operation an incurable situation is identified, the en bloc resection is abandoned and a palliative resection is performed in a manner similar to that described for tumors of the middle and upper thoracic esophagus. The hospital mortality for the group undergoing a curative en bloc resection is similar to those undergoing a palliative transhiatal resection. If preoperative staging has shown that the patient is a candidate for palliative resection, a transhiatal esophagectomy is performed (Fig. 23-89). A standard left thoracotomy with intrathoracic anastomosis for lower lesions or an Ivor Lewis combined approach for higher lesions is not advocated because of (1) the proven need to resect long lengths of the esophagus to eradicate submucosal spread, (2) the higher morbidity associated with a thoracic anastomotic leak, and (3) the high incidence of esophagitis secondary to reflux following an intrathoracic anastomosis.

En Bloc Esophagogastrectomy Versus Transhiatal Resection for Carcinoma of the Lower Esophagus and Cardia

Strategies for treatment of esophageal carcinoma limit the role of surgery to removing the primary tumor, with the hope that adjuvant therapy will increase cure rates by destroying systemic disease. This approach emphasizes the concept of biological determinism, i.e., the outcome of treatment in esophageal cancer is determined at the time of diagnosis and that surgical therapy aimed at removing more than the primary tumor is not helpful. Lymph node metastases are considered simply markers of systemic disease; the systematic removal of involved nodes is not considered beneficial. The belief that removal of the primary tumor by transhiatal esophagogastrectomy results in the same survival rates as a more extensive en bloc resection is based on the same kind of reasoning.

In the transhiatal procedure there is no specific attempt made to remove lymph node-bearing tissue in the posterior mediastinum. By contrast, the en bloc esophagectomy removes the tumor covered on all surfaces with a layer of normal tissue (see Fig. 23-88). A long length of foregut above and below the lesion is resected to incorporate submucosal spread of the tumor. Consistent with this is resection of the proximal two-thirds of the stomach in patients with a tumor in the lower third of the thoracic esophagus or cardia. Appropriate cervical mediastinal and abdominal lymph node dissections are included using an en bloc technique to remove potentially involved regional nodes. Arguments to support the more extensive esophagectomy, gastrectomy, and lymph node dissection are listed in Table 23-16.

Data collected by Hagen et al. showed that en bloc esophagogastrectomy resulted in significantly better 5-year survival than transhiatal esophagogastrectomy despite the fact that postoperative pathologic classification of the removed specimens from the patients who underwent transhiatal esophagogastrectomy showed that these patients tended to have been in an earlier disease stage (Fig. 23-90). Comparison of the actuarial survival following en bloc esophagogastrectomy and transhiatal esophagogastrectomy based on pathologic classification and stage of disease is shown in Figs. 23-91, 23-92, and 23-93. A clear survival advantage is observed in patients with early lesions following en bloc esophagogastrectomy, where the 5-year survival was 75 percent. This was less so for intermediate and late lesions. Of interest is the fact that the results after transhiatal esophagogastrectomy were similar for early and intermediate disease, 19.8 percent and 21.0 percent at 2 years, respectively, and not much better than the 8.6 percent for late disease.

These studies showed that for early cancers of the lower esophagus and cardia, en bloc esophagogastrectomy results in significantly better survival rates than transhiatal esophagogastrectomy. This finding cannot be explained by a bias in the stage of disease resected, a difference in operative mortality, or death from nontumor causes. Rather, it appears to be due to the type of operation performed. The results in patients with intermediate and advanced disease are less clear; a survival advantage was present in advanced disease but not in intermediate disease. We suspect that the explanation for this outcome is the imprecision in assigning a stage of esophageal cancer to patients or the evaluation of too few patients with intermediate disease. Ideally, the question as to which procedure is the best should be resolved by a prospective randomized study.

Extent of Resection to Cure Disease Confined to the Mucosa

The development of surveillance programs for the detection of early squamous cell carcinoma in endemic areas and for early adenocarcinoma in patients with Barrett's esophagus has given rise to controversy over how to manage tumors confined to the mucosa. Some authors have endoscopically resected squamous carcinomas after using endoscopic ultrasound to determine that the depth of the tumor was limited to the mucosa. Surprisingly, large areas of squamous mucosa can be resected without perforation or bleeding, leaving the smooth surface of the muscularis mucosa intact. Reepithelization of the large artificially induced ulcer is usually complete in 3 weeks. In order not to miss a squamous cancer that has invaded deeper than expected, it is important to examine the deep margins of the resected specimen carefully and to perform periodic endoscopic follow-up examinations with vital staining techniques. This technique is not appropriate for multiple and widespread or circumferential squamous lesions because of the risk of developing a stricture during the healing process. In this situation, those acquainted with endoscopic resection would advocate an esophagectomy.

Several studies have shown that intraepithelial carcinoma, i.e., carcinoma in situ or high-grade dysplasia, and intramucosal tumors, i.e., invasive cancer limited by the muscular mucosa, are quite different in their biological behavior from submucosal tumors regardless of their histologic characteristics, that is, regardless of whether they are squamous cell carcinoma or adenocarcinoma arising in Barrett's mucosa. Vessel invasion and lymph node metastasis do not occur in severe dysplasia, are uncommon in the intramucosal tumors, but are the rule in submucosal tumors. Consequently, the 5-year survival for intramucosal tumors is significantly better than for submucosal tumors. These findings indicate that both severe dysplasia and intramucosal cancers represent early malignant lesions of the esophagus. A critical issue to be resolved is whether an intramucosal tumor can be correctly discriminated from a submucosal tumor before surgery. The results of using endoscopic ultrasound for determining the depth of tumors confined to the esophageal wall are of questionable accuracy. The resolution of present-day endoscopic ultrasonographic systems is not sufficient to predictably differentiate the fine detail of tumor infiltration when it is limited to the esophageal wall. Currently there is no dependable way, before surgery, of determining whether a tumor extends beyond the muscularis mucosa.

Another complicating factor is that up to 30 percent of patients with intramucosal tumors have lymph node metastases, although the number of involved nodes per patient is usually five or less. Akiyama and others have reported that even though the number of involved nodes may be small, they can spread to distant nodal regions including cervical and abdominal nodes.

For these reasons—the difficulty of determining preoperatively the depth to which a tumor has penetrated into the esophageal wall, the knowledge that tumors which invade through the muscularis mucosa into the submucosa have a 60 percent or more incidence of lymph node metastasis—as well as the fact that the distance from the epithelium to the muscularis mucosa is only 1 to 2 mm less than to the submucosa, it seems prudent to perform an en bloc esophagectomy for the treatment of intramucosal tumors. This also holds true for patients with Barrett's esophagus and high-grade dysplasia, since 50 percent of these patients will harbor an unknown intramucosal tumor. In both situations, a mediastinal and abdominal lymph node dissection is performed as well, but the proximal stomach and spleen is not resected since the possibility of extensive submucosal spread and splenic node metastases is minimal. Gastrointestinal continuity is reestablished by pulling the stomach up into the neck and performing an esophagogastrostomy. This approach, however, is not universally accepted.

As the understanding of the pathology of esophageal cancer improves and experience with its resection increases, evidence is accumulating that the best chance for cure of patients with an intramural tumor in the distal esophagus or cardia is an en bloc esophagectomy and proximal gastrectomy with gastrointestinal continuity reestablished with a colon interposition. For patients with a tumor in the upper or cervical esophagus, the best chance for cure is an en bloc esophagectomy and a cervical lymph node dissection with gastrointestinal continuity reestablished with a gastric pull-up. Table 23-17 presents a summary of the extent of resection for tumors extending various depths into the esophageal wall.

Alternative Therapies

Radiation Therapy

Primary treatment with radiation therapy does not produce results comparable with those obtained with surgery. Currently, the use of radiotherapy is restricted to patients who are not candidates for surgery. Palliation of dysphagia is short-term, generally lasting only 2 to 3 months. Furthermore, the length and course of treatment are difficult to justify in patients with a limited life expectancy. Consequently there is a reluctance to treat patients with advanced disease.

Adjuvant Chemotherapy

The proposal to use adjuvant chemotherapy in the treatment of esophageal cancer began when it became evident that most patients develop postoperative systemic metastasis without local recurrence. This observation led to the hypothesis that undetected systemic micrometastasis had been present at the time of diagnosis and if effective systemic therapy was added to local regional therapy, survival should improve.

Recently this hypothesis has been supported by the observation of epithelial tumor cells in the bone marrow in 37 percent of patients with esophageal cancer who were resected for cure. These patients had a greater prevalence of relapse at 9 months after surgery compared to those patients without such cells. Such studies emphasize that hematogenous dissemination of viable malignant cells occurs early in the disease and that systemic chemotherapy may be helpful if the cells are sensitive to the agent. On the other hand, systemic chemotherapy may be a hindrance, because of its immunosuppressive properties, if the cells are resistant. Unfortunately, current technology is not able to test tumor cell sensitivity to chemotherapeutic drugs. This requires that the choice of drugs be made solely on the basis of their clinical effectiveness against grossly similar tumors.

The decision to use preoperative rather than postoperative chemotherapy was based on the ineffectiveness of chemotherapeutic agents when used after surgery and animal studies suggesting that agents given before surgery were more effective. The claim that patients who receive chemotherapy before resection are less likely to develop resistance to the drugs is unsupported by hard evidence. The claim that drug delivery is enhanced because blood flow is more robust before patients undergo surgical dissection is similarly

flawed, since if enough blood reaches the operative site to heal the wound or anastomosis, then the flow should be sufficient to deliver chemotherapeutic drugs. There are, however, data supporting the claim that preoperative chemotherapy in patients with esophageal carcinoma can, if effective, facilitate surgical resection by reducing the size of the tumor. This is particularly beneficial in the case of squamous cell tumors above the level of the carina. Reducing the size of the tumor may provide a safer margin between the tumor and the trachea and allow an anastomosis to a tumor-free cervical esophagus just below the cricopharyngeus. Involved margin at this level usually requires a laryngectomy to prevent subsequent local recurrence.

Preoperative Chemotherapy

Three randomized prospective studies with squamous cell carcinoma have shown no survival benefit with preoperative chemotherapy over surgery alone (Table [23-18](#)). Similar studies for adenocarcinoma have not been done. For squamous cell tumors a complete response to chemotherapy occurred only in 6 percent of patients.

With the exception of the potential to improve resectability of tumors located above the carina, the benefits cited by those in favor of preoperative chemotherapy are questionable. Preoperative chemotherapy alone can potentially down-stage the tumor, particularly squamous cell carcinoma. It can also potentially eliminate or delay the appearance of metastasis. There is no evidence, however, that it can prolong survival of patients with resectable carcinoma of the esophagus. Most failures are due to distant metastatic disease, underscoring the need for improved systemic therapy. Postoperative septic and respiratory complications are more common in patients receiving chemotherapy.

Preoperative Combination Chemo- and Radiotherapy

Preoperative chemoradiotherapy using cisplatin and *5-fluorouracil (5-FU)* in combination with radiotherapy has been reported by several investigators to be beneficial in both adenomatous and squamous cell carcinoma of the esophagus. There have been six randomized prospective studies: four with squamous cell carcinoma, one with both squamous and adenocarcinoma, and one with only adenocarcinoma (Table [23-19](#)). Only one showed any survival benefit with preoperative chemoradiotherapy over surgery alone. Most authors report substantial morbidity and mortality to the treatment. However, many have been encouraged by the observation that some patients, who had a complete response, had remained free of recurrence at 3 years.

Caution must be exercised in trying to isolate the effects of chemotherapy, because the addition of preoperative radiation therapy to chemotherapy elevates the complete response rate and inflates the benefit of chemotherapy. With chemoradiation the complete response rates for adenocarcinoma range from 17 to 24 percent (Table [23-20](#)). When radiation is removed, the complete response falls to 0 percent to 5 percent, which suggests that the effects of chemotherapy are negligible. If radiotherapy is the factor responsible for improved response rate, surgery alone could do the job as well, since numerous studies in the past have shown that the combination of surgery and radiation does not provide any survival advantages.

The real question is, should a patient with carcinoma of the esophagus go through three cycles of chemotherapy on the 5 percent chance that they may get a complete response in the primary tumor and in the face of the paucity of evidence that such a response controls systemic disease? Studies have shown that the rates of infection, anastomotic breakdown, incidence of adult respiratory distress syndrome, and long-term use of a respirator were greater in patients receiving adjuvant therapy as compared with surgery alone.

Current data support giving chemoradiotherapy as a mater of routine in a limited number of clinical settings including: (1) preoperatively to reduce tumor size in a young person with surgically incurable squamous cell carcinoma above the carina, and (2) chemotherapy as salvage therapy for patients who have not had previous chemotherapy and develop recurrent systemic disease after surgical resection (Fig. [23-94](#)). Adjuvant therapy in patients not in those categories should be limited to the setting of a controlled clinical trial. At present, the strongest predictors of outcome of patients with esophageal cancer are the anatomic extent of the tumor at diagnosis and the completeness of tumor removal by surgical resection. After incomplete resection of an esophageal cancer the 5-year survival rates are 0 to 5 percent. In contrast, after complete resection, independent of stage of disease, 5-year survival ranges from 15 to 40 percent according to selection criteria and stage distribution. The importance of early recognition and adequate surgical resection cannot be overemphasized.

SARCOMAS OF THE ESOPHAGUS

Sarcomas and carcinosarcomas are rare neoplasms, accounting for approximately 0.1 to 1.5 percent of all esophageal tumors. They present with the symptom of dysphagia, which does not differ from the dysphagia associated with the more common epithelial carcinoma. Tumors located within the cervical or high thoracic esophagus can cause symptoms of pulmonary aspiration secondary to esophageal obstruction. Large tumors originating at the level of the tracheal bifurcation can produce symptoms of airway obstruction and syncope by direct compression of the tracheobronchial tree and heart (Fig. [23-95](#)). The duration of dysphagia and age of the patients affected with these tumors are similar to those with carcinoma of the esophagus.

A barium swallow usually shows a large polypoid intraluminal esophageal mass, causing partial obstruction and dilatation of the esophagus proximal to the tumor. The smooth polypoid nature of the lesion, although not itself diagnostic, is distinctive enough to suggest the presence of a sarcoma rather than the more common ulcerating, stenosing carcinoma.

Esophagoscopy commonly shows an intraluminal necrotic mass. When biopsy is attempted, it is important to remove the necrotic tissue until bleeding is seen on the tumor's surface. When this is not done, the biopsy specimen will show only tissue necrosis. Even when viable tumor is obtained on biopsy, it has been our experience that it cannot be definitively identified as either carcinoma, sarcoma, or carcinosarcoma on the basis of the histology of the portion biopsied. Biopsy results cannot be totally relied on to identify the presence of sarcoma, and it is often the polypoid nature of the lesion which arouses suspicion that it may be something other than carcinoma.

Polypoid sarcomas of the esophagus, in contrast to infiltrating carcinomas, remain superficial to the muscularis propria and are less likely to metastasize to regional lymph nodes. In one series of 14 patients, local extension or tumor metastasis would have prevented a potentially curative resection only in 5. Thus the presence of a large polypoid tumor should not deter the surgeon from resecting the lesion.

Sarcomatous lesions of the esophagus can be divided into epidermoid carcinomas with spindle cell features, such as carcinosarcoma, and true sarcomas that arise from mesenchymal tissue, such as leiomyosarcoma, fibrosarcoma, and rhabdomyosarcoma. Based on current histologic criteria for diagnosis, fibrosarcoma and rhabdomyosarcoma of the esophagus are extremely rare lesions and may not in fact exist.

Surgical resection of polypoid sarcoma of the esophagus is the treatment of choice, since radiation therapy has little success and the tumors remain superficial, with local invasion or distant metastases occurring late in the course of the disease. As with carcinoma, the absence of both wall penetration and lymph node metastases is necessary for curative treatment, and surgical resection is consequently responsible for the majority of the reported 5-year survivals. Resection also provides an excellent means of palliating the patient's symptoms. The surgical technique for resection and the subsequent restoration of the gastrointestinal continuity is similar to that described for carcinoma.

Four of the eight patients with carcinoma survived for 5 years or longer. Even though this number is small, it suggests that resection produces better results in epithelial carcinoma with spindle cell features than in squamous cell carcinoma of the esophagus. Similarly, with leiomyosarcoma of the esophagus, the same scattered reports exist with little information on survival. Of seven patients with leiomyosarcoma, two died from their disease—one in 3 months and the other 4 years and 7 months after resection. The other five patients were reported to have survived more than 5 years.

It is difficult to evaluate the benefits of resection for leiomyoblastoma (Fig. [23-96](#)) of the esophagus, due to the small number of reported patients with tumors in this location. Most leiomyoblastomas occur in the stomach, and 38 percent of these patients succumb to the cancer in 3 years. Fifty-five percent of patients with extragastric leiomyoblastoma also die from the disease, within an average of 3 years. Consequently, leiomyoblastoma should be considered a malignant lesion and apt to behave like a leiomyosarcoma. The presence of nuclear hyperchromatism, increased mitotic figures (more than 1 per high-power field), tumor size larger than 10 cm, and clinical symptoms of longer than 6 months' duration are associated with a poor prognosis.

BENIGN TUMORS AND CYSTS

Benign tumors and cysts of the esophagus are relatively uncommon. From the perspectives of both the clinician and the pathologist, benign tumors may be divided into those that are within the muscular wall and those that are within the lumen of the esophagus.

Intramural lesions are either solid tumors or cysts, and the vast majority are leiomyomas. They are made up of varying portions of smooth muscle and fibrous tissue. Fibromas, myomas, fibromyomas, and lipomyomas are closely related and occur rarely. Other histologic types of solid intramural tumors have been described, such as lipomas, neurofibromas, hemangiomas, osteochondromas, granular cell myoblastomas, and glomus tumors, but they are medical curiosities.

Intraluminal lesions are polypoid or pedunculated growths which usually originate in the submucosa, develop mainly into the lumen, and are covered with normal stratified squamous epithelium. The majority of these tumors are composed of fibrous tissue of varying degrees of compactness with a rich vascular supply. Some are loose and myxoid, e.g., myxoma and myxofibroma; some are more collagenous, e.g., fibroma; and some contain adipose tissue, such as fibrolipoma. These different types of tumor are frequently collectively designated as fibrovascular polyps or simply as polyps. Pedunculated intraluminal tumors should be removed. If the lesion is not too large, endoscopic removal with a snare is feasible.

Leiomyoma

Leiomyomas constitute more than 50 percent of benign esophageal tumors. The average age at presentation is 38, which is in sharp contrast to that seen with esophageal carcinoma. Leiomyomas are twice as common in males. Because they originate in smooth muscle, 90 percent are located in the lower two-thirds of the esophagus. They are usually solitary, but multiple tumors have been found on occasion. They vary greatly in size and shape. Tumors as small as 1 cm in diameter or as large as 10 lb have been removed.

Typically, leiomyomas are oval. During their growth, they remain intramural, having the bulk of their mass protruding toward the outer wall of the esophagus. The overlying mucosa is freely movable and normal in appearance. Neither their size nor location correlates with the degree of symptoms. Dysphagia and pain are the most common complaints, the two symptoms occurring more frequently together than separately. Bleeding directly related to the tumor is rare, and when hematemesis or melena occurs in a patient with an esophageal leiomyoma, other causes should be investigated.

A barium swallow is the most useful method to demonstrate a leiomyoma of the esophagus (Fig. [23-97](#)). In profile the tumor appears as a smooth, semilunar, or crescent-shaped filling defect which moves with swallowing, is sharply demarcated, and is covered and surrounded by normal mucosa. Esophagoscopy should be performed to exclude the reported observation of a coexistence with carcinoma. The freely movable mass, which bulges into the lumen, should not be biopsied because of an increased chance of mucosal

perforation at the time of surgical enucleation.

Despite their slow growth and limited potential for malignant degeneration, leiomyomas should be removed unless there are specific contraindications. The majority can be removed by simple enucleation. If during removal the mucosa is inadvertently entered, the defect can be repaired primarily. After tumor removal, the outer esophageal wall should be reconstructed by closure of the muscle layer. The location of the lesion and the extent of surgery required will dictate the approach. Lesions of the proximal and middle esophagus require a right thoracotomy, whereas distal esophageal lesions require a left thoracotomy. Videothoracoscopic approaches have been reported. The mortality rate associated with enucleation is less than 2 percent, and success in relieving the dysphagia is near 100 percent. Large lesions or those involving the gastroesophageal junction may require esophageal resection.

Esophageal Cyst

Cysts may be congenital or acquired. Congenital cysts are lined wholly or partly by columnar ciliated epithelium of the respiratory type, by glandular epithelium of the gastric type, by squamous epithelium, or by transitional epithelium. In some, epithelial lining cells may be absent. Confusion over the embryologic origin of congenital cysts has led to a variety of names, such as enteric, bronchogenic, and mediastinal cysts. Acquired retention cysts also occur, probably as a result of obstruction of the excretory ducts of the esophageal glands.

Enteric and bronchogenic cysts are the most common and arise as a result of developmental abnormalities during the formation and differentiation of the lower respiratory tract, esophagus, and stomach from the foregut. During its embryologic development, the esophagus is lined successively with simple columnar, pseudostratified ciliated columnar, and finally stratified squamous epithelium. This sequence probably accounts for the fact that the lining epithelium may be any or a combination of these; the presence of cilia does not necessarily indicate a respiratory origin.

Cysts vary in size from small to very large and are usually located intramurally in the middle to lower third of the esophagus. Their symptoms are similar to those of a leiomyoma. The diagnosis similarly depends on radiographic and endoscopic findings. Surgical excision by enucleation is the preferred treatment. During removal, a fistulous tract connecting the cysts to the airways should be looked for, particularly in patients who have had repetitive bronchopulmonary infections.

ESOPHAGEAL PERFORATION

Perforation of the esophagus constitutes a true emergency. It most commonly occurs following diagnostic or therapeutic procedures. Spontaneous perforation, referred to as Boerhaave's syndrome, accounts for only 15 percent of cases of esophageal perforation, foreign bodies for 14 percent, and trauma for 10 percent. Pain is a striking and consistent symptom and strongly suggests that an esophageal rupture has occurred, particularly if located in the cervical area following instrumentation of the esophagus or substernally in a patient with a history of resisting vomiting. If subcutaneous emphysema is present, the diagnosis is almost certain.

Spontaneous rupture of the esophagus is associated with a high mortality rate because of the delay in recognition and treatment. Although there usually is a history of resisting vomiting, in a small number of patients the injury occurs silently, without any antecedent history. When the chest radiogram of a patient with an esophageal perforation shows air or an effusion in the pleural space, the condition is often misdiagnosed as a pneumothorax or pancreatitis. An elevated serum amylase caused by the extrusion of saliva through the perforation may fix the diagnosis of pancreatitis in the mind of an unwary physician. If the chest radiogram is normal, a mistaken diagnosis of myocardial infarction or dissecting aneurysm is often made.

Spontaneous rupture usually occurs into the left pleural cavity or just above the gastroesophageal junction. Fifty percent of patients have concomitant gastroesophageal reflux disease, suggesting that minimal resistance to the transmission of abdominal pressure into the thoracic esophagus is a factor in the pathophysiology of the lesion. During vomiting, high peaks of intragastric pressure can be recorded, frequently exceeding 200 mmHg, but since extragastric pressure remains almost equal to intragastric pressure, stretching of the gastric wall is minimal. The amount of pressure transmitted to the esophagus varies considerably depending on the position of the gastroesophageal junction. When it is in the abdomen and exposed to intraabdominal pressure, the pressure transmitted to the esophagus is much less than when it is exposed to the negative thoracic pressure. In the latter situation, the pressure in the lower esophagus will frequently equal intragastric pressure if the glottis remains closed. Cadaver studies have shown that when this pressure exceeds 150 mmHg, rupture of the esophagus is apt to occur. When a hiatal hernia is present and the sphincter remains exposed to abdominal pressure, the lesion produced is usually a Mallory-Weiss mucosal tear, and bleeding rather than perforation is the problem. This is due to the stretching of the supradiaphragmatic portion of the gastric wall. In this situation, the hernia sac represents an extension of the abdominal cavity, and the gastroesophageal junction remains exposed to abdominal pressure.

Diagnosis

Abnormalities on the chest radiogram can be variable and should not be depended upon to make the diagnosis. This is because the abnormalities are dependent on three factors: (1) the time interval between the perforation and the radiographic examination, (2) the site of perforation, and (3) the integrity of the mediastinal pleura. Mediastinal emphysema, a strong indicator of perforation, takes at least 1 h to be demonstrated and is present in only 40 percent of patients. Mediastinal widening secondary to edema may not occur for several hours. The site of perforation can also influence the radiographic findings. In cervical perforation, cervical emphysema is common and mediastinal emphysema rare; the converse is true for thoracic perforations. Frequently, air will be visible in the erector spinae muscles on a neck radiogram before it can be palpated or seen on a chest radiogram (Fig. 23-98). The integrity of the mediastinal pleura influences the radiographic abnormality in that rupture of the pleura results in a pneumothorax, a finding that is seen in 77 percent of patients. In two-thirds of patients the perforation is on the left side, in one-fifth it is on the right side, and in one-tenth it is bilateral. If pleural integrity is maintained, mediastinal emphysema (rather than a pneumothorax) appears rapidly. A pleural effusion secondary to inflammation of the mediastinum occurs late. In 9 percent of patients the chest radiogram is normal.

The diagnosis is confirmed with a contrast esophagogram, which will demonstrate extravasation in 90 percent of patients. The use of a water-soluble medium such as Gastrografin is preferred. Of concern is that there is a 10 percent false-negative rate. This may be due to obtaining the radiographic study with the patient in the upright position. When the patient is upright, the passage of water-soluble contrast material can be too rapid to demonstrate a small perforation. The studies should be done with the patient in the right lateral decubitus position (Fig. 23-99). In this position the contrast material fills the whole length of the esophagus, allowing the actual site of perforation and its interconnecting cavities to be visualized in almost all patients.

Management

The key to optimum management is early diagnosis. The most favorable outcome is obtained following primary closure of the perforation within 24 h, resulting in 80 to 90 percent survival. Figure 23-100 is an operative photograph taken through a left thoracotomy of an esophageal rupture following a pneumatic dilation for achalasia. The most common location for the injury is the left lateral wall of the esophagus just above the gastroesophageal junction. To get adequate exposure of the injury a dissection similar to that described for esophageal myotomy is performed. A flap of stomach is pulled up and the soiled fat pad at the gastroesophageal junction is removed. The edges of the injury are trimmed and closed using a modified Gambee stitch (Fig. 23-101A). The closure is reinforced by the use of a pleural patch or construction of a Nissen fundoplication (Fig. 23-101B).

Mortality associated with immediate closure varies between 8 and 20 percent. After 24 h survival decreases to less than 50 percent and is not influenced by the type of operative therapy, that is, drainage alone or drainage plus closure of the perforation. If the time delay prior to closing a perforation approaches 24 h and the tissues are inflamed, division of the cardia and resection of the diseased portion of the esophagus are recommended. The remainder of the esophagus is mobilized and as much normal esophagus as possible is saved and brought out as an end cervical esophagostomy. In some situations the retained esophagus may be so long that it loops down into the chest. The contaminated mediastinum is drained and a feeding jejunostomy tube is inserted. The recovery from sepsis is often immediate, dramatic, and reflected by a marked improvement in the patient's condition over a 24-h period. On recovery from the sepsis the patient is discharged and returns on a subsequent date for reconstruction with a substernal colon interposition. Failure to apply this aggressive therapy can result in a mortality rate in excess of 50 percent in patients in whom the diagnosis has been delayed.

Nonoperative management of esophageal perforation has been advocated in select situations. The choice of conservative therapy requires skillful judgment and necessitates careful radiographic examination of the esophagus. This course of management usually follows an injury occurring during dilation of esophageal strictures or pneumatic dilations of achalasia. Conservative management should not be used in patients who have free perforations into the pleural space. Cameron proposed three criteria for the nonoperative management of esophageal perforation: (1) the barium swallow must show the perforation to be contained within the mediastinum and drain well back into the esophagus (Fig. 23-102), (2) symptoms should be mild, and (3) there should be minimal evidence of clinical sepsis. If these conditions are met, it is reasonable to treat the patient with hyperalimentation, antibiotics, and cimetidine to decrease acid secretion and diminish pepsin activity. Oral intake is resumed in 7 to 14 days, dependent on subsequent radiographic examinations.

CAUSTIC INJURY

Accidental caustic lesions occur mainly in children and, in general, rather small quantities of caustics are taken. In adults or teenagers, the swallowing of caustic liquids is usually deliberate, during suicide attempts, and greater quantities are swallowed. Alkalies are more frequently swallowed accidentally than acids because strong acids cause an immediate burning pain in the mouth.

Pathology

The swallowing of caustic substances causes both an acute and a chronic injury. During the acute phase, care focuses on controlling the immediate tissue injury and the potential for perforation; during the chronic phase, the focus is on treatment of strictures and disturbances in pharyngeal swallowing. In the acute phase the degree and extent of the lesion are dependent on several factors: the nature of the caustic substance, its concentration, the quantity swallowed, and the time the substance is in contact with the tissues. Acids and alkalies affect tissue in a different manner. Alkalies dissolve tissue and, therefore, penetrate more deeply, while acids cause a coagulative necrosis which limits their penetration. Animal experiments have shown that there is a correlation between the depth of lesion and the concentration of sodium hydroxide (NaOH) solution. When a solution of 3.8% comes into contact with the esophagus for 10 s, it causes a necrosis of the mucosa and the submucosa but spares the muscular layer. A concentration of 22.5% penetrates the whole esophageal wall and into the periesophageal tissues. Cleansing products can contain up to 90% NaOH. The strength of esophageal contractions varies according to the level of the esophagus, being weakest at the striated muscle-smooth muscle interface. Consequently, clearance from this area may be somewhat slower, allowing caustic substances to remain in contact with the mucosa longer. This can explain why the esophagus is preferentially and more severely affected at this level than in the lower portions.

The lesions caused by lye injury occur in three phases. First is the acute necrotic phase, lasting 1 to 4 days after injury. During this period, coagulation of intracellular proteins results in cell necrosis and the living tissue surrounding the area of necrosis develops an intense inflammatory reaction. Second is the ulceration and granulation phase, starting 3 to 5 days after injury. During this period the superficial necrotic tissue sloughs, leaving an ulcerated, acutely inflamed base, and granulation tissue fills the defect left by the sloughed mucosa. This phase lasts 10 to 12 days, and it is during this period that the esophagus is the weakest. Third is the phase of cicatrization and scarring, which begins the third week following injury. During this period the previously formed connective tissue begins to contract, resulting in narrowing of the esophagus. Adhesions between granulating areas

occurs, resulting in pockets and bands. It is during this period that efforts must be made to reduce stricture formation.

Clinical Manifestations

The clinical picture of an esophageal burn is determined by the degree and extent of the lesion. In the initial phase, complaints consist of pain in the mouth and substernal region, hypersalivation, pain on swallowing, and dysphagia. The presence of fever is strongly correlated with the presence of an esophageal lesion. Bleeding can occur, and frequently the patient vomits. These initial complaints disappear during the quiescent period of ulceration and granulation. During the cicatrization and scarring phase, the complaint of dysphagia reappears and is due to fibrosis and retraction resulting in narrowing of the esophagus. Of the patients who develop strictures, 60 percent do so within 1 month, and 80 percent within 2 months. If dysphagia does not develop within 8 months, it is unlikely that a stricture will occur. Serious systemic reactions such as hypovolemia and acidosis resulting in renal damage can occur in cases where the burns have been caused by strong acids. Respiratory complications such as laryngospasm, laryngedema, and occasionally pulmonary edema can occur, especially when strong acids are aspirated.

Inspection of the oral cavity and pharynx can indicate that caustic substances were swallowed, but does not reveal that the esophagus has been burned. Conversely, esophageal burns can be present without apparent oral injuries. Because of this poor correlation, early esophagoscopy is advocated to establish the presence of an esophageal injury. To lessen the chance of perforation, the scope should not be introduced beyond the proximal esophageal lesion. The degree of injury can be graded according to the criteria listed in Table [23-21](#). Even if the esophagoscopy is normal, strictures may appear later. Radiographic examination is not a reliable means to identify the presence of early esophageal injury but is important in later follow-up to identify strictures. The most common locations of caustic injuries are shown in Table [23-22](#).

Treatment

Treatment of a caustic lesion of the esophagus is directed toward management of both the immediate and late consequences of the injury. The immediate treatment consists of limiting the burn by administering neutralizing agents. To be effective, this must be done within the first hour. Lye or other alkali can be neutralized with half-strength vinegar, lemon juice, or orange juice. Acid can be neutralized with milk, egg white, or antacids. Sodium bicarbonate is not used because it generates CO₂, which might increase the danger of perforation. Emetics are contraindicated, since vomiting renews the contact of the caustic substance with the esophagus and can contribute to perforation if too forceful. Hypovolemia is corrected and broad-spectrum antibiotics are administered to lessen the inflammatory reaction and prevent infectious complications. If necessary a feeding jejunostomy tube is inserted to provide nutrition. Oral feeding can be started when the dysphagia of the initial phase has regressed.

In the past, surgeons waited until the appearance of a stricture before starting treatment. Currently, dilations are started the first day after the injury, with the aim of preserving the esophageal lumen by removing the adhesions that occurred in the injured segments. This approach, however, is controversial in that dilations can traumatize the esophagus, causing bleeding and perforation, and there are data indicating that excessive dilations cause increased fibrosis secondary to the added trauma. The use of steroids to limit fibrosis has been shown to be effective in animals, but their effectiveness in human beings is debatable.

Extensive necrosis of the esophagus frequently leads to perforation and is best managed by resection. When there is extensive gastric involvement, the esophagus is nearly always necrotic or severely burned, and total gastrectomy and near-total esophagectomy are necessary. The presence of air in the esophageal wall is a sign of muscle necrosis and impending perforation and is a strong indication for esophagectomy.

Management of acute injury is summarized in the algorithm in Fig. [23-103](#). Some authors have advocated the use of an intraluminal esophageal stent (Fig. [23-104](#)) in patients who are operated upon and found not to have evidence of extensive esophagogastric necrosis. In these patients, a biopsy of the posterior gastric wall should be performed in order to exclude occult injury. If histologically there is a question of viability, a second-look operation should be done within 36 h. If a stent is inserted it should be kept in position for 21 days, and removed after a satisfactory barium esophagogram. Esophagoscopy should be done and if strictures are present, dilations initiated.

Once the acute phase has passed, attention is turned to the prevention and management of strictures. Both antegrade dilation with a Hurst or Maloney bougie and retrograde dilation with a Tucker bougie have been satisfactory. Occasionally, particularly with severe strictures, the patient is instructed to swallow a string, over which metal Sippy dilators are passed until an adequate lumen can be obtained for passage of a mercury bougie. In a series of 1079 patients, early dilations started during the acute phase gave excellent results in 78 percent, good results in 13 percent, and poor results in 2 percent. Fifty-five patients died during the treatment. In contrast, of 333 patients whose strictures were dilated when they became symptomatic, only 21 percent had excellent results, 46 percent good, and 6 percent poor, with 3 dying during the process. The length of time the surgeon should persist with dilation before consideration of esophageal resection is problematic. An adequate lumen should be reestablished within 6 months to 1 year, with progressively longer intervals between dilations. If during the course of treatment an adequate lumen cannot be established or maintained, that is, smaller bougies must be used, operative intervention should be considered. Surgical intervention is indicated when there is (1) complete stenosis in which all attempts from above and below have failed to establish a lumen, (2) marked irregularity and pocketing on barium swallow, (3) the development of a severe periesophageal reaction or mediastinitis with dilatation, (4) a fistula, (5) the inability to dilate or maintain the lumen above a 40 French bougie, or (6) a patient who is unwilling or unable to undergo prolonged periods of dilation.

The variety of abnormalities seen requires that creativity be used when considering esophageal reconstruction. Skin tube esophagoplasties are now used much less frequently than they were in the past, and are mainly of historical interest. Currently the stomach, jejunum, and colon are the organs used to replace the esophagus through either the posterior mediastinum or the retrosternal route. A retrosternal route is chosen when there has been a previous esophagectomy or there is extensive fibrosis in the posterior mediastinum. When all factors are considered, the order of preference for an esophageal substitute is (1) colon, (2) stomach, and (3) jejunum. Free jejunal grafts based on the superior thyroid artery have provided excellent results. Whatever method is selected, it must be emphasized that these procedures cannot be taken lightly; minor errors of judgment or technique may lead to serious or even fatal complications.

Critical in the planning of the operation is the selection of cervical esophagus, pyriform sinus, or posterior pharynx as the site for proximal anastomosis. The site of the upper anastomosis depends on the extent of the pharyngeal and cervical esophageal damage encountered. When the cervical esophagus is destroyed and a pyriform sinus remains open, the anastomosis can be made to the hypopharynx (Fig. [23-105](#)). When the pyriform sinuses are completely stenosed, a transglottic approach is used to perform an anastomosis to the posterior oropharyngeal wall (Fig. [23-106](#)). This allows excision of supraglottic strictures and elevation and anterior tilting of the larynx. In both of these situations, the patient must relearn to swallow. Recovery is long and difficult and may require several endoscopic dilations and, often, reoperations. Sleeve resections of short strictures are not successful because the extent of damage to the wall of the esophagus can be greater than realized, and almost invariably the anastomosis is carried out in a diseased area.

The management of a bypassed damaged esophagus after injury is problematic. If the esophagus is left in place, ulceration from gastroesophageal reflux or the development of carcinoma must be considered. The extensive dissection necessary to remove the esophagus, particularly in the presence of marked periesophagitis, is associated with significant morbidity. Leaving the esophagus in place preserves the function of the vagus nerves, and in turn the function of the stomach. On the other hand, leaving a damaged esophagus in place can result in multiple blind sacs and subsequent development of mediastinal abscesses years later. Most experienced surgeons recommend that the esophagus be removed unless the operative risk is unduly high.

DIAPHRAGMATIC HERNIAS

With the advent of clinical radiology, it became evident that a diaphragmatic hernia was a relatively common abnormality and was not always accompanied by symptoms. Three types of esophageal hiatal hernia were identified: (1) the sliding hernia, type I, characterized by an upward dislocation of the cardia in the posterior mediastinum (Fig. [23-107A](#)); (2) the rolling or paraesophageal hernia, type II, characterized by an upward dislocation of the gastric fundus alongside a normally positioned cardia (Fig. [23-107B](#)); and (3) the combined sliding-rolling or mixed hernia, type III, characterized by an upward dislocation of both the cardia and the gastric fundus (Fig. [23-107C](#)). The end stage of type I and type II hernias occurs when the whole stomach migrates up into the chest by rotating 180° around its longitudinal axis, with the cardia and pylorus as fixed points. In this situation the abnormality is usually referred to as an intrathoracic stomach (Fig. [23-107D](#)).

Incidence and Etiology

The true incidence of a hiatal hernia in the overall population is difficult to determine because of the absence of symptoms in a large number of patients who are subsequently shown to have a hernia. When radiographic examinations are done in response to gastrointestinal symptoms, the incidence of a sliding hiatal hernia is seven times higher than that of a paraesophageal hernia. The age distribution of patients with paraesophageal hernias is significantly different from that observed in sliding hiatal hernias. The median age of the former is 61; of the latter, 48. Paraesophageal hernias are more likely to occur in women by a ratio of 4:1.

Structural deterioration of the phrenoesophageal membrane over time may explain the higher incidence of hiatal hernias in the older age group. These changes involve thinning of the upper fascial layer of the phrenoesophageal membrane (i.e., the supradiaphragmatic continuation of the endothoracic fascia) and loss of elasticity in the lower fascial layer (i.e., the infradiaphragmatic continuation of the transversalis fascia). Consequently, the phrenoesophageal membrane yields to stretching in the cranial direction due to the persistent intraabdominal pressure and the tug of esophageal shortening on swallowing. The upper fascial layer is formed only by loose connective tissue and is of little importance. The lower fascial layer is thick, stronger, and more important. It divides into an upper and lower leaf about 1 cm before attaching intimately with the esophageal adventitia. Due to stretching in the cranial direction, the attachment of the lower leaf protrudes upward and can frequently be identified in the thoracic cavity (Fig. [23-108](#)).

These observations point to the conclusion that the development of a hiatal hernia is an age-related phenomenon secondary to repetitive upward stretching of the phrenoesophageal membrane. A paraesophageal hernia rather than a sliding hernia develops when there is a defect, perhaps congenital, in the esophageal hiatus anterior to the esophagus. The persistent posterior fixation of the cardia to the preaortic fascia and the median arcuate ligament is the only essential difference between a sliding and a paraesophageal hernia. When an anterior defect in the hiatus occurs in association with a loss of fixation of the cardia, a mixed, or type III, hernia develops.

Clinical Manifestations

The clinical presentation of a paraesophageal hiatal hernia differs from that of a sliding hernia. There is usually a higher prevalence of symptoms of dysphagia and postprandial fullness with paraesophageal hernias, but the typical symptoms of heartburn and regurgitation present in sliding hiatal hernias can also occur. Both are caused by gastroesophageal reflux secondary to an underlying mechanical deficiency of the cardia. The symptoms of dysphagia and postprandial fullness in patients with a paraesophageal hernia are explained by the compression of the adjacent esophagus by a distended cardia, or twisting of the gastroesophageal junction by the torsion of the stomach that occurs as it becomes progressively displaced in the chest.

About one-third of patients with a paraesophageal hernia complain of hematemesis, which is due to recurrent bleeding from ulceration of the gastric mucosa in the herniated portion of the stomach. Respiratory complications are frequently associated with a paraesophageal hernia and consist of dyspnea from mechanical compression and recurrent pneumonia from aspiration. With time the stomach migrates into the chest and can cause intermittent obstruction due to the rotation that has occurred. In contrast, many patients with paraesophageal hiatal hernia are asymptomatic or complain of very minor symptoms. However, the presence of a paraesophageal hernia is life-threatening in one-fifth of patients in

that the hernia can lead to sudden catastrophic events, such as excessive bleeding or volvulus with acute gastric obstruction or infarction. With mild dilatation of the stomach, the gastric blood supply can be markedly reduced, causing gastric ischemia, ulceration, perforation, and sepsis.

The symptoms of sliding hiatal hernias are usually due to functional abnormalities associated with gastroesophageal reflux and include heartburn, regurgitation, and dysphagia. These patients have a mechanically defective LES, giving rise to the reflux of gastric juice into the esophagus and the symptoms of heartburn and regurgitation. The symptom of dysphagia occurs from the presence of mucosal edema, Schatzki's ring, stricture, or the inability to organize peristaltic activity in the body of the esophagus as a consequence of the disease.

There are a group of patients with sliding hiatal hernias not associated with reflux disease who have dysphagia without any obvious endoscopic or manometric explanation. Video barium radiograms have shown that the cause of dysphagia in these patients is an obstruction of the swallowed bolus by diaphragmatic impingement on the herniated stomach. Manometrically, this is reflected by a double-humped high-pressure zone at the gastroesophageal junction (Fig. 23-109). The first pressure rise is due to diaphragmatic impingement on the herniated stomach and the second to the true distal esophageal sphincter. These patients usually have a mechanically competent sphincter, but the impingement of the diaphragm on the stomach can result in propelling the contents of the supradiaphragmatic portion of the stomach up into the esophagus and pharynx, resulting in complaints of pharyngeal regurgitation and aspiration. Consequently this abnormality is often confused with typical gastroesophageal reflux disease. Surgical reduction of the hernia results in relief of the dysphagia in 91 percent of patients.

Diagnosis

A radiogram of the chest with the patient in the upright position can diagnose a hiatal hernia if it shows an air-fluid level behind the cardiac shadow (Fig. 23-110). This is usually caused by a paraesophageal hernia or an intrathoracic stomach. The accuracy of the upper gastrointestinal barium study in detecting a paraesophageal hiatal hernia is greater than for a sliding hernia, since the latter can often spontaneously reduce. The paraesophageal hiatal hernia is a permanent herniation of the stomach into the thoracic cavity, so that a barium swallow provides the diagnosis in virtually every case. Attention should be focused on the position of the gastroesophageal junction, when seen, to differentiate it from a type II hernia (see Fig. 23-107 B, C).

Fiberoptic esophagoscopy is very useful in the diagnosis and classification of a hiatal hernia because the scope can be retroflexed. In this position, a sliding hiatal hernia can be identified by noting a gastric pouch lined with rugal folds extending above the impression caused by the crura of the diaphragm (Fig. 23-111), or measuring at least 2 cm between the crura, identified by having the patient sniff, and the squamous-columnar junction on withdrawal of the scope (Fig. 23-112). A paraesophageal hernia is identified on retroversion of the scope by noting a separate orifice adjacent to the gastroesophageal junction into which gastric rugal folds ascend (Fig. 23-113). A sliding-rolling or mixed hernia can be identified by noting a gastric pouch lined with rugal folds above the diaphragm, with the gastroesophageal junction entering about midway up the side of the pouch (Fig. 23-114).

Pathophysiology

It has been assumed for a long time that a sliding hiatal hernia is associated with an incompetent distal esophageal sphincter, whereas a paraesophageal hiatal hernia constitutes a pure anatomic entity and is not associated with an incompetent cardia. Accordingly, surgical therapy for patients with a sliding hernia has been directed toward restoration of the physiology of the cardia, but for patients with a paraesophageal hernia, treatment consisted of simply reducing the stomach into the abdominal cavity and closing the crura.

Physiologic testing with 24-h esophageal pH monitoring has shown increased esophageal exposure to acid gastric juice in 60 percent of the patients with a paraesophageal hiatal hernia compared with the observed 71 percent incidence in patients with a sliding hiatal hernia. It is now recognized that paraesophageal hiatal hernia can be associated with pathologic gastroesophageal reflux.

Physiologic studies have also shown that the competency of the cardia depends on an interrelationship between distal esophageal sphincter pressure, the length of the sphincter that is exposed to the positive-pressure environment of the abdomen, and the overall length of the sphincter. A deficiency in any one of these manometric characteristics of the sphincter is associated with incompetency of the cardia regardless of whether a hernia is present. Patients with a paraesophageal hernia who have incompetent cardias have been shown to have a distal esophageal sphincter with normal pressure but a shortened overall length and displacement outside the positive-pressure environment of the abdomen (Fig. 23-115). In a sliding hernia, even though the sphincter appears to be within the chest on a radiographic barium study, it can still be exposed to abdominal pressure because of the surrounding hernial sac, which functions as an extension of the abdominal cavity. A high insertion of the phrenoesophageal membrane into the esophagus gives adequate length of the distal esophageal sphincter exposed to abdominal pressure. A low insertion gives inadequate length (Fig. 23-116). The importance of the anatomic length of esophagus within the hernial sac has been emphasized by Bombeck, Dillard, and Nyhus in their careful postmortem dissections of the hiatus in 55 patients. Of these patients, 8 had a hiatal hernia. Five of the eight had no evidence of esophagitis and, therefore, a competent cardia, and in these 5 patients, the phrenoesophageal membrane inserted 2 to 5 cm (with a mean of 3.6 cm) above the gastroesophageal junction. The other three patients had evidence of esophagitis and therefore an incompetent cardia, and in these patients the membrane inserted 1 cm or less (with a mean of 0.5 cm) above the gastroesophageal junction. This difference was significant and underscores the importance of an adequate length of intraabdominal esophagus in maintaining competency of the cardia even in the presence of a hiatal hernia.

In contrast to a paraesophageal hernia, where the sphincter remains fixed in the abdomen, in a mixed (type III) hernia the sphincter moves extraperitoneally into the thorax through the widened hiatus along with a portion of the lesser curvature of the stomach and cardia and forms part of the wall of the hernial sac. Consequently, the LES lies outside the abdominal cavity and is unaffected by its environmental pressures. The loss of normal esophageal fixation that occurs in a type I (sliding) hernia or a type III (mixed) hernia results in the body of the esophagus being less able to carry out its propulsive function. This contributes to a greater exposure of the distal esophagus to refluxed gastric juice when components of an incompetent cardia are present. The causes for a mechanical incompetency of the cardia are similar regardless of the type of hernia and are identical with those in patients who have an incompetent cardia and no hiatal hernia.

Treatment

Because of the catastrophic life-threatening complications of bleeding, infarction, and perforation that are part of the natural history of a paraesophageal hiatal hernia in about 25 percent of patients, surgical repair is indicated, even in the elderly.

In the report of Skinner and Belsey, 6 of 21 patients with a paraesophageal hernia treated medically died from the complications of strangulation, perforation, or exsanguinating hemorrhage secondary to acute dilatation of the herniated intrathoracic stomach. These catastrophes occurred for the most part without warning. With this in mind, patients with a paraesophageal hernia are counseled to have elective repair of their hernia regardless of the severity of their symptoms or the size of the hernia. If surgery is delayed and repair is done on an emergency basis, there is a 19 percent operative mortality, compared with less than 1 percent for an elective repair.

Based on pathophysiologic studies on patients with a paraesophageal hiatal hernia, the repair of a paraesophageal hernia should include an antireflux procedure to correct the sphincter characteristics associated with a mechanically incompetent cardia. This is particularly important when the operation is performed on an emergency basis, without preoperative studies. If time permits, preoperative evaluation with 24-h esophageal pH monitoring and esophageal manometry allows the identification of patients with competent cardias. Such patients are candidates for a simple anatomic repair, provided it can be done without surgical dissection of the cardia. If dissection of the cardia is necessary, an antireflux procedure should be added to the repair. Operative repair of sliding hiatal hernias is driven by symptoms of or complications of gastroesophageal reflux disease, unless it is determined that impingement on the stomach by the diaphragm is the cause of the patient's symptoms.

MISCELLANEOUS LESIONS

Plummer-Vinson Syndrome

This uncommon clinical syndrome is characterized by dysphagia associated with atrophic oral mucosa, spoon-shaped fingers with brittle nails, and chronic anemia. It characteristically occurs in middle-aged edentulous women. Because iron-deficiency anemia is a common finding, another name for this condition is sideropenic dysphagia. The syndrome is more common in the Scandinavian countries than in the United States, and its presentation is variable. Not all patients exhibit the classic syndrome; some lack iron-deficiency anemia and others have the typical clinical stigmata but lack dysphagia or the presence of an esophageal web.

Clinical observation suggests that the esophageal web once thought to be a component of the syndrome in some patients may actually be a drug-induced lesion, caused by ingestion of ferrous sulfate, a drug commonly prescribed in cases of iron-deficiency anemia. Ferrous sulfate is known to cause esophageal injury, and a number of patients may have had a drug-induced esophageal injury develop at the site where the web is commonly observed. Not knowing the cause of the esophageal abnormality, early observers reported the web as part of the syndrome. Malignant lesions of the oral mucosa, hypopharynx, and esophagus have been noted to occur in up to 100 percent of patients when followed long-term.

Videoradiographic study as well as endoscopic findings have demonstrated a fibrous web just below the cricopharyngeus muscle as the cause of dysphagia in these patients. Treatment consists of dilation of the web and iron therapy to correct the nutritional deficiency.

Schatzki's Ring

Schatzki's ring is a thin submucosal circumferential ring in the lower esophagus at the squamocolumnar junction, often associated with a hiatal hernia. Its significance and pathogenesis are unclear (Fig. 23-117). The ring was first noted by Templeton, but Schatzki and Gary defined it as a distinct entity in 1953. Its prevalence varies from 0.2 to 14 percent in the general population, depending on the technique of diagnosis and the criteria used. Stiennon believed the ring to be a pleat of mucosa formed by infolding of redundant esophageal mucosa due to shortening of the esophagus. Others believe the ring to be congenital, and still others suggest it is an early stricture resulting from inflammation of the esophageal mucosa caused by chronic reflux.

Schatzki's ring is a distinct clinical entity having different symptoms, upper gastrointestinal function studies, and response to treatment when compared with patients with a hiatal hernia but without a ring. Twenty-four-hour esophageal pH monitoring has shown that patients with a Schatzki's ring have a lower incidence of reflux than hiatal hernia controls. They also have better LES function. This, together with the presence of a ring, could represent a protective role to prevent gastroesophageal reflux.

Symptoms associated with Schatzki's ring are episodes of short-lasting dysphagia during hurried ingestion of solid foods. Its treatment has varied from dilation alone to dilation with antireflux measures, antireflux procedure alone, incision, and even excision of the ring. Little is known about the natural progression of Schatzki's rings. Chen et al., using radiologic techniques, showed progressive stenosis of rings in 59 percent of patients, whereas Schatzki found that the rings decreased in diameter in 29 percent of patients and remained unchanged in the rest.

Symptoms in patients with a ring are caused more by the presence of the ring than by gastroesophageal reflux. Most patients with a ring but without proven reflux respond to one

dilation, while most patients with proven reflux require repeated dilations. In this regard, the majority of Schatzki's ring patients without proven reflux have a history of ingestion of drugs known to be damaging to the esophageal mucosa. Bonavina et al. have suggested drug-induced injury as the cause of stenosis in patients with a ring but without a history of reflux. Since rings also occur in patients with proven reflux, it is likely that gastroesophageal reflux also plays a part. This is supported by the fact that there is less drug ingestion in the history of these patients. Schatzki's ring is probably an acquired lesion that can lead to stenosis from chemical-induced injury by pill lodgment in the distal esophagus or from reflux-induced injury to the lower esophageal mucosa.

The best form of treatment of a symptomatic Schatzki's ring in patients who do not have reflux consists of esophageal dilation for relief of the obstructive symptoms. In patients with a ring who have proven reflux and a mechanically defective sphincter, an antireflux procedure is necessary to obtain relief and avoid repeated dilation.

Mallory-Weiss Syndrome

In 1929, Mallory and Weiss described four patients with acute upper gastrointestinal bleeding who were found at autopsy to have mucosal tears at the gastroesophageal junction. This syndrome, characterized by acute upper gastrointestinal bleeding following repeated vomiting, is considered to be the cause of up to 15 percent of all severe upper gastrointestinal bleeds. The mechanism is similar to spontaneous esophageal perforation: an acute increase in intraabdominal pressure against a closed glottis in a patient with a hiatal hernia.

Mallory-Weiss tears are characterized by arterial bleeding, which may be massive. Vomiting is not an obligatory factor, as there may be other causes of an acute increase in intraabdominal pressure, such as paroxysmal coughing, seizures, and retching. The diagnosis requires a high index of suspicion, particularly in the patient who develops upper gastrointestinal bleeding following prolonged vomiting or retching. Upper endoscopy confirms the suspicion by identifying one or more longitudinal fissures in the mucosa of the herniated stomach as the source of bleeding.

In the majority of patients the bleeding will stop spontaneously with nonoperative management. In addition to blood replacement, the stomach should be decompressed and antiemetics administered, as a distended stomach and continued vomiting aggravate further bleeding. A Sengstaken-Blakemore tube will not stop the bleeding, as the pressure in the balloon is not sufficient to overcome arterial pressure. Only occasionally will surgery be required to stop blood loss. The procedure consists of laparotomy and high gastrotomy with oversewing of the linear tear. Mortality is uncommon and recurrence is rare.

Scleroderma

Scleroderma is a systemic disease accompanied by esophageal abnormalities in approximately 80 percent of patients. In most, the disease follows a prolonged course. Renal involvement occurs in a small percentage of patients and signals a poor prognosis. The onset of the disease is usually in the third or fourth decade of life, occurring twice as frequently in women as in men.

Small vessel inflammation appears to be an initiating event, with subsequent perivascular deposition of normal collagen, which may lead to vascular compromise. In the gastrointestinal tract, the predominant feature is smooth muscle atrophy. Whether the atrophy in the esophageal musculature is a primary effect or occurs secondary to a neurogenic disorder is unknown. The results of pharmacologic and hormonal manipulation, with agents that act either indirectly via neural mechanisms or directly on the muscle, suggest that scleroderma is a primary neurogenic disorder. Methacholine, which acts directly on smooth muscle receptors, causes a similar increase in LES pressure in normal controls and in patients with scleroderma. Edrophonium, a cholinesterase inhibitor which enhances the effect of acetylcholine when given to patients with scleroderma, causes an increase in LES pressure that is less marked in these patients than in normal controls, suggesting a neurogenic rather than myogenic etiology. Muscle ischemia due to perivascular compression has been suggested as a possible mechanism for the motility abnormality in scleroderma. Others have observed that, in the early stage of the disease, the manometric abnormalities may be reversed by reserpine, an agent that depletes catecholamines from the adrenergic system. This suggests that in early scleroderma an adrenergic overactivity may be present which causes a parasympathetic inhibition, supporting a neurogenic mechanism for the disease. In advanced disease manifested by smooth muscle atrophy and collagen deposition, reserpine no longer produces this reversal. Consequently, from a clinical perspective, the patient can be described as having a poor esophageal pump and a poor valve.

The diagnosis of scleroderma can be made manometrically by the observation of normal peristalsis in the proximal striated esophagus, with absent peristalsis in the distal smooth muscle portion (Fig. [23-118](#)). The LES pressure is progressively weakened as the disease advances. Because many of the systemic sequelae of the disease may be nondiagnostic, the motility pattern is frequently used as a specific diagnostic indicator. Gastroesophageal reflux commonly occurs in patients with scleroderma, since they have both hypotensive sphincters and poor esophageal clearance. This combined defect can lead to severe esophagitis and stricture formation. The typical barium swallow shows a dilated, barium- filled esophagus, stomach, and duodenum, or a hiatal hernia with distal esophageal stricture and proximal dilatation (Fig. [23-119](#)).

Traditionally, esophageal symptoms have been treated with H2 blockers, antacids, elevation of the head of the bed, and multiple dilations for strictures, with generally unsatisfactory results. The degree of esophagitis is usually severe and leads to marked esophageal shortening. Consequently a Collis gastroplasty in combination with a Belsey antireflux repair is the usual procedure for the surgical management of this problem. Surgery reduces esophageal acid exposure but does not return it to normal because of the poor clearance function of the body of the esophagus. Only 50 percent of the patients have a good to excellent result. If the esophagitis is severe or there has been a previous failed antireflux procedure and the disease is associated with delayed gastric emptying, a gastric resection with Roux-en-Y esophagojejunostomy and a Hunt-Lawrence pouch has proved the best option.

Acquired Fistulas

The esophagus lies in immediate contact with the membranous portion of the trachea and left bronchus, predisposing to the formation of fistula to these structures. Most acquired esophageal fistulas are to the tracheobronchial tree and secondary to either esophageal or pulmonary malignancy. Traumatic fistulas and those associated with esophageal diverticula account for the remainder. Fistulas associated with traction diverticula are usually due to mediastinal inflammatory disease, and traumatic fistulas usually occur secondary to penetrating wounds, lye ingestion, or iatrogenic injury.

These fistulas are characterized by paroxysmal coughing following the ingestion of liquids and by recurrent or chronic pulmonary infections. The onset of cough immediately after swallowing suggests aspiration, whereas a brief delay (30 to 60 s) suggests a fistula.

Spontaneous closure is rare, owing to the presence of malignancy or a recurrent infectious process. Surgical treatment of benign fistulas consists of division of the fistulous tract, resection of irreversibly damaged lung tissue, and closure of the esophageal defect. To prevent recurrence, a pleural flap should be interposed. Treatment of malignant fistulas is difficult, particularly in the presence of prior irradiation. Generally, only palliative treatment is indicated. This can be best done by using a specially designed esophageal endoprosthesis that bridges and occludes the fistula, allowing the patient to eat. Rarely, esophageal diversion, coupled with placement of a feeding jejunostomy, can be used as a last resort.

TECHNIQUES OF ESOPHAGEAL RECONSTRUCTION

Options for esophageal substitution include gastric advancement, colonic interposition, and either jejunal free transfer or advancement into the chest. Rarely, combinations of these grafts will be the only possible option. The indications for esophageal resection and substitution include malignant and end-stage benign disease. The latter includes reflux- or drug-induced stricture formation that cannot be dilated without damage to the esophagus, a dilated and tortuous esophagus secondary to severe motility disorders, lye-induced strictures, and multiple previous antireflux procedures. The choice of esophageal substitution has significant impact upon the technical difficulty of the procedure and influences the long-term outcome.

Partial Esophageal Resection

Low-lying benign lesions, with preserved proximal esophageal function, are best treated with the interposition of a segment of proximal jejunum into the chest and primary anastomosis. A jejunal interposition can reach to the inferior border of the pulmonary hilum with ease, but the architecture of its blood supply rarely allows the use of the jejunum above this point. Because the anastomosis is within the chest, a thoracotomy is necessary.

The jejunum is a dynamic graft and contributes to bolus transport, whereas the stomach and colon function more as a conduit. The stomach is a poor choice in this circumstance because of the propensity for the reflux of gastric contents into the upper esophagus following an intrathoracic esophagogastrectomy. It is now well recognized that this occurs, and can lead to incapacitating symptoms and esophageal destruction in some patients. Short segments of colon, on the other hand, lack significant motility and have a propensity for the development of esophagitis above the anastomosis.

Replacement of the cervical portion of the esophagus while preserving the distal portion is occasionally indicated in cervical esophageal or head and neck malignancy and following the ingestion of lye. Free transfer of a portion of jejunum to the neck has become a viable option and is successful in the majority of cases. Revascularization is achieved via use of the internal mammary artery and the internal mammary or innominate vein. Removal of the sternoclavicular joint aids in performing the vascular and distal esophageal anastomosis (Fig. [23-120](#)).

Reconstruction after Total Esophagectomy

Neither the intrathoracic stomach nor the intrathoracic colon functions as well as the native esophagus after an esophagogastrectomy. The choice between these organs will be influenced by several factors, such as the adequacy of their blood supply and the length of resected esophagus that they are capable of bridging. If the stomach shows evidence of disease, or has been contracted or reduced by previous gastric surgery, the length available for esophageal replacement may not be adequate. The presence of diverticular disease, unrecognized carcinoma, or colitis prohibits the use of the colon. The blood supply of the colon is more affected than the blood supply of the stomach by vascular disease, which may prevent its use. Of the two, the colon provides the longest graft. The stomach can usually reach to the neck if the amount of lesser curve resected does not interfere with the blood supply to the fundus. Gastric interposition has the advantage that only one anastomosis is required. On the other hand, there is greater potential for aspiration of gastric juice or stricturing of the cervical anastomosis from chronic reflux when stomach is used for replacement.

Patients following an esophagogastrectomy may have discomfort during or shortly after eating. The most common symptom is a postprandial pressure sensation or a feeling of being stuffed, which probably results from the loss of the gastric reservoir. This symptom is less common when the colon is used as an esophageal substitute, probably because the distal third of the stomach is retained in the abdomen and the interposed colon provides an additional reservoir function.

King and Hölscher have reported a 40 percent and 50 percent incidence of dysphagia after reestablishing gastrointestinal continuity with the stomach following esophagogastrectomy. This incidence is similar to Orringer's results after using the stomach to replace the esophagus in patients with benign disease. More than half of the patients

experienced dysphagia postoperatively; two-thirds of this group required postoperative dilation and one-fourth have persistent dysphagia and require home dilation. By contrast, dysphagia is uncommon and the need for dilation is rare following a colonic interposition. Isolauri reported on 248 patients with colonic interpositions and noted a 24 percent incidence of dysphagia 12 months after the operation. When it occurred, the most common cause was recurrent mediastinal tumor. The high incidence of dysphagia with the use of the stomach is probably related to the esophagogastric anastomosis in the neck and the resulting difficulty of passing a swallowed bolus.

Another consequence of the transposition of the stomach into the chest is the development of postoperative duodenogastric reflux, probably due to pyloric denervation, and adding a pyloroplasty may worsen this problem. Following gastric advancement, the pylorus lies at the level of the esophageal hiatus, and a distinct pressure differential develops between the intrathoracic gastric and intraabdominal duodenal lumina. Unless the pyloric valve is extremely efficient, the pressure differential will encourage reflux of duodenal contents into the stomach. Duodenogastric reflux is less likely to occur following colonic interposition, because there is sufficient intraabdominal colon to be compressed by the abdominal pressure and the pylorus and duodenum remain in their normal intraabdominal position.

Although there is general acceptance of the concept that an esophagogastric anastomosis in the neck results in less postoperative esophagitis and stricture than one at a lower level, reflux esophagitis following a cervical anastomosis does occur, albeit at a slower rate than when the anastomosis is at a lower level. Most patients undergo cervical esophagogastrostomy for malignancy; thus the long-term sequelae of an esophagogastric anastomosis in the neck are not of concern. Patients who have had a cervical esophagogastrostomy for benign disease, however, may develop problems associated with the anastomosis in the fourth or fifth postoperative year severe enough to require anastomotic revision. This is less likely in patients who have had a colonic interposition for esophageal replacement. Consequently, in patients who have a benign process or a potentially curable carcinoma of the esophagus or cardia, a colonic interposition is used to obviate the late problems associated with a cervical esophagogastrostomy. Colonic interposition for esophageal substitution is a more complex procedure than gastric advancement, with the potential for greater perioperative morbidity, particularly in inexperienced hands.

Composite Reconstruction

Occasionally a combination of colon, jejunum, and stomach is the only reconstructive option available. This situation may arise when there has been previous gastric or colonic resection, when dysphagia has recurred after a previous esophageal resection, or following postoperative complications such as ischemia of an esophageal substitute. Although not ideal, combinations of colon, jejunum, and stomach used to restore gastrointestinal continuity function surprisingly well, and allow alimentary reconstruction in an otherwise impossible situation.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 24 - Stomach

Stanley W. Ashley
Denis Evoy
John M. Daly

ANATOMY

Gross Anatomy

The stomach is an asymmetric dilatation of the proximal gastrointestinal tract, responsible for the initial digestion and storage of food. The stomach's capacity in the adult is approximately 1.5 to 2.0 L. It lies in the epigastrium, beginning at the gastroesophageal junction and extending distally to the proximal duodenum. The long axis first passes downward and then forward, upward, and to the right. It has two curvatures: the greater curvature, which constitutes its left and inferior margin, and the lesser curvature, its right and superior margin.

The stomach has been divided into anatomic regions based on its external and functional anatomy. The cardia of the stomach is poorly defined externally but represents the region immediately adjacent to the gastroesophageal junction. The fundus is the domed portion of the stomach to the left and superior to the cardia. The corpus or body extends from the fundus distally to the incisura angularis, a notch in the lesser curvature seen at the junction of the middle and distal thirds of the stomach; the greater curvature margin of the corpus is poorly defined by external anatomy. These landmarks are useful during operations on the stomach (Fig. 24-1). Functionally, the antral mucosa usually begins on the lesser curvature at the junction of the proximal and middle thirds and extends distally to the pylorus. This difference in functional and external anatomy is important when resection of the functional antrum is the goal of operation; the lesser curvature margin should extend considerably proximally to the incisura angularis.

Visceral peritoneum covers all but a small posterior portion of the stomach near the entry of the esophagus. The mucosal lining has a velvet appearance and is lined by numerous longitudinal folds, the rugae.

Embryology

The stomach arises during the fourth week of embryonic life as a dilatation of the primitive foregut. As gestation progresses, growth of the left gastric wall is accelerated relative to the right, rotating the organ so that at full term it assumes a more transverse, J-shaped orientation. This developmental pattern is responsible for the asymmetry in the length of the lesser and greater curvatures and causes the left vagus nerve to assume an anterior position while the right lies behind the distal esophagus. As they course along the stomach, the left and right nerves then lie in the anterior and posterior leaflets of the lesser omentum.

Anatomic Relationships

The close anatomic relationship of the stomach to a variety of upper abdominal organs is of considerable clinical significance. The left lobe of the liver extends over the anterior wall of the proximal stomach and frequently must be retracted laterally to obtain adequate exposure of the gastroesophageal junction. The spleen lies superior and to the left of the stomach, sharing its blood supply with that of the greater curvature, and is at risk of iatrogenic injury during operations on the proximal stomach. The stomach has anterior and posterior relations with the diaphragm and is subject to several forms of diaphragmatic herniation. The lesser sac separates the stomach from the pancreas, but this space may be obliterated by inflammation. As a result, pancreatic pseudocysts are often densely adherent to the posterior gastric wall; in this instance, cystogastrostomy (surgical drainage into the gastric lumen) represents the most appropriate therapy. Splenic and pancreatic enlargement can directly impinge on the gastric lumen, producing symptoms of early satiety. The transverse colon lies immediately inferior to the stomach and neoplasms of either organ may extend to involve the other. Gastric neoplasms may extend directly into any adjacent organ.

Blood Supply

The stomach has a particularly rich blood supply from four major and several minor arteries, all of which arise from the celiac axis (Fig. 24-2). The lesser curvature is supplied by the right gastric artery, arising from the hepatic artery, and by the left gastric artery, originating directly from the celiac axis. It is not unusual for the left hepatic artery to originate from the left gastric artery, in which case hepatic perfusion can be compromised when the left gastric artery is divided. The greater curvature is supplied by the right and left gastroepiploic arteries, which arise from the gastroduodenal and splenic arteries, respectively. The short gastric vessels arise directly from the splenic artery and supply the fundus and proximal body of the stomach. These arteries are closely interconnected through a dense submucosal vascular plexus, permitting division of up to three of the stomach's major arteries without significantly compromising perfusion. This vascular redundancy facilitates the mobilization of the stomach for reconstruction of the esophagus because it may be based on a single vascular pedicle, although the right gastric and gastroepiploic arteries are usually both preserved. It also explains the profuse bleeding that can occur with relatively small mucosal lesions.

The venous drainage of the stomach parallels the arterial supply, eventually draining into the portal circulation. The coronary vein, which corresponds to the left gastric artery, assumes particular significance in portal hypertension when it serves as a conduit between the portal and systemic circulations via the esophageal plexus, forming esophageal varices. The vein of Mayo or prepyloric vein is a typically prominent vein on the anterior pylorus, useful in identifying this region at the time of operation.

Lymphatics

The lymphatics also follow the vascular distribution and are extensively interconnected (Fig. 24-3). Malignant disease involving the gastric lymphatics often spreads to nodal groups well beyond the area of primary drainage. The primary drainage site of the proximal lesser curvature is the superior gastric lymph nodes near the gastroesophageal junction, while the proximal greater curvature is served by splenic and omental nodal groups. The distal stomach drains into suprapyloric and subpyloric nodes. Secondary nodal regions include the celiac axis, the porta hepatis, and the pancreas. This anatomy has received increasing attention as a result of reports suggesting that a more extensive lymph node dissection during gastrectomy for gastric cancer may be associated with improved survival rates.

Innervation

The autonomic nervous system plays an important role in controlling gastric secretory and motor function. The left and right vagus nerves, carrying parasympathetic efferent and afferent fibers, lie on the anterior and posterior surfaces of the esophagus, respectively (Fig. 24-4). The left or anterior vagus is usually smaller and more closely approximated to the anterior surface of the distal esophagus; the larger posterior trunk is typically found in a more intermediate position between the esophagus and the aorta. Often small fibers branch from the main trunks on the esophagus and travel separately to innervate acid-secreting cells of the proximal stomach. The "criminal" nerve of Grassi is one such fiber that arises from the posterior trunk and may play a role in recurrent ulceration after vagotomy if the nerve is sectioned distal to where it leaves the vagus. Below the diaphragm, the anterior vagus gives off a hepatic branch, and the celiac branch arises from the posterior vagus. The remaining fibers, the nerves of Latarget, travel in the anterior and posterior leaflets of the lesser omentum and give off branches to the adjacent corpus. At the antrum the final fibers innervating the antral and pyloric musculature are referred to as the "crow's foot." Sympathetic nerves also reach the stomach via the celiac plexus. They travel to the gastric wall in association with the major vessels.

Histology

The wall of the stomach is composed of four distinct layers: mucosa, submucosa, muscularis, and serosa. The surface of the mucosa is lined by a columnar epithelium believed to play a protective role, secreting both mucus and bicarbonate. The surface is spotted or cobblestoned with pits that lead into gastric glands that differ in each of the three functional regions of the stomach (Fig. 24-5). In the cardia, the transition zone between the squamous epithelium of the esophagus and the simple columnar epithelium of the stomach, the gastric glands are lined by clear-staining, mucus-secreting cells. The cardiac glands probably serve to lubricate foodstuffs and also to protect the mucosa of this region from acid injury. The glands of the fundus and body, oxyntic glands, are lined by acid-secreting parietal cells and pepsinogen-secreting chief cells (Fig. 24-6). Several glands may enter into a single pit. They also contain undifferentiated stem cells, mucus neck cells, and a variety of argentaffin endocrine cells. The glands of the pylorus and antrum are lined with mucus-secreting cells and gastrin-producing G cells. The gastric glands consist of six major cell types: surface, mucous neck, progenitor, chief, parietal, and endocrine. The surface epithelial cells are distinguished by abundant mucous granules within their apical surface. These cells are designed to protect the epithelium from ingestants and from the injurious effects of gastric acid. They are also the source of a sodium-rich alkaline secretion. The mucous neck cells line the entrance to the gastric glands. Cells at the base of the gastric pits serve as stem or progenitor cells for the development of new surface cells and also the cells of the gastric glands. Chief cells are the source of pepsinogen, a proteolytic enzyme that is converted to its active form, pepsin, at a pH below 2.5. A variety of endocrine cells exist within the gastric gland.

The parietal cells, secreting both hydrogen ions and intrinsic factor, contain numerous mitochondria, consistent with the intense energy requirements for acid secretion. They possess deeply invaginated intracellular secretory canaliculi that connect with the luminal or apical membrane and, in the resting state, numerous intracellular membranous structures called tubulovesicles (Fig. 24-7). With secretory stimulation, the tubulovesicles fuse with the canaliculi, markedly increasing the luminal surface area for secretion of acid.

The chief cells contain an abundant rough endoplasmic reticulum for packaging of protein. Zymogen granules, containing inactive pepsinogen, are concentrated within the apical cytoplasm for luminal secretion via exocytosis. In contrast, antral G cells contain many basally located granules, which fuse with the basolateral membrane, releasing gastrin into the circulation.

The muscularis mucosa is a thin layer of smooth muscle that separates the mucosa, consisting of the epithelium and its underlying lamina propria, from the submucosa. The submucosa is composed of connective tissue and a rich plexus of arteries, veins, lymphatics, and nerves. External to this, the muscularis is composed of an inner layer of oblique muscles and a middle layer of circular muscle and is surrounded by longitudinal muscle. Proximally, the three layers are easily recognized, and through the process of receptive relaxation, they permit the remarkable dilatation of the stomach that occurs after ingestion of a meal. The circular layer is most developed in the antrum and the pylorus, where it serves to grind and mix the meal, regulating the release of food particles into the duodenum.

PHYSIOLOGY

The stomach performs two interrelated functions in the initial phase of digestion. It begins the process of food breakdown, exposing a solid meal to the proteolytic action of acid and pepsin, and grinding and diluting the mixture to form a more uniformly consistent chyme. It acts as a reservoir, relaxing to accommodate the meal and regulating the release of swallowed food into the duodenum in small boluses. The process of early digestion requires that solid foodstuffs be stored for a prolonged period of time (4 h) as they undergo reduction in size and preliminary breakdown into basic metabolic constituents.

The storage function of the stomach is greatly enhanced by the process of receptive relaxation. This is an event in which the upper portion of the stomach relaxes as the intake of food is anticipated. Solid food settles and layers within the greater curvature of the fundic area of the stomach. Liquids pass rapidly from the stomach along its lesser curve (the magenstrasse or canalis gastricus), thereby leaving the solid mass quite undisturbed. Processing of the food mass is initiated by a skimming from the outermost layers of the gastric bolus. Salivary digestion occurs within the middle of the bolus, and gastric digestion at its periphery. Food particles are reduced in size by a grinding action of the antrum as well as

digestion and dilution by the gastric secretions. The storage function of the stomach is enhanced by the antrum and pylorus, which constantly return material to the proximal stomach until it is ready for delivery to the duodenum.

Gastric Secretion

Digestion, which occurs in the stomach, is designed to reduce the size of food particles and begin the dispersion of nutrients so that they can be absorbed. There is considerable redundancy in these actions and most of the digestion occurs beyond the pyloric sphincter. The action of gastric acid and pepsin, in conjunction with the grinding that occurs in the distal stomach, increases the efficiency of the process and allows consumption of larger and less frequent meals.

Acid Secretion

The main component of gastric secretion is hydrochloric acid and its accompanying water. Acid is secreted by parietal cells in the glands of the fundus and body of the stomach. The acid pump mechanism is the hydrogen-potassium–stimulated adenosine triphosphatase (H^{+}/K^{+} -ATPase) enzyme system that exchanges cellular H^{+} for luminal K^{+} (Fig. 24- 8). Adenosine triphosphate (ATP) is hydrolyzed to catalyze the exchange, and the H^{+} is concentrated approximately 2.5 million times in the secretory canaliculi. K^{+} and Cl^{-} are secreted into the lumen down their electrochemical gradients by passive diffusion across the parietal cell apical membrane. The OH^{-} generated during the formation of the H^{+} from water is converted to HCO_3^{-} by carbonic anhydrase. The bicarbonate is released into the submucosal capillaries, in exchange for Cl^{-} , producing the so-called “alkaline tide,” which can be measured in venous blood from the stomach during acid secretion.

The regulation of gastric secretion is an extremely complex phenomenon, involving endocrine, neural, paracrine, and even autocrine mechanisms. The three major stimulants appear to be gastrin, acetylcholine, and histamine. Gastrin reaches the parietal cell after it is released into the systemic circulation from the antrum and proximal duodenum. Acetylcholine is secreted by cholinergic nerve terminals in close proximity to the acid- secreting mucosa. Histamine acts in a paracrine fashion following its secretion from mast cells in the lamina propria of the fundus and body.

The parietal cell has been shown to possess receptors for histamine, acetylcholine, and gastrin, the three major stimulants of acid secretion. Each of these is believed to act through second messenger mechanisms to increase protein phosphorylation in the parietal cell (Fig. 24-9); this in turn increases the activity of the ATPase. Histamine acts via adenylate cyclase, and gastrin and acetylcholine are believed to increase intracellular calcium through a pathway involving the phosphatidylinositols. Pharmacologically, the histamine is directed at H_2 receptors, and the acetylcholine receptor on parietal cells is of the muscarinic variety. The resulting phosphorylation in turn activates the H^{+}/K^{+} -ATPase, increasing acid secretion. The response to a combination of these stimulants is greater than the sum of the responses to each agent administered individually, a phenomenon referred to as potentiation. A reduction in the amount of stimulation at one receptor also reduces the responsiveness to the other agonists, an interaction exploited in both medical and surgical management of acid-peptic disease (e.g., blockade of the histamine receptor decreases the responsiveness to acetylcholine and gastrin). The parietal cell also possesses inhibitory receptors for somatostatin and the prostaglandins.

Between meals the stomach secretes 2 to 5 mEq of acid per hour, although there is some diurnal variation, with a maximum during the early evening hours. This resting secretion is believed to be related to vagal tone and basal histamine secretion. The response to a meal has been divided conceptually into three phases: cephalic, gastric, and intestinal.

The cephalic phase, originally described by Pavlov, is triggered by the sight, smell, and taste of food. Acid secretion is mediated through the vagus nerve. The acetylcholine stimulates parietal cells directly, inhibits somatostatin release, and directly induces gastrin release from antral G cells. The cephalic phase produces approximately 10 mEq of acid per hour.

The gastric phase, in which gastrin is the most important mediator, is the result of food entering the stomach. Antral distention, amino acids and small peptides, and the increase in luminal pH produced by the buffering capacity of the meal all increase gastrin release. Gastrin is a family of polypeptides of varying lengths, formed from the breakdown of preprogastrin, the 101 amino acid precursor. A 34 amino acid breakdown product (G34) is the major circulating form, while G17 appears to be the predominant hormonal stimulus to secretion. In addition to its effects on acid secretion, gastrin is a trophic hormone, stimulating growth not only of the gastric mucosa but also of duodenum, pancreas, and colon. The gastric phase produces approximately 15 to 25 mEq of acid per hour; it is limited not only because gastric distention also causes somatostatin release but also because acidification eventually inhibits gastrin release. When luminal pH falls to 2.0, gastrin release is completely abolished.

The entry of chyme into the duodenum initiates the intestinal phase of secretion. The effect is small and variable, accounting for approximately 5 percent of meal-induced secretion. It is probably the result of absorbed amino acids and gastrin release from the proximal duodenum. Of greater significance are the inhibitory effects of the duodenum. A variety of hormones, including secretin, peptide YY, *gastric inhibitory peptide (GIP)*, and neurotensin, may be important in this process. The complete integration of this redundant regulatory system is only partially understood.

Gastric secretion has several other important consequences. The acidic conditions sterilize the proximal gastrointestinal tract; only a few fastidious organisms, such as *Helicobacter pylori*, can withstand the low luminal pH. In contrast, when acid secretion is inhibited, the stomach and small intestine are rapidly colonized with enteric organisms. This usually is of no consequence, but achlorhydric patients have been shown to be more susceptible to salmonellosis and cholera. Patients undergoing operation on the proximal gastrointestinal tract are vulnerable to infection and in preparation should receive prophylactic antibiotics.

The stomach also plays a role in iron and calcium absorption; after gastrectomy, iron deficiency anemia and bone disease are not uncommon. Gastric acid appears to play an important role in the mobilization and solubilization of dietary iron and calcium. The acidic conditions contribute to the conversion of iron to the ferrous state, in which it is best absorbed in the duodenum.

Other Secretory Functions

In addition to hydrochloric acid, the parietal cells also secrete intrinsic factor in response to many of the same stimuli. Secretion is usually in considerable excess (approximately 100×) of that needed for cobalamin (vitamin B₁₂) absorption in the terminal ileum. Gastrectomy and atrophic gastritis, conditions that can produce a reduction in parietal cell mass, can be associated with vitamin B₁₂ deficiency and megaloblastic anemia.

Pepsinogen is secreted by the chief cells in response to many of the same stimuli that affect the parietal cells; the most important stimulus is acetylcholine. Like acid secretion, pepsinogen secretion seems to be inhibited by somatostatin. Pepsinogen is autocleaved to active pepsin under acidic conditions; it is most active at pH 2.5 and denatures at pH 7.0 or greater. Pepsin catalyzes the hydrolysis of a variety of peptide bonds and initiates the digestion of collagen and other proteins. In human beings, two distinct forms, PG I and II, have been identified. Despite extensive investigation, there is no clear evidence of any correlation with gastric disease states.

The surface epithelial cells of the mucosa secrete a combination of mucus and bicarbonate. The bicarbonate is generated by carbonic anhydrase within the epithelium and then secreted across the apical membrane in exchange for Cl^{-} . It is believed that the combination of mucus and bicarbonate plays an important role in protecting the mucosa from acid injury. Recent evidence suggests that as a consequence of the bicarbonate secretion, a pH gradient is established in the mucous layer that coats the luminal surface. At its luminal surface the mucus is acidic, but as H^{+} ions diffuse through the layer toward the epithelium they are buffered by the bicarbonate, protecting the mucosa from acid injury.

Gastric Motility and Emptying

Gastric motor activity involves two distinct patterns that are distinguished anatomically and functionally (Fig. 24-10). The first pattern involves the smooth muscle of the proximal one-third of the stomach. Myocytes of this region demonstrate no spontaneous myoelectric activity and instead relax in response to the stretch produced by increasing gastric volumes. This receptive relaxation means that a meal can be accommodated with little increase in intragastric pressure. Relaxation is believed to occur via a vagally mediated reflex and permits the stomach to act as a reservoir for several hours after a meal. Contractile activity in the proximal stomach gradually increases, and food is propelled distally by the pressure gradient.

The motility pattern of the distal stomach is quite different. Beginning at a pacemaker site on the greater curvature, a series of myoelectric complexes pass distally at a rate of three times per minute. Action potentials are sometimes superimposed on these spontaneous depolarizations, resulting in a peristaltic wave that propagates distally. Vagal activity, gastrin, and motilin all increase the frequency of the action potentials, and secretin, glucagon, and *GIP* seem to reduce it.

The movement of the meal is influenced by each of the motility patterns (Fig. 24-11). Initially food enters the stomach and is stored in its proximal portion, where it is gradually digested and diluted. As pressure increases proximally, the bolus enters the antrum, where it is propelled toward the pylorus by a peristaltic wave. The pylorus actually closes several seconds before the arrival of the wave, allowing only a small amount of liquid and suspended food particles to enter the duodenum, while the main mass of food is repulsed back into the proximal stomach. This process, referred to as trituration, serves to grind the food into smaller pieces and mix it further with gastric secretions.

Sphincters

The stomach is bounded proximally and distally by sphincters. The *lower esophageal sphincter (LES)* is a high- pressure zone of the distal esophagus just proximal to the cardia. It normally relaxes only in response to a peristaltic wave from above, allowing the passage of a meal and swallowed secretions into the proximal stomach. Between relaxations, the high pressure prevents the reflux of gastric contents into the esophagus. LES physiology is discussed in detail in Chap. 23.

In contrast with the *LES*, which has no anatomic correlate, the pyloric sphincter is a thickening of smooth muscle that not only regulates gastric emptying but also prevents reflux of duodenal contents into the stomach. The consequence of duodenogastric reflux after surgical disruption of the sphincter, alkaline reflux gastritis, is discussed later, in the section Postgastrectomy Syndromes.

DIAGNOSIS OF GASTRIC DISEASE

Symptoms and Signs

Patients with disorders of the stomach present with a variety of symptoms and signs, depending on the condition. None of the symptoms and signs alone is specific enough to make a diagnosis, but a constellation of them can strongly suggest a diagnosis of gastric disease.

The loss of appetite is not unusual in a variety of conditions, ranging from viral syndromes to malignant tumors. Any form of gastric obstruction produces early satiety and accompanying anorexia. This may occur with both ulcer disease and neoplasms of the stomach. Infiltrating tumors tend to interfere with receptive relaxation, producing early satiety and a sense of fullness after consuming only small quantities of food. Significant anorexia with weight loss is often a symptom of a malignant tumor, although anorexia may not develop until the tumor is considerably advanced.

Nausea and vomiting accompany a variety of gastric disorders, although they are most common with gastric outlet obstruction. Nausea is a sensation that defies precise definition but is probably best described as a feeling of impending emesis. Vomiting itself occurs as the gastric contents are forcefully ejected from the alimentary tract in a retrograde direction. This complex reflex occurs when a contraction of the abdominal musculature is combined with elevation of the cardia (to open the gastroesophageal sphincter) and closure of the pylorus. Retching is essentially unsuccessful vomiting. Nausea and vomiting occur with peptic ulcer and neoplasms of the stomach, even in the absence of obstruction. Nausea and vomiting related to gastric outlet obstruction typically occur more than 1 h after a meal. Vomiting of undigested material eaten 12 h previously is characteristic of such disorders. Unlike in most other conditions, vomiting often relieves the pain of peptic ulcer. Compared with obstructions of the gastric outlet, the vomiting that accompanies more distal obstructions is typically bile-stained and may be accompanied by more significant abdominal distention.

Reflux of gastric contents is associated with heartburn, respiratory symptoms from aspiration, and “water brash” (a sour taste from reflux of gastric contents into the mouth). Although regurgitation may occur with gastric outlet obstruction, it is most commonly a result of disease of the *LES* (see Chap. 23).

Pain of gastric origin is typically localized to the epigastrium and may radiate to the back. A variety of disorders can produce discomfort, and the character of the pain is far from diagnostic. The gastric mucosa is devoid of pain fibers, and many disorders may be relatively painless until far advanced. Pain from gastric cancer is typically continuous and is often increased by food. Patients with a duodenal ulcer develop burning epigastric pain several hours after meals, and the pain is characteristically relieved by antacids and food. In contrast, patients with gastric ulcers have more severe pain, which occurs soon after meals and is not relieved by antacids or food consumption. Ulcer perforation is associated with severe generalized pain.

Blood loss is a common manifestation of gastric disease. Chronic mild blood loss may be occult, detected only by testing the stool with a chemical reagent. More significant upper gastrointestinal bleeding typically produces melena, which is produced by degradation of blood in the gut. Gastric bleeding can be so massive that on occasion patients present with hematochezia (bright red blood per rectum), although this is more typical of lower gastrointestinal bleeding. Hematemesis is bloody vomitus and may be fresh or “coffee-ground,” implying some contact with gastric acid.

Gastric neoplasms can produce mild chronic blood loss, and ulcer disease and gastritis may be associated with hemorrhage of varying magnitude. Mallory-Weiss tears of the gastric mucosa, typically associated with vomiting, and Dieulafoy’s lesion, small erosions from pressure ulceration from a mucosal artery, are characterized by massive hemorrhage. If upper gastrointestinal bleeding is suspected in the patient with melena or hematochezia, a nasogastric tube should be placed. This confirms bleeding from an esophageal or gastric lesion but, if negative, does not rule out a source distal to the pylorus. Most patients with suspected upper gastrointestinal hemorrhage should undergo endoscopy, which localizes the lesion and permits directed therapy and biopsy procedures.

Weight loss may be a sign of gastric disease. Benign gastric ulcer is more often associated with weight loss than duodenal ulcer, because food in the stomach aggravates the pain. Gastric malignant tumor is associated with significant weight loss and, in advanced cases, produces severe cachexia.

Distention of the stomach (gastric dilatation) may produce significant symptoms if it develops acutely. Acute distention develops in a variety of settings but is most common after abdominal operations and in the critically ill. The stomach is typically massively enlarged with gas and may be associated with activation of the autonomic nervous system, producing pallor, tachypnea, bradycardia, and hypotension. Percussion of the abdomen reveals tympany that may extend into the pelvis. Nasogastric suction will usually relieve the problem, but if allowed to progress, acute distention can produce necrosis and perforation. In contrast, distention associated with gastric outlet obstruction typically develops more slowly and may be almost asymptomatic. The characteristic physical finding is a succussion splash, audible with palpation, produced by liquid sloshing in the chronically distended stomach.

Tenderness is elicited by deep palpation of the epigastrium. Tenderness is more common with duodenal ulcer but may be associated with gastric ulcer, gastritis, and malignant tumor. It is neither a sensitive nor a specific finding. Perforated ulcers are associated with a chemical peritonitis related to irritation by the acidic gastric contents. Patients typically present with marked abdominal tenderness and “boardlike rigidity” from the guarding that occurs.

Gastric cancer has often metastasized by the time of diagnosis, and a palpable epigastric mass is not unusual. Several eponyms have been applied to common sites of such disease: Virchow’s node, an enlarged left supraclavicular lymph node; Sister Joseph’s nodule, a palpable mass at the umbilicus suggesting diffuse involvement of the peritoneal surfaces or carcinomatosis; Irish’s node, an involved left axillary lymph node; Krukenberg’s tumor, ovarian involvement with metastatic disease, and Blumer’s shelf, an anterior ridge palpable on rectal examination produced by drop metastasis into the cul-de-sac.

Diagnosis

Radiography

Radiography may be a useful initial diagnostic maneuver in the evaluation of patients with upper gastrointestinal complaints (Fig. 24-12). There has been some debate about the relative merits of radiography versus endoscopy. Radiography is more easily tolerated and is probably the procedure of choice in younger patients. In contrast, the patient over 50 years of age needs to have malignant tumor ruled out by endoscopic biopsy. Several criteria have been used to distinguish benign from malignant ulcers radiographically, such as the position of the ulcer relative to the gastric wall. Benign ulcer craters tend to extend outside the lumen, whereas the malignant ulcer is typically limited to the tumor that projects into the gastric lumen. In benign ulcers, the gastric folds radiate from the ulcer crater; in malignant disease, the normal orientation of the rugae persists. Such criteria are not completely reliable, and endoscopy with biopsy is usually indicated.

Radiography is not very sensitive for acute lesions involving only the mucosa; endoscopy is usually preferred for the diagnosis of upper gastrointestinal bleeding. Routine studies will reveal approximately two-thirds of chronic gastric ulcers and only a slightly smaller proportion of duodenal lesions (Fig. 24-13). With special techniques such as double contrast and compression, this sensitivity can be increased to 90 percent. Contrast radiography is not usually indicated in the evaluation of suspected acute perforations; a flat plate of the abdomen demonstrating free air is usually sufficient evidence to proceed with surgery, and the specific lesion can be identified at the time of operation. A contrast study will identify most advanced gastric neoplasms; it is an unreliable screening test for early lesions and usually must be followed by endoscopic biopsy.

Endoscopy

Over the past thirty years, improvements in endoscopic techniques for evaluation and treatment have revolutionized the management of lesions of the upper gastrointestinal tract. The gastroscope reveals even the most subtle mucosal lesions and permits biopsy procedures when diagnosis cannot be made on the basis of the endoscopic appearance alone. It is particularly useful in the evaluation of malignant ulcers; biopsy specimens are taken from multiple areas of the ulcer. It helps to rule out the presence of a second primary lesion. Seven or eight biopsy specimens should be taken, sampling both the crater base and the rim; additional specimens do not significantly improve the chances of identifying a malignant tumor. Cytology may increase the yield.

Endoscopic biopsy can also be used for the diagnosis of *H. pylori* infection, which appears to be a significant factor in peptic ulcer and gastritis. A combination of special stains (Giemsa or Warthin-Starry) to identify the organism histologically and assay for the presence of urease, which is produced by *H. pylori*, are typically used to establish the diagnosis in the biopsy specimen.

Endoscopy is particularly useful in the evaluation of upper gastrointestinal bleeding. The technique is safe and establishes the diagnosis in more than 90 percent of patients. Endoscopic therapy has been demonstrated with esophageal varices and peptic ulcer to reduce the incidence of rebleeding, decrease requirements for transfusion and operation, and in some instances, improve survival. A variety of effective techniques have been developed for endoscopic control of gastric and duodenal bleeding. These include thermal coagulation with the neodymium:yttrium-aluminum-garnet (Nd:YAG) laser, bipolar electrocautery, heater probe, and injection therapy with epinephrine or sclerosing agents.

Gastric Secretory Analysis

Gastric secretory analysis has traditionally been used to assess acid secretory capacity. A nasogastric tube is inserted and the patient is placed in a semirecumbent position on the left side. Basal acid output (BAO) and maximum acid output (MAO) are analyzed. Aspirations are then done by hand syringe every 5 min for 1 h. The aspirates are pooled in 15-min aliquots. At the end of the final aspirate, the stomach is stimulated to secrete by the intravenous administration of histalog in a dose of 2 mg/kg, or pentagastrin in a dose of 6 mg/kg. Aspiration is continued as described above, with four 15-min collections obtained over a 1-h period. The volume of the collections is measured, and an aliquot is titrated electrometrically to determine its content of H⁺. The rate of secretion is then expressed as the number of milliequivalents produced per hour during the basal or prestimulatory phase and during maximal and peak output. MAO is obtained by averaging the output of the two final 15-min periods. Peak output is the highest rate of secretion obtained during a 15-min period after stimulation. BAO is normally 2 to 3 mEq/h; secretory output (MAO) is in the range of 10 to 15 mEq/h. Patients with duodenal ulcer generally have higher values, and those with gastric ulcer may have lower values, but there is remarkable overlap between either condition and the normal range. Identifying patients with peptic ulcer disease using acid secretory data alone is usually not possible. For this reason most clinicians have found routine use of gastric analysis to be of little value in the workup of patients suspected of having gastric or duodenal ulcers. One beneficial situation of gastric analysis is its use as a screen for the *Zollinger-Ellison (ZE)* syndrome. In this illness, characterized by flagrant ulcer disease, BAO may be in excess of 50 mEq/h. Secretory studies may also be indicated in patients who develop recurrent ulceration after previous surgery for peptic ulcer, and may help to identify the patient who has had an incomplete vagotomy.

Gastric Emptying and Motility Studies

The saline load test has been used as a rough qualitative measure of gastric emptying; 750 mL saline solution is infused into the stomach and the nasogastric tube is clamped. At 30 min, the remaining fluid is aspirated. A residual volume of more than 400 mL suggests a significant delay in emptying.

Gastric emptying can be more precisely evaluated using gamma scintigraphy (Fig. 24-14). Liquids or solids can be labeled with technetium (^{99m}Tc) and abnormalities in the emptying of each assessed. These studies, sometimes combined with gastroduodenal manometry, are useful in the diagnosis and treatment of gastric motility disorders, e.g., postoperative atony and diabetic gastropathy.

ACID-PEPTIC DISEASE

Classification

Acid-peptic disease of the stomach and duodenum includes erosive gastritis and peptic ulcer. Each of these lesions is associated with an imbalance in the normal interplay between acid-pepsin and mucosal defense mechanisms. A complete discussion of acid-peptic disease should include the various disorders related to acid-pepsin secretion by the stomach. Although gastroesophageal reflux is such a disorder, it primarily affects the esophagus (see Chap. 23).

The distinction between gastritis and ulceration is an important one. In general terms, gastritis is inflammation confined to the mucosa of the stomach and can occur in acute and chronic forms. Chronic gastritis is important in its role in the pathogenesis of other disorders. Acute gastritis is a frequently encountered problem in the care of the critically ill. It occurs typically after major physical or thermal trauma, shock, sepsis, head injury, and ingestion of a variety of chemical agents, such as aspirin and alcohol. This lesion is generally classified under the generic term “stress erosion.”

True ulcers extend through the mucosa into the submucosa and muscularis mucosa (Fig. 24-15). Although this extension may occur occasionally in acute erosive gastritis, initiating massive hemorrhage from a large submucosal vessel, ulceration is usually chronic in nature. Chronic gastric and duodenal ulcers are distinguished by the presence of an established inflammatory reaction.

Peptic Ulcer

Peptic ulcer disease continues to affect between 3.5 to 7.5 million people annually in the United States, with an annual direct cost of almost \$4 billion, and even greater indirect costs from loss of work and diminished quality of life. There are approximately half a million new cases of duodenal ulcer and 90,000 of gastric ulcer diagnosed each year, and approximately 10 percent of the population is affected at some point in life. Compared with duodenal ulcer, gastric ulcer is a disease of the elderly; the average patient is 10 years older than patients with duodenal ulcers, and the incidence in those over 65 years of age, particularly women, is increasing.

Duodenal ulcer usually occurs in the proximal duodenum within 1 to 2 cm of the pylorus, the portion of the intestine first exposed to gastric secretion. Such ulcers have been associated with hypersecretion of acid, although this is not the case in all patients. Since acid is usually neutralized rapidly in the duodenum, ulcers occurring distal to the bulb are atypical, and an unusual cause, such as Zollinger-Ellison syndrome, must be suspected.

Three locations of gastric ulcer are recognized. Type I gastric ulcer, the most common, is located in the proximal antrum, but it may occur anywhere throughout the corpus and antrum. Its pathogenesis has been associated with a disturbance in mucosal defense. Although some acid is required for such ulcers, they are associated with hyposecretion. Type II gastric ulcer, which arises secondary to duodenal ulcer with pyloric stenosis, and Type III, the prepyloric and pyloric channel ulcer, develop in the setting of acid hypersecretion and are assumed to share a common etiology with duodenal ulcer.

Pathogenesis

The paradigm for peptic ulcer pathogenesis is an imbalance in the aggressive activity of acid and pepsin and the defensive mechanisms that resist mucosal digestion. Many different derangements have been identified, and it remains unclear which abnormalities are most important. It has become clear in the past decade that three different etiologies underlie virtually all ulcers: infection with *H. pylori*, use of nonsteroidal anti-inflammatory drugs (NSAIDs), and massive acid hypersecretion secondary to gastrinoma (the ZE syndrome).

Perhaps the most dramatic change regarding peptic ulcer has been the recognition of an association with *H. pylori*, leading many experts to conclude that ulcer is an infectious disease. *H. pylori*-induced gastritis is found in more than 95 percent of duodenal and 80 percent of gastric ulcer patients, in the absence of NSAIDs use or gastrinoma. Although the relationship between *H. pylori* and ulceration is difficult to prove, the most convincing evidence is that eradication of infection is associated with a near elimination of recurrence of peptic ulcer following treatment. The infection causes a number of physiologic abnormalities, each of which may be related to ulcer formation. *H. pylori* infection is associated with a decrease in the resistance of the mucus layer to acid permeation, a property referred to as its hydrophobicity. There is a strong link between infection and antritis or duodenitis, common denominators in patients with gastric and duodenal ulcer, respectively. Only a minority (10 percent over a lifetime) of people harboring *H. pylori* develop ulcers.

The association between NSAIDs and gastric ulcer has long been recognized; recent data have suggested an important role for NSAIDs in the pathogenesis of duodenal ulcer. The effect of NSAIDs in stress erosion is primarily topical; chronic ulcer seems to be most directly related to the systemic effects of these agents. Suppression of prostaglandin synthesis appears to be the common denominator. Prostaglandins inhibit acid secretion and have important roles in stimulating mucus and bicarbonate secretion and mucosal blood flow. The prevalence of peptic ulcers in chronic NSAID users is reported to range from 10 to 30 percent. Although most gastric ulcer patients have an element of chronic gastritis in association with their ulcer, studies of gastrectomy specimens reveal that habitual aspirin users may develop ulcers in otherwise normal mucosa.

Zollinger-Ellison syndrome is uncommon, occurring in 0.1 to 1.0 percent of all patients with peptic ulcer, but it is the best understood form of gastroduodenal mucosal injury. Ulceration results from massive hypersecretion of acid, which is stimulated by ectopic gastrin production from a nonbeta islet cell tumor, the gastrinoma. These tumors produce several forms of gastrin; the predominant form in the tumor is gastrin-17, whereas gastrin-34, with its longer half-life, is the major circulating form. Gastrinomas may be located in the pancreas but have also been identified in the duodenum. They are assumed to represent ectopic lesions. The cells are histologically distinct from those in the antrum that normally produce gastrin, and gastrin cells have not been identified in the normal pancreas. Accompanying islet cell hyperplasia has been recognized, although its significance is not yet clear. In about 20 percent of patients there is an association with the multiple endocrine neoplasia Type I (MEN I) syndrome, which includes, in addition to islet cell tumors, hyperparathyroidism and pituitary tumors. In this group, the disease has a genetic basis. At least 20 percent of gastrinomas are multiple, and more than two-thirds are malignant. They are very slow-growing, indolent tumors. The parietal cell mass is expanded enormously as a result of the trophic effects of gastrin and has been estimated to be at least three to six times as large as that in normal individuals.

Although *H. pylori* or NSAIDs probably play a role in most patients with chronic ulcer, not all patients with these risk factors develop disease. A variety of other abnormalities, specific to the individual and to the type of ulcer, may contribute to the development of peptic ulceration (Table 24-1).

Duodenal Ulcer

Patients with duodenal ulcer hypersecrete more acid, at rest and in response to stimulation, than do normal control subjects. There is considerable overlap, and there appears to be no direct relationship, between the degree of acid hypersecretion and the severity of the ulcer diathesis.

Multiple sources for this increase in acid secretion have been identified. Duodenal ulcer patients have more parietal cells, and chief cell numbers increase in parallel. Explanations have been provided for this increase in secretory mass. In some patients there is probably a genetic basis for the increase; in others it may be the result of an increase in the release of trophic factors such as gastrin. The hypergastrinemia associated with *H. pylori* infection may be significant in this regard. It may be that there is an increase not only in cell numbers but also in the capacity of the individual cell to secrete. There is an increase in basal secretion and in the response to exogenous stimuli. Some studies have demonstrated an increase in peak secretion with meals, and others demonstrate a more prolonged response.

In addition to these disorders in acid secretion, some duodenal ulcer patients have a motility abnormality. There is more rapid gastric emptying of meals, particularly liquids, and acid and food in the duodenum slow emptying to a lesser extent than in control subjects. As a result, the proximal duodenum is exposed to more acid.

Recent studies have clearly demonstrated a defect in duodenal acid disposal. Most patients with duodenal ulcer have a reduction in basal and peak duodenal bicarbonate secretion, paralleled by a loss of mucosal buffering capacity. This appears to be the most prevalent physiologic abnormality in duodenal ulcer. Defects in a variety of other duodenal defense mechanisms have been proposed. Impaired motility of the proximal duodenum, decreased production of prostaglandins, reduced mucosal blood flow, and defects in mucus or bicarbonate secretion have been hypothesized, but are difficult to demonstrate in humans. Chronic duodenitis, associated with *H. pylori*, may affect any of these parameters, setting the stage for ulceration.

Smoking is also associated with an increased incidence and impaired healing of duodenal ulcers. This may be a result of an associated decrease in prostaglandin synthesis, enhanced acid secretion, or reductions in duodenal and pancreatic bicarbonate secretion.

Although *H. pylori* and NSAIDs contribute to the pathogenesis of duodenal ulceration, a variety of other abnormal physiologic patterns have been described. These include hypersecretion of acid, rapid gastric emptying, and a defect in duodenal acid disposal or mucosal defense. The common denominator in most patients appears to be *H. pylori* infection. Recent evidence suggests that apart from NSAIDs-induced and some atypical ulcers, this disorder would not exist without *H. pylori*.

Gastric Ulcer

In contrast to duodenal ulcer patients, gastric ulcer patients secrete in a range from normal to barely detectable. Cases have been reported with no detectable acid secretion or achlorhydria, but it is believed that at least some acid-pepsin is required for ulceration, and it is clear that the defect in these patients must include a significant disorder of defense.

It has been suggested that the most basic abnormality in gastric ulcer is reflux of duodenal contents (biliary and pancreatic secretions) into the stomach, resulting in gastritis and eventual ulceration. This appears to be the result of pyloric sphincter dysfunction. Normally the pylorus has a low resting pressure that increases in response to acid, fat, amino acids, and cholecystokinin from the duodenum. In some gastric ulcer patients, low basal pressures have been documented, and in others pressure increases in response to duodenal infusion of acid or fat are significantly less than those of control subjects. Cigarette smoking, which has been linked to the initiation and delayed healing of gastric ulcer, also increases duodenogastric reflux.

The physiologic link between reflux and gastric ulceration has not been clearly established. Bile acids (particularly deoxycholate and taurocholate), lysolecithin, and pancreatic secretions are the agents in duodenal contents speculated to have the most negative effects. It is believed that they damage the mucosa topically, disturbing the surface mucus layer and somehow producing a low-grade gastritis. Aspirin seems to have similar effects. Chronic gastritis is the widely assumed intermediary step between repeated injury to the gastric mucosal barrier by reflux of duodenal contents and development of gastric ulceration. This concept of gastritis producing ulceration is unproved, and chronic atrophic gastritis is very common in the elderly, found in approximately 40 percent of persons over 50 years of age. Although chronic ulcer is usually associated with gastritis, the primary event has not been established clearly. Gastritis may be limited to the area around the ulcer, but it usually persists after ulcer healing. The extension of the usual antral pattern of gastritis into the fundus helps to explain the generally low rates of acid secretion in gastric ulcer patients. Some increase in basal hydrogen ion back-diffusion may play a role. Gastrin cells are apparently spared the gastritis because, in response to the reduced acidity, fasting serum gastrin levels are usually slightly elevated and may double the normal increase after a meal.

Gastric ulcer seems to result from a defect in gastric mucosal defense against digestion by acid-pepsin. Hypersecretion cannot be incriminated in its pathogenesis. There is evolving evidence that pyloric dysfunction, acting as a result of duodenogastric reflux, allows endogenous agents to exert injurious effects on the mucosa. These agents, in combination with acid-pepsin, produce ulceration in areas of reduced mucosal resistance. NSAIDs and *H. pylori* infection predispose to this reduction.

Clinical Manifestations and Diagnosis

Classically, duodenal ulcer pain is relieved by food or alkali and usually develops several hours after a meal when food has passed the duodenum and the crater is exposed to unbuffered gastric secretions. In contrast, gastric ulcer is associated with a gnawing or burning epigastric pain brought on by or closely following the secretory stimulus of eating. Symptoms in these two disorders are very nonspecific, and even the correlation of pain with the actual presence of peptic ulceration is a poor one. Although intractable pain has generally been considered an indication for surgery in peptic ulcer, because of its nonspecific nature, such pain is difficult to define. The actual physiology of ulcer pain is not understood, although two explanations have been proposed. Acidic luminal contents may irritate afferent nerves within the ulcer crater itself or, alternatively, peristaltic waves passing through the ulcer might produce discomfort. The relative importance of these two possibilities has not been determined. Pain symptoms in gastric and duodenal ulcer usually is chronic and recurrent. Although these ulcers cannot be differentiated on the basis of clinical findings, the fact that the mean age of gastric ulcer patients is approximately 10 years older than that of patients with duodenal lesions might help in the differentiation. Gastric ulcers have a peak incidence from the ages of 50 to 65 years, whereas most duodenal ulcers develop in the fourth decade of life. Other common symptoms include nausea and weight loss, even in the absence of pyloric obstruction, and mild epigastric tenderness.

In the patient with pain, diagnosis is usually fairly straightforward. Routine laboratory studies add little to the diagnostic workup. Because of the overlap in secretory rates between ulcer patients and control subjects, and because secretory studies have not been found to be useful in selecting therapy, they usually are not indicated. The two mainstays of diagnosis are upper gastrointestinal radiography and endoscopy. The decision of when to study a patient with dyspepsia or epigastric pain is a complex issue requiring consideration of the character, severity, and duration of symptoms. In the young patient, it may be more cost-effective to initiate a course of medical therapy and determine whether the symptoms resolve; the risk of malignant tumor precludes such an approach in the elderly.

The choice between radiography and endoscopy is not a simple one. With optimal double-contrast studies, more than 90 percent of gastric and duodenal ulcer craters will be detected, a sensitivity rate comparable to that achieved with endoscopy. Endoscopy is indicated in the case of a poor-quality x-ray study. The question of malignancy in gastric ulcer complicates the decision. Because 3 to 7 percent of gastric malignancies appear benign on x-ray study, endoscopy and biopsy procedures in all cases has been recommended. A variety of tests for *H. pylori* are now available. Biopsies can be examined histologically or by urease assay. Alternatively, a noninvasive breath test may be performed. With an endoscopically or radiographically proved duodenal ulcer, testing is often eliminated in favor of empiric treatment because the probability of a positive test for *H. pylori* exceeds 90 percent.

Complications

Peptic ulcer may produce one of three main complications: hemorrhage, perforation, or obstruction. These can develop without any premonitory symptoms but typically appear as an abrupt change from preexisting dyspepsia. All complications result from the extension of ulceration and the accompanying inflammation deeper into the wall of the gastroduodenum.

When the crater extends into a major vessel, significant hemorrhage can result. Fifteen to twenty percent of patients with peptic ulcer at some point develop gross bleeding; occult blood loss is even more common. Emergent bleeding requiring operation is often the result of posterior erosion of a duodenal ulcer into the gastroduodenal artery. Bleeding gastric ulcers present with hematemesis or melena in roughly equal proportion, whereas duodenal ulcers tend to produce melena alone. Widely disparate symptoms may stem from the resultant hypovolemia; patients have presented with transient ischemic attacks, myocardial infarction, and intestinal ischemia. Diagnostic maneuvers can be performed at the time the patient is being stabilized. The diagnosis of upper gastrointestinal hemorrhage is confirmed by passage of a nasogastric tube. Bleeding peptic ulcer accounts for only one-third of massive upper gastrointestinal bleeds; endoscopy is therefore indicated to identify the nature and site of the lesion.

When the ulcer erodes through the full thickness of the gastroduodenum, it may produce a perforation or penetrate into adjacent structures. This occurs in 5 to 10 percent of patients with peptic ulcer. With perforation, the spilled gastric juice incites both peritonitis and consequent catastrophic abdominal pain, marked tenderness, and paralytic ileus. This peritoneal irritation also is responsible for the accompanying leukocytosis and hypovolemia secondary to fluid sequestration. Pneumoperitoneum is present in 75 percent of patients. Upright films of the chest and abdomen reveal free air beneath the diaphragm (Fig. 24-16). If a perforation seals quickly, patients may seek medical attention only after a localized intraabdominal abscess develops. If the perforation is diverted by the falciform ligament into the right colic gutter, it may infrequently be confused with appendicitis. Penetration into the biliary tract or colon can produce a fistula.

Gastric outlet obstruction develops, usually in the context of chronic ulcer disease, when secondary edema or scarring occludes the lumen. Obstruction develops in less than 5 percent of patients, usually with duodenal ulcer, but occasionally with gastric ulcer. Onset is insidious, but patients present with nausea, vomiting, and abdominal distention. Vomiting of hydrochloric acid may produce severe dehydration and a metabolic alkalosis. This is perpetuated by a paradoxical aciduria as the kidneys retain bicarbonate with sodium to maintain pH in the absence of the hydrochloric acid lost in the vomitus. The diagnosis can be documented with barium x-ray, the saline load test, or sequential scintigraphy with technetium-labeled liquids or solids. Endoscopy can differentiate atony from true obstruction when the diagnosis is in question.

Zollinger-Ellison Syndrome

In patients with gastrinoma, symptoms tend to be more severe, unrelenting, and less responsive to therapy than those of usual ulcers. Although most ulcers are located in the proximal duodenum, gastric and more distal ulceration also occurs. Lesions are usually single and small, but multiple and giant ulcers have been described with greater frequency than in common duodenal ulcer. Diarrhea is also a frequent symptom and may occur in the absence of gross ulceration. This is a consequence of large quantities of hydrochloric acid entering the small bowel. There are frequently morphologic abnormalities throughout the small bowel, with stunted villi and mucosal inflammatory infiltrates. Steatorrhea may also be present, and probably results because the low pH inactivates pancreatic lipase, impairing hydrolysis of dietary fats. In addition, bile salts are precipitated at low pH, reducing the formation of the micelles required for lipid absorption. A low pH also appears to interfere with vitamin B₁₂ absorption, although intrinsic factor secretion by the stomach is normal.

The diagnosis of gastrinoma has been increasingly refined. A diagnostic search for ZE syndrome is warranted in patients with multiple, giant, or distal ulcers, ulcer disease refractory to the usual medical therapy, and recurrences after adequate surgery for peptic ulcer. Diagnostic techniques include acid secretory studies and contrast radiography. These patients have elevated acid secretion, and basal acid output is usually greater than 60 percent of maximal. There is considerable overlap with normal control subjects and patients with duodenal ulcer. Upper gastrointestinal x-ray films may reveal ulceration, prominent rugal folds indicating hypertrophy, dilated small intestine, and even the occasional tumor in the duodenum. Determination of basal serum gastrin levels by radioimmunoassay can be diagnostic, but elevations may also be detected in several other pathologic conditions.

In pernicious anemia, which is an atrophic and inflammatory gastric process that reduces the number of parietal cells but typically spares the antrum, hypergastrinemia results from both an increase in the number of gastrin cells and a loss of the normal acid feedback inhibition of secretion. Patients with renal insufficiency also develop elevated serum gastrin, although the mechanism has not yet been defined and gastrin clearance is normal even in anephric patients. Antral gastrin-cell hyperplasia or hyperfunction occurs in a small proportion of patients with duodenal ulcer. These patients have hypergastrinemia and hypersecretion, although an actual increase in the number of gastrin cells has not been documented.

Three provocative tests have improved the specificity of biochemical diagnosis. In the secretin test, 2 U/kg of secretin is given intravenously over 30 seconds, and serum gastrin is measured 5 min before, immediately before, and at 5-min intervals for one-half hour after injection. In normal patients and in patients with duodenal ulcer, secretin has no significant effect, but in patients with ZE syndrome there is a dramatic increase in serum gastrin levels. The mechanism for this is unknown, although it has been suggested that secretin may release gastrin by a direct local effect on blood flow to the tumor. In the calcium infusion test, calcium gluconate is given intravenously at a rate of 5 mg/kg for a 3-h period, and serum gastrin levels are determined 30 min before and at half-hour intervals for 4 h. In ZE patients, there is more than a 400 pg/mL increase, possibly through release of peptide hormone through the calmodulin system, and much smaller increases occur in patients with duodenal ulcer and other conditions. A combination test using both secretin and calcium is even more sensitive. A third provocative test, the use of a standard meal, produces little change from basal levels in gastrinoma patients when compared with the marked increase produced in control subjects or in patients with gastrin-cell hyperplasia. Gastrin secretion by the gastrinoma is independent of normal control mechanisms.

Computerized tomographic scanning and angiography are useful, though not in the diagnosis of gastrinoma but in preoperative tumor localization. In some cases venous sampling has provided additional information for localization, although most surgeons will explore a patient without recourse to such invasive diagnostic methods.

Treatment

Consistent with the general concept of a disturbance in the interplay between acid secretion and mucosal defense underlying peptic ulcer, the medical and surgical therapies represent attempts to restore a balance. Most treatments have approached this task by methods designed to reduce acid secretion, and some have been devised to improve mucosal resistance.

Medical Therapy

In the past, a variety of measures, generally unrelated to the pathophysiology of these lesions, were adapted empirically as part of the treatment regimen. Dietary therapy with frequent feeding of bland foods, has not been shown to be particularly effective and has little if any effect on gastric acid secretion. Milk, another home remedy, is actually a strong secretory stimulus. Hospitalization for peptic ulcer was believed to improve healing, although recent evidence suggests that there is little added benefit. Cigarette smoking has clearly been shown to retard ulcer healing and should be avoided. Aspirin and NSAIDs have a detrimental effect on mucosal resistance and should not be prescribed. Coffee strongly stimulates acid secretion, and alcohol may damage the mucosa; moderation in consumption should be suggested.

Pharmacologic therapy restores the balance of secretion by three general mechanisms: neutralization of gastric secretion, inhibition of secretion, and protection of the gastric mucosa from further injury. Table 24-2 summarizes the classes of drugs available. A combination of drugs acting by different mechanisms may have more than an additive effect in improving ulcer healing.

Antacids are the oldest drugs used. They reduce gastric acidity by reacting with hydrochloric acid to form a salt and water, inhibit peptic activity through the increase in pH, and bind bile acids, which may have a special role in the treatment of gastritis and gastric ulcer. Antacids differ greatly in their buffering strength, absorption, taste, and side effects. Magnesium antacids tend to be the best buffers but cause significant diarrhea by a cathartic action. Aluminum antacids precipitate with phosphorus, resulting in occasional hypophosphatemia and, in addition, may produce significant constipation. Calcium compounds can cause a delayed acid secretory rebound, believed to result largely from the effect of absorbed calcium ion. For each equivalent of hydrochloric acid neutralized, bicarbonate is released in the tissue, and all these drugs may produce a systemic alkalosis. Normal renal function usually prevents this from becoming significant. Antacids are best taken an hour after meals because food tends to prolong gastric emptying. Many patients have found large, frequent doses (30 mL of liquid antacid four to seven times daily) to be unacceptable.

The H₂-receptor antagonists are a group of agents that have revolutionized the treatment of peptic ulcer by directly reducing acid secretion. There are two types of gastric histamine receptors: H₁ receptors, located on smooth muscle cells, and H₂ receptors, located on the parietal cells; it is the latter that this class of drugs was devised to block. Because of the

independent but mutually augmenting parietal cell receptors for histamine, acetylcholine, and gastrin, inhibition of the histamine receptor alone also reduces the effects of acetylcholine and gastrin. These drugs all consist of a five-member organic ring and a side chain similar to histamine itself. Cimetidine was the first widely used H₂-receptor antagonist. Alteration of the ring or side chain has produced agents of increasing antagonistic potency and more specific gastric action. The physiologic effects include a decrease in both basal and stimulated acid secretion by what is believed to be competitive inhibition at the parietal cell level. Pepsin output is also reduced. The decrease in antral acidity elevates serum gastrin levels. These agents also reduce gastric blood flow, a finding that could explain the observation that they may be less effective than antacids in preventing stress erosions. Cimetidine is a potent inhibitor of acid secretion with a half-life of approximately 2 h, requiring a dosage schedule of four tablets per day. Patient compliance is much better with cimetidine than with antacids, but a number of significant side effects, including azoospermia, gynecomastia, and a reversible central nervous system toxicity, have been described. Ranitidine and famotidine, longer acting agents, appear to be less likely to cause some of the side effects of cimetidine.

Anticholinergic agents act to inhibit the action of acetylcholine at muscarinic receptors. In the stomach they act directly at the parietal cell level. Atropine and propantheline bromide are anticholinergic agents that, at equal doses, are even more potent than H₂-receptor antagonists. Side effects, including urinary retention, blurred vision, dry mouth, delayed gastric emptying, and mental disturbances, restrict use of these agents except at lower, less effective doses. Pirenzepine, a selective M1 anticholinergic, inhibits secretion with minimal side effects.

An even more potent class of secretory inhibitors is the substituted benzimidazoles, which selectively inhibit parietal cell H⁺/K⁺-ATPase, the enzyme responsible for acid secretion. Omeprazole is a highly potent agent that is effective even in patients with refractory ulcers and gastrinoma. It is a weak base and accumulates selectively within the acidic environment of parietal cell. Long-term side effects may become evident, but at present this drug represents the most effective secretory inhibitor.

Several different agents are effective by improving gastric mucosal resistance. The prostaglandins are naturally occurring fatty acids that prevent or heal mucosal injury by several mechanisms. The methylated E2 analogs are absorbed orally and significantly inhibit gastric acid secretion; at doses required for this effect, they also produce dramatic diarrhea. At smaller doses they still exert what has been called a "cytoprotective" action, probably by increasing mucosal blood flow and bicarbonate and mucus secretion. Misoprostol, the most widely used analog, appears to be comparable to H₂-receptor antagonists in the healing of duodenal ulcers.

Sucralfate is related structurally to heparin, although it has no anticoagulant effects. Sucralfate's unique mechanism of action has proved effective in the treatment of established ulcer disease, and it is the only agent currently in widespread clinical use that acts by enhancing resistance. It is an aluminum salt of sulfated sucrose that dissociates under the acidic conditions in the stomach. The sucrose polymerizes and binds to proteins in the ulcer crater, producing a kind of protective coating that lasts up to 6 h. It inhibits peptic activity and appears to bind and concentrate endogenous basic fibroblast growth factor, stabilizing and concentrating this substance, which is important in mucosal healing. The liberated aluminum hydroxide may have some slight additional antacid effect; little of this is absorbed, and side effects are minimal.

Colloidal bismuth coats exposed protein and has been shown to protect against acute mucosal lesions in a variety of experimental models. An antipeptic activity has been postulated to explain its effectiveness. It also has activity against *H. pylori*, although complete eradication typically employs combined therapy.

For eradication of *H. pylori*, two different regimens are used: bismuth- based triple therapy and proton pump inhibitor—based antibiotic therapy. In the former, bismuth is combined with tetracycline and metronidazole; it is highly effective and inexpensive. Alternatively, a proton pump inhibitor, such as omeprazole, is combined with two of the three following antibiotics: clarithromycin, amoxicillin, and metronidazole.

Surgical Therapy

Several operations have been devised for the treatment of acid-peptic disease. Some of these procedures have been commonly used in the elective setting when operation is performed for recurrent disease and intractable pain. The recognition that *H. pylori* plays a role in many of these patients, and that eradication of the infection prevents recurrence in more than 90 percent of patients, has almost eliminated the need for elective surgical procedures. Occasionally elective operation will be required for the noncompliant patient or for the patient who requires NSAIDs for control of other conditions. Nonhealing of a gastric ulcer also may be an indication for elective operation. For the most part, surgery for peptic ulcer is now performed for complications in an emergent or semi-urgent setting.

Treatment for Specific Entities

Duodenal Ulcer

Many patients are able to self-medicate their disease with nonprescription medication, never seeking formal medical therapy. Of those who seek professional treatment, more than 95 percent will have their disease controlled by pharmacologic therapy designed to reduce acid secretion. Antacids are extremely effective, but their inconvenience has made the H₂-blockers the drugs of choice. Combination therapy usually is only necessary in refractory patients; antacids or cimetidine alone will heal all but about 25 percent of duodenal ulcers within 4 weeks. Sucralfate is as effective as cimetidine. Colloidal bismuth, in a few clinical trials, also appears to equal the H₂-receptor antagonists. There is debate as to whether an actual ulcer crater need be documented by x-ray film or endoscopy before starting therapy; it has been suggested that the symptoms and potential complications should be the most important factors in determining treatment. Many ulcers heal, and most symptoms disappear with placebo. Though pharmacologic treatment is more expensive, the more rapid response and greater probability of healing justify the use of antiulcer medication. Patients should be advised to stop smoking and to avoid NSAIDs if possible.

None of these regimens affects the natural progression of the disease process itself. When therapy stops, ulcers will recur within 6 months in about 80 percent of patients. Continuation of cimetidine at lower doses, perhaps only once daily, after healing has been shown to reduce significantly the risk of recurrence, and prior to the recognition of the role of *H. pylori* in this disorder, some patients were maintained on such therapy. Treatment of *H. pylori* with either the bismuth-based or proton pump inhibitor regimen prevents recurrence in more than 95 percent of cases, and reinfection is unusual. There is some controversy about whether every duodenal ulcer patient should be treated with antibiotics; most clinicians now base therapy on the results of testing for the organism, but it may be judged more cost effective to treat empirically in the future.

Classically the four indications for surgery include: intractability, complications of hemorrhage, perforation, and obstruction. Elective operation for intractability has become an increasingly rare occurrence. The choice of an operative procedure, and its ultimate results, depends critically on the indication for surgery and the specific pathophysiology of the indicating condition. In the treatment of intractable pain, surgery is elective, the patient is adequately prepared for the procedure, time is not a crucial factor, and there is little justification for exposing the patient to the potential for severe early or late postoperative complications. Under these circumstances, highly selective cell vagotomy is clearly the procedure of choice. The lumen of the gastrointestinal tract is not entered, reducing the risk of septic complications. Because the normal functional anatomy of the stomach and pylorus are preserved and distal vagal innervation is not interrupted, the potential for the serious long-term side effects of dumping and diarrhea is greatly reduced. Some have argued that this is now the only definitive operation that should be performed under any circumstances.

Hemorrhage is the principal cause of death from duodenal ulcer. These patients are sometimes critically ill and often have significant associated medical problems delaying the decision to operate, although most studies indicate that early operation is the only way to reduce mortality and morbidity. Upper endoscopy should be performed early to establish the diagnosis and initiate therapy. Once bleeding has stopped, H₂blockers reduce the rebleeding rate. Indications for operation include continued bleeding of more than 6 units and recurrent bleeding after the hemorrhage has been controlled endoscopically. At operation, the initial maneuver should be to attend directly to the bleeding lesion through a pyloroduodenotomy. Most surgeons would then perform some form of definitive ulcer operation, although there is some evidence now that, with appropriate medical therapy, there is no need for such a procedure. In the past vagotomy and antrectomy were performed in patients with a good risk profile. This is probably not indicated today because of the long-term side effects and the lower recurrence rates after treatment of *H. pylori*. Highly selective vagotomy should be performed after closure of the pyloroduodenotomy, if the patient can tolerate the longer operating time. In the higher-risk patient, vagotomy and pyloroplasty can be performed quickly and simply.

Formerly, perforation was treated by simple closure, usually reinforced with omentum (the Graham patch). This is a rapid, effective treatment and is still recommended for patients with preoperative shock, perforation exceeding 48 h, and significant coexistent medical problems. Before development of therapy for *H. pylori*, 80 percent of those treated with simple closure developed a recurrence, and one-third required another operation. With proper patient selection, definitive treatment can be performed, and vagotomy, pyloroplasty, and parietal cell vagotomy have been used. The latter seems to be receiving greater support, although there may be a trend back toward patching alone because of the development of effective medical therapy to prevent recurrence.

Treatment for obstruction is designed to relieve the blockage. Highly selective vagotomy and dilatation of the pylorus have been used. Although there has been some hesitation in using vagotomy in the presence of already disturbed motility patterns, it should probably be included in the definitive operation. Vagotomy with antrectomy, resecting the scarred pylorus, is the procedure of choice, although vagotomy with gastroenterostomy is preferred by some surgeons.

Gastric Ulcer

Trials of medical therapy for chronic gastric ulcer have not been as successful as those for duodenal ulcer. Several conclusive clinical studies have shown cimetidine to be significantly more effective than placebo in healing gastric ulcers, although other investigators were unable to replicate this finding. Two considerations are important in evaluating the therapy for chronic ulcer: healing and pain relief. There is evidence of better pain relief with cimetidine in most studies of gastric ulcer; antacids have not been shown to be effective, but the physiologic reasons for this difference are still unclear. The anticholinergics also have not proved particularly useful in chronic gastric ulcer. Sucralfate appears to speed healing, but in one study it was found to be less effective than cimetidine. Omeprazole appears to be equally efficacious. Current recommendations suggest that these patients be started on an H₂ blocker. If *H. pylori* is involved, triple therapy is also instituted. If pain relief has not been achieved after 2 weeks, a second medication should be added. At 8 weeks, follow-up x-ray films or endoscopy should be performed. If the ulcer is not completely healed, these studies should be repeated again at 12 and 15 weeks. Patients whose ulcers have not healed by this time should be offered surgery.

Gastric ulcer has been treated with surgery somewhat earlier than duodenal ulcer. There are several reasons for this approach. One is the fear of malignancy in ulcers that appear benign. Improvements in the endoscopic diagnosis of cancer have reduced this risk. Most data indicate that gastric ulcer patients are more often hospitalized for their illness than duodenal ulcer patients, and their complications more often require operation, suggesting that gastric ulcer is a more virulent disease. The older gastric ulcer patient has a distinctly higher mortality for a complication of the ulcer than the corresponding younger patient with a duodenal lesion. Surgery is generally recommended for the recurrent ulcer refractory to medical therapy, the ulcer that recurs during treatment with cimetidine, the ulcer that fails to heal within 12 to 15 weeks, or one of the complications of ulcer disease.

Prepyloric ulcers or those in association with duodenal ulcers share a common cause with duodenal lesions and should be treated as such. Parietal cell vagotomy for such lesions is associated with high recurrence rates, although some surgeons believe that even if it is combined with antisecretory medication, this is preferable to a major resection. It is unnecessary to remove a large amount of stomach for elective treatment of gastric lesions. For the usual Type I gastric ulcer, distal gastric resection to remove the ulcer with or without vagotomy has proved effective. The ulcer itself should be excised completely and submitted for pathologic analysis. Because gastric ulcer does not primarily result from a secretory disturbance, the use of vagotomy has been questioned. For proximal gastric ulcers, in the fundus or the corpus, total gastrectomy is often used because proximal gastrectomy is invariably associated with emptying problems. The specific operations for complications of gastric ulcer are similar in rationale to those performed for duodenal ulcer.

Zollinger-Ellison Syndrome

Treatment of gastrinoma continues to evolve. In the past, recognition that these tumors were frequently multicentric and metastatic, though slow- growing, led to the use of the only operation believed to reduce acid secretion reliably and consistently: the total gastrectomy. With the advent of cimetidine, most clinicians adopted pharmacologic therapy in preference to such a major procedure with its considerable accompanying morbidity and mortality. This trend was strengthened with the development of even more effective omeprazole therapy. Recently there has been a move back to surgery for several reasons. First, using a combination of intraoperative endoscopy and ultrasonography to visualize duodenal lesions, and thorough exploration of the head of the pancreas and hepatoduodenal ligament to localize others, much higher success rates have been reported for resection. In addition, compliance problems have raised some doubts about the desirability of continuing medical therapy. Most patients without evidence of liver metastases should be explored. Because of the low grade of this malignancy, even if it is unresectable, if acid secretion is effectively inhibited medically, long-term survival is possible. Little other effective treatment, e.g., radiotherapy or chemotherapy, is available.

Acute Gastritis

Improvements in resuscitation and care of the critically ill patient, along with prophylactic measures directed specifically at reducing gastric injury, have considerably reduced the incidence and severity of acute erosive gastritis. Stress erosions are usually multiple, small, punctate lesions situated in the proximal acid- secreting portion of the stomach, although they may extend into the antrum and even the duodenum. They occur in three clinical settings, two of which are associated with a reduction in the ability of the gastric mucosa to protect itself against injury. The first is that of patients with severe illness, trauma, burns, or sepsis, virtually all of whom develop these lesions, although they achieve clinical significance only in the small percentage of lesions that extend into larger submucosal vessels, producing life- threatening hemorrhage. In the setting of thermal injury, acute erosions have been referred to as Curling's ulcers, although no real difference exists. Patients in this critically ill group have multiple reasons for a depression of normal defense mechanisms, although mucosal ischemia seems to be a predominant factor. The second setting for acute erosions is in the context of drug and chemical ingestion. Aspirin, a variety of other NSAIDs, and alcohol may produce an acute erosive gastritis. These agents directly alter gastric mucosal resistance, allowing back-diffusion of acid and further damage. The third underlying condition, central nervous system trauma, is not clearly related to a defect in mucosal defense. These patients have elevated levels of serum gastrin and, most likely, a secondary increase in acid secretion. This lesion, Cushing's ulcer, is characteristically deeper than other acute erosions and more frequently perforates.

Pathogenesis

Several different illnesses appear to predispose to the development of acute erosive gastritis. Most of these reduce the ability of the stomach to protect itself against acute injury rather than increasing the amount of acid secretion. Experimental evidence suggests that in the critically ill patient with hemorrhagic shock and sepsis there may actually be a reduction in acid secretion. In general, Schwarz's "no acid, no ulcer" dictum remains valid, and complete neutralization of gastric acidity prevents the development of these lesions.

Given that the presence of luminal acid is a necessary but not sufficient prerequisite for the development of stress lesions, it seems reasonable to postulate that there might be a decreased mucosal resistance to this acid. The gastric mucosa possesses a unique resistance to luminal acid and pepsin; virtually any other region of the gastrointestinal tract ulcerates rapidly when exposed to gastric juices. Davenport suggested that there was normally a gastric mucosal barrier to the back-diffusion of acid into the tissue. He noted that substances such as aspirin, bile salts, and alcohol reduced this normal resistance, allowing the backward flux of hydrogen ions that might produce histamine release, vasodilatation, and eventual bleeding (Fig. 24-17). This concept of a barrier has been supported by a number of subsequent experimental studies with chemically induced erosions. Other models for stress lesions suggest that barrier disruption may not be an essential component of the pathologic process in all types of injury; hemorrhagic shock and endotoxemia, for example, can produce lesions without overt evidence of a back-diffusion. Even in these instances, diffusion of a smaller magnitude probably occurs. In severely traumatized or septic patients, endogenous bile salts may chemically disrupt the barrier. Clinical observations suggest that reflux of bile from duodenum to stomach is more common in critically ill patients, probably as a result of the loss of normal intestinal motility. The combination of acid, bile salts, and mucosal ischemia is remarkably ulcerogenic.

Mucosal ischemia is a critical pathogenic factor. Many of these patients have experienced an episode of shock from hemorrhage, sepsis, or cardiac dysfunction, and decreased mucosal blood flow is a common denominator in many experimental models of stress erosion. The exact role of blood flow in maintaining mucosal defense remains uncertain. The leading hypothesis is that it somehow functions to dispose of or to buffer acid entering the tissue. Ischemia reduces this capacity, lowering intramucosal pH. It may also reduce mucosal resistance by producing a deficit in mucosal adenosine triphosphate and other high-energy phosphate intermediates.

Systemic and gastric acid-base balance are also significant. Acidosis has been shown to reduce the ability of the gastric mucosa to protect itself against injury in animal studies. The secretory status of the mucosa may be a critical factor. With acid secretion, bicarbonate is released into the tissue and blood (the "alkaline tide"), and this appears to play a role in buffering back- diffusing acid. In animal studies, the actively secreting stomach is more resistant to luminal acid than it is under conditions of secretory inhibition.

A variety of other mechanisms are important to mucosal defense and may somehow be impaired in patients who develop stress ulcers. It has been demonstrated that the mucus layer over surface epithelial cells may serve as a layer in which actively secreted bicarbonate neutralizes luminal acid before it can reach and damage the cells. There is some evidence that the maintenance of this pH gradient is altered in these patients. A reduction in the rate of epithelial renewal, or restitution, may also be a factor. Prostaglandins exert a protective effect in the gastric mucosa, and the inhibition of their production by aspirin and other NSAIDs may be a primary mechanism for gastric injury. Prostaglandins in high doses inhibit acid secretion, but even at lower concentrations they seem to exert a "cytoprotective" effect. They increase mucosal blood flow, but they also enhance bicarbonate and mucus secretion.

The recognition of these pathogenic mechanisms has provided a rationale for the successful prevention of stress erosions in critically ill patients. The prerequisite of acid secretion provides a basis for the regular use of antacid or H₂-receptor blockade, the efficacy of which has been demonstrated in a number of clinical trials. The correction of abnormalities in cardiac output and intravascular volume and the treatment of any septic focus may be critical in the prevention of mucosal ischemia. Adequate nutritional therapy is suggested by the concept of a mucosal energy deficit; the correction of systemic acid-base balance may play a role.

Clinical Manifestations and Diagnosis

The predominant clinical manifestation of erosive gastritis is gastrointestinal bleeding. Prodromal signs, such as abdominal pain, are infrequent. Although erosion formation can often be demonstrated within 24 h of the acute insult, massive bleeding usually occurs after 7 to 10 days when the superficial erosion extends into the larger submucosal vessels. Only rarely do these lesions perforate.

If a high index of suspicion is maintained in conditions that predispose to this illness, diagnosis is fairly straightforward. Routine upper gastrointestinal series are of little value: the critical condition of these patients often precludes a good-quality study, and the erosions are usually too superficial to visualize. Upper endoscopy is the procedure of choice and is diagnostic in more than 90 percent of patients. If the diagnosis remains in doubt, radionuclide scanning and visceral angiography may prove useful.

These patients are usually critically ill, and it is essential to stabilize and correct any predisposing conditions at the same time that diagnostic maneuvers are being performed. Hypovolemia and coagulopathy should be identified and treated as early as possible. Nasogastric intubation, preferably with a large-lumen Ewald tube, decompresses the stomach, eliminating the stimulating effects of distention and blood on acid secretion. In addition, it provides information about the rate of bleeding, clears the stomach for endoscopy, and allows saline lavage, which alone will control the bleeding in more than 80 percent of patients. Although suggested by some, the efficacy of lavage with iced solutions and levaterenol for their vasoconstrictor effects has not been demonstrated. Once the diagnosis is established, additional steps in medical and possibly surgical management should be initiated.

Treatment

If gastric evacuation and lavage are successful in halting the bleeding, antacid administration is useful in preventing the development of further hemorrhage. The pH of gastric contents should be checked regularly, probably every hour, and magnesium or aluminum antacids that are insoluble and remain in the stomach for prolonged periods should be instilled to keep the pH above 5. There is some evidence that antacids may be superior to H₂ blockers in the prevention of bleeding in critically ill patients. Omeprazole, the H⁺/K⁺-ATPase inhibitor, is as effective as antacids. Sucralfate, which binds to the ulcer bed and promotes healing, may also be utilized. The use of sucralfate has been associated with a decrease in pulmonary infections, presumably resulting from aspiration of gastric contents that become colonized when acid is neutralized with the antisecretory agents.

Intraarterial infusion of vasopressin controls hemorrhage in approximately 80 percent of patients. Any of the variety of endoscopic techniques for controlling upper gastrointestinal hemorrhage may also prove useful. Surgery should be considered when the bleeding exceeds 6 to 8 units over 48 h, but the underlying condition of the patient should enter into this decision. Mortality for surgery in this group of patients is in the range of 40 percent, and there is much controversy over the type of operation that offers the highest probability of success. No prospective trial has been performed, but some surgeons have advocated near-total gastrectomy in preference to the lesser procedures of vagotomy and pyloroplasty, with oversewing of the bleeding erosions, or partial gastrectomy with vagotomy. Generally, if the lesser procedures are successful, they are associated with a reduced morbidity and mortality but a greater incidence of rebleeding. Vagotomy is effective by reducing acid secretion and perhaps also by acutely decreasing mucosal blood flow. Oversewing a few bleeding erosions is often effective because, although there is usually a diffuse gastritis, only a few lesions have progressed into the deeper submucosal vessels, and it is these lesions that are responsible for the significant hemorrhage. Satisfactory results have been obtained with the use of vagotomy and pyloroplasty with oversewing of bleeding erosions as the initial procedure, reserving total resection for those who rebleed.

GASTRIC NEOPLASMS

Malignant Tumors

Gastric Cancer

Benign tumors of the stomach are rare. Over 90 percent of gastric tumors are malignant, and gastric adenocarcinoma (commonly referred to as gastric cancer) comprises 95 percent of the total number of gastric malignancies. Lymphoma (4 percent), leiomyosarcoma (1 percent), and other rare entities, e.g., squamous cell carcinoma, angiosarcoma, carcinosarcoma, and metastasis from adjacent or distant primary sites, constitute the balance.

Epidemiology

Gastric cancer is a biologically aggressive tumor that is often incurable when discovered in its symptomatic phase. Gastric cancer occurs worldwide, but its frequency varies greatly, and few cancers show as wide a variation in incidence as gastric cancer. The disease is rarely encountered in the United States. Malaysia, Chile, Iceland, and Japan have the highest incidence worldwide. The incidence of gastric cancer in Japan is greater than in the United States by a factor of eight and is the leading cause of cancer- related death in Japan.

In 1996 there were an estimated 22,000 new cases of gastric cancer in the United States and 14,000 deaths from the disease. The United States has experienced a rapid decline in gastric cancer-related deaths, from a rate of 30 per 100,000 in 1930 to 8 per 100,000 in 1996. The decline has been in antral tumors of the intestinal type, but the reason for this

favorable trend is not known.

Most populations show a 2:1 male to female ratio for gastric cancer. There is a higher incidence among blacks than whites in the United States. The peak incidence occurs in the sixth and seventh decades of life. There is a strong socioeconomic trend, with a higher frequency in lower socioeconomic groups. Although rates have decreased over the past 30 years, there has been an increase in the incidence of gastric cancer in the cardia. This is associated with a concomitant rise in the incidence of esophageal cancer.

Etiology

Diet has been the most studied risk factor for gastric cancer. Consumption of preserved, smoked, and cured foods results in a high concentration of nitrites, which form mutagenic compounds when exposed to bile acids in a stimulated gastric cavity. Colonization of the achlorhydric stomach by bacteria may also reduce dietary nitrates to nitrites and convert dietary amines in the presence of nitrite into carcinogenic *N*-nitroso compounds. A high-salt diet is a risk factor, and intake of fresh fruit and vegetables is protective; vitamin C and other antioxidants inhibit the conversion of nitrites into mutagenic compounds.

Substantial evidence has accumulated for an increased risk for gastric cancer with *H. pylori* infection, a relationship supported by numerous epidemiologic studies. The presence of immunoglobulin G antibodies to *H. pylori* in a given population is correlated with local gastric cancer incidence and mortality rates using linear regression analysis. There is a sixfold increased risk of gastric cancer in populations with 100 percent *H. pylori* infection compared with populations that have no infection (Table [24-3](#)).

Gastric ulcer disease is a marker for later development of gastric cancer with a relative increase in risk of 1.8. A contributing factor to both is atrophic gastritis induced by *H. pylori*. Other conditions associated with gastric cancer include cigarette smoking, previous partial gastrectomy (latency period of 15 years or more), radiation exposure, aflatoxin ingestion, family history, pernicious anemia, blood group A (relative risk 1.2 compared to blood group O), certain occupational exposures, and Epstein- Barr virus in the mucoid histologic subtype and in gastric remnant tumors. The number of partial gastrectomies being performed has fallen dramatically with the introduction of H₂- receptor antagonists and gastric proton pump inhibitors. The achlorhydric state produced by these agents used in the long term does not seem to have translated into an increased incidence of gastric cancer. It appears that duodenal ulcer disease actually is protective against gastric cancer.

Pathology

Gastric cancer is not a single entity but consists of several tumor types. Prognosis is more dependent on the depth of invasion through the stomach wall and the presence and extent of lymph node involvement than on any tumor classification.

Gastric cancers are divided into four subtypes by way of macroscopic appearance (Table [24-4](#)). Scirrhou cancer (linitis plastica) involves all layers of the stomach and there is a large connective tissue element, so much so that a histologic diagnosis can be obscured at endoscopic biopsy. The stomach is not distensible even when air is introduced at gastroscopy, and the condition is known as “leather bottle stomach.” Scirrhou cancer carries an extremely poor prognosis. Gross pathology is related to the histologic grade of the tumor. Scirrhou cancer is always undifferentiated, and polypoid tumors are likely to be well differentiated. There are multiple variations of the macroscopic appearances of gastric tumors between the stereotypic deep ulcerating tumor and the exophytic fungating papillary type. Lauren divided gastric cancer into two major histologic types: intestinal and diffuse. Intestinal types of gastric cancer arise on a field change of intestinal metaplasia (Fig. [24-18](#)). Intestinal metaplasia is the replacement of stomach epithelium by goblet and Paneth’s cells. Some believe that *H. pylori* initiates the stepwise sequence of superficial gastritis, atrophic gastritis, intestinal metaplasia, dysplasia, carcinoma in situ, and eventually invasive carcinoma. Only dysplasia is regarded as a positive predictor of gastric cancer. Severe dysplasia indicates imminent or concurrent gastric cancer and should be an indication for gastric resection. Most gastric tumors (85 percent) arise in a hypochlorhydric stomach. When intestinal metaplasia is found at the antrum, the risk of gastric cancer is proportional to the amount of metaplasia.

Intestinal type gastric cancer has been described as epidemic in type and predominates over the diffuse type in high-risk areas, and it is more closely associated with *H. pylori* infection than the diffuse type. The fall in incidence of gastric cancer in the United States has been due to a decrease in the intestinal type. The diffuse type of gastric cancer seems to be less related to environmental influences; it has increased in relative incidence with the falling incidence of the intestinal type, and it occurs more often in young people. The diffuse type is not associated with intestinal metaplasia, is not localized to the antrum, and arises out of single-cell mutations within normal gastric glands (Fig. [24-19](#)). The diffuse type, stage for stage, has a worse prognosis than the intestinal type. There are other classifications for gastric cancer based on histologic features such as papillary, tubular, mucinous, and signet-ring. The signet-ring cell type is caused by intracellular secretion of mucin and is associated with scirrhou gastric cancer. Gastric cancer has a great propensity to metastasize early by direct, lymphatic, hematologic, and transcoelomic routes.

The molecular events that occur in the progression of a gastric epithelial cell from benign to malignant are being investigated. Briefly, genetic instability, telomerase activity, and p53 mutation may be early events; epidermal growth factor overexpression and oncogene activation may occur later in the malignant process.

Clinical Manifestations

Anorexia with weight loss is the most common sign of gastric cancer, occurring in over 95 percent of those diagnosed with the disease. Patients are relatively asymptomatic until there is extensive involvement of the gastric wall and adjacent viscera, or widespread metastases. Massive hematemesis occurs in less than 5 percent of patients, although findings of anemia and occult blood in the stool are common. Nausea and vomiting may occur when distal lesions obstruct the pylorus. Dysphagia is a dominant symptom when cancer arises from the cardia of the stomach. Pain is a late and uncommon complaint. Abdominal tenderness is a rare finding, but a palpable abdominal mass is common (50 percent). Hepatomegaly may occur, suggesting metastatic spread. Peritoneal seeding may cause massive ascites, or involvement of the ovaries (Krukenberg’s tumor) or pelvic cul-de- sac (creating a Blumer’s shelf) by gravitational metastases. These late manifestations may lead to pelvic pain and constipation. A palpable lymph node in the left supraclavicular fossa (Virchow’s node) or a metastatic deposit to the umbilicus (Sister Joseph’s nodule) are also classic clinical signs of advanced gastric malignancy.

Diagnosis and Staging

While presenting complaints may be vague, a thorough history may provide indicators alerting the clinician to a possible diagnosis of gastric cancer. The most important specific investigation is upper gastrointestinal flexible endoscopy, permitting a biopsy procedure and tissue diagnosis. An upper gastrointestinal barium study (Fig. [24-20](#)) (single or air contrast technique) may be regarded as complementary, particularly if the scirrhou variety of tumor is suspected. Both techniques have a sensitivity and specificity above 90 percent. Upper abdominal computed tomography (CT) with the use of intravenous and oral contrast should then be used for the preoperative staging of gastric malignancy. TNM staging of gastric cancer is shown in Table [24-5](#), and preoperative CT staging in Table [24-6](#).

Endoscopic ultrasonography provides accurate information about the depth of penetration of tumor through the stomach wall. Laparoscopy is increasingly used as a staging tool to determine the presence of small intraperitoneal or liver metastases not seen on CT scan.

Early Gastric Cancer

Early gastric cancer is defined as gastric cancer confined to the mucosa or submucosa, regardless of lymph node involvement. This tumor stage was identified more than 30 years ago in Japan because of the inception of screening programs to detect gastric malignancy at an early stage. Detection of early gastric cancer ranges from 8 to 25 percent in the United States, and 35 to 50 percent of all gastric cancers in Japan because of the more effective screening programs. Early gastric cancer is divided into several types and subtypes (Table [24-7](#)) (Fig. [24-21](#)).

Seventy percent of early gastric cancers are well differentiated, and 30 percent are poorly differentiated; in 3 percent of early gastric cancers there is lymph node involvement. Early cancers of the exophytic type (I and IIA) are well differentiated, and the types with an ulcerating morphology are poorly differentiated. Five-year survival for patients when cancer is confined to the mucosa is 99 percent, when confined to the submucosa, 93 percent, and when lymph node involvement is present, survival drops precipitously to near 70 percent. The lymph drainage of the stomach has been detailed very thoroughly by the Japanese Research Society for Gastric Cancer, and an important concept is that the nodal staging of a tumor is dependent on the distance of the involved lymph nodes from the primary tumor. Systemic recurrence is strongly associated with lymph node involvement.

Treatment

Treatment of patients with gastric cancer is primarily surgical. Radiation and chemotherapy have little to offer in the way of palliation. Resection offers the only possibility of cure for gastric cancer and also provides the best palliation (Table [24-8](#)). The best palliation for patients with duodenal obstruction or bleeding is often resection, even if incomplete; the need for accurate staging is obviated by staging at laparotomy in many cases (Fig. [24-22](#)).

It is essential that a curative procedure for gastric cancer provides resection margins free of cancer. The most common operation performed for cure is a radical subtotal gastrectomy. This operation includes resection of the gastrocolic omentum and ligation and division of the right gastric, the right gastroepiploic, and the left gastric arteries at their origin, with removal of associated lymph nodes and mesentery. Approximately 2 cm of proximal duodenum is included in the resection; 50 to 85 percent of the stomach is removed, and continuity is reestablished using a gastrojejunostomy (Fig. [24-23](#)). Splenectomy and total gastrectomy may be required when the lesion is extensive, when it is in the proximal portion of the stomach, or when it directly invades within the hilum of the spleen. In the absence of a screening program, many patients present with an advanced tumor, and surgical therapy undertaken at this time is usually palliative with little hope of cure. The results of surgery for early disease are excellent.

There is a dynamic existing between extensive resection ensuring tumor clearance and the accompanying morbidity and mortality of extensive surgery. While some studies suggest that a more extensive resection of the stomach, nodes, and adjacent structures improves long-term outcome, most reports in Western countries note little additional benefit to more extensive operations. Gastrectomy is a major procedure with an overall operative mortality of 8 percent. Traditional subtotal gastrectomy includes removal of the perigastric lymph nodes within 3 cm of the stomach serosa (D1 resection). A D2 level resection involves removing gastric lymph nodes beyond this 3 cm margin; the nodes involved are found along the left gastric artery, celiac axis, common hepatic artery, and splenic artery and hilum. A D3 level resection involves removal of lymph nodes along the hepatoduodenal ligament, pancreatic head, and root of the small bowel mesentery. The specific operation performed depends on the location and extent of the gastric tumor, and the extent of lymphadenectomy varies accordingly. Splenectomy and distal pancreatectomy result in significant morbidity and should be avoided.

The precise role of adjuvant and neoadjuvant chemotherapy and radiotherapy has yet to be strictly defined in the treatment of gastric cancer. Because of the high probability of relapse after gastric resection, there is interest in defining an effective adjuvant regime. Single agents have no benefit as adjuvant therapy. Chemotherapeutic agents used in previous studies include 5-fluorouracil, methyl-lomustine (methyl-CCNU), Adriamycin, mitomycin C, and cytosine arabinoside. Radiotherapy combined with chemotherapy may yet emerge as an efficacious adjuvant treatment. The role of neoadjuvant (preoperative) therapy is less well established.

Lymphoma (Lymphosarcoma)

Gastric lymphoma can occur as an isolated neoplasm confined to the stomach or it may be the manifestation of widespread infiltrative disease involving lymphatic and other organ

systems. The lesion may present as a tumor mass or, more commonly, as a thickening of the rugal epithelial folds secondary to lymphocytic infiltration within the submucosa. Anorexia and weight loss are the most common presenting complaints. Early satiety is a prominent symptom as the gastric wall becomes thickened and the lumen is progressively compromised by the neoplastic infiltrate. Bleeding is uncommon. Definitive diagnosis is made by endoscopy and biopsy.

Bulky lesions with associated gastric outlet obstruction are best treated by subtotal gastric resection and postoperative irradiation. Radiation therapy alone, however, provides a long- standing remission that may be equal to that of resection alone in most cases. Radiation has emerged as the treatment of choice because of its low morbidity. A combined approach is associated with a 5-year survival rate of 85 percent when the malignant process is limited to the stomach. Involvement of the stomach by generalized lymphosarcoma is usually treated by combined radiation and chemotherapy. Gastrectomy is undertaken only when complications ensue such as perforation or when the stomach is the major source of disabling symptoms, e.g., obstruction.

Leiomyosarcoma

Leiomyosarcoma originates in the smooth muscle and is the least common of gastric malignancies. Frequently it grows to a very large size before detection because of its outward growth away from the gastric lumen. Distal spread is late, and even massive tumors that become adherent to the liver or pancreas can be resected, resulting in prolonged survival. Leiomyosarcomas are not usually responsive to radiation or chemotherapy. They generally are detected following a gastrointestinal hemorrhage from a breakdown of the overlying gastric epithelium or as a consequence of malnutrition secondary to compromise of gastric capacity and are often palpable on abdominal examination. Preoperative assessment is improved by abdominal CT scans done with intravenous and oral contrast. Resection is the preferred treatment. In order to excise the tumor completely, resection of the adjacent colon or spleen may be required. Extensive lymphadenectomy is not required because these tumors metastasize hematogenously to the liver and within the peritoneal cavity, but not usually via the lymphatics. Long-term survival is correlated inversely with tumor grade, size, and the status of histologic involvement of the margin.

Benign Tumors

Polyps

Polyps are the most common benign tumors of the stomach. There are two types: inflammatory and adenomatous. Adenomatous polyps are more important as they are true neoplasms and may have malignant potential. They can be distinguished from inflammatory polyps because of their long stalk and tendency to occur in the atrophic mucosa of patients with pernicious anemia. Occasionally adenomatous polyps will arise in the stomach in conjunction with the small bowel polyposis of the Peutz-Jeghers syndrome or the familial polyposis of Gardner's syndrome. Inflammatory polyps are usually sessile outgrowths within the antrum or fundus of the stomach. They are asymptomatic except when they are adjacent to, and prolapse through, the pylorus. Hypertrophic gastritis (Ménétrier's disease) may also be associated with multiple inflammatory polypoid lesions within the fundic area of the stomach. These lesions can be distinguished from multiple gastric adenomatous polyposis by biopsy and histologic examination. They do not require surgical extirpation. Gastric adenomatous polyps should be biopsied and excised by endoscopic ensnarement. Malignant polyps should be treated as gastric cancer by surgical resection.

Leiomyoma

The origin and differentiation of gastric stromal tumors has attracted much attention as immunohistochemical techniques have advanced. Gastric stromal tumors have been traditionally classified as smooth muscle tumors (leiomyomas). The origin of these tumors is usually from smooth muscle or epithelium. The correlation between histologic appearance and immunohistochemical markers is variable, making a definitive classification difficult. The majority of gastric stromal tumors are of smooth muscle origin. Gastric stromal tumors, regardless of type, have a similar natural history. Small leiomyomas are commonly found within the smooth muscle of the gastric wall at autopsy or during palpation of the stomach at laparotomy. They are of little clinical significance until they enlarge to greater than 4 cm in diameter. At this point they begin to compromise the blood supply to the overlying gastric epithelium. This leads to ulceration and proteolytic digestion of the core of the neoplasm that itself may have undergone central necrosis. This process may culminate in a massive upper gastrointestinal hemorrhage that may require emergency gastric resection for control. More typically patients present with anemia and guaiac-positive stools. Such lesions, when large, cannot be easily distinguished from their malignant counterparts and should be treated by gastrectomy with a minimum proximal margin of 4 cm. Smaller lesions (<4 cm) can be shelled out of the gastric wall or removed by wedge resection. Tumor size and histologic characteristics, including the number of mitoses per high-power field, capsular invasion, and pleomorphism are important characteristics distinguishing the benign and malignant nature of these lesions.

Lipoma

Lipomas of the stomach are asymptomatic submucosal lesions that are a radiographic curiosity distinguished by their smooth contour. Endoscopy reveals their submucosal position and often pale yellow color. They need not be biopsied or excised.

Ectopic Pancreas

Rarely, a portion of ectopic pancreas will reside within the antrum of the stomach. While this lesion is usually submucosal, it will often present within the gastric lumen as an umbilicated dimple. Excision may be necessary to exclude malignancy or when patients present with unremitting dyspeptic symptoms that are refractory to antiulcer therapy.

OTHER GASTRIC LESIONS

Hypertrophic Gastritis (Ménétrier's Disease)

Ménétrier's disease is a rare inflammatory disease of the gastric epithelium that is characterized by hypertrophic gastric folds within the proximal stomach. In advanced stages the epithelium assumes the appearance of large, multiple polypoid outgrowths. Histologic examination reveals that the thickened folds consist of a hypertrophy of the gastric glandular epithelium as well as a remarkable increase in size of the submucosa, which is edematous and contains a large number of small round cells. The latter finding suggests that the disease may have an autoimmune component. Ménétrier's disease is characterized clinically by the massive amount of plasma proteins that can be lost through an epithelium that is normally impermeable to large molecules.

Most cases of hypertrophic gastritis can be managed nonoperatively with treatment directed toward maintaining good nutrition and symptomatic relief of the vague gastric complaints offered by these patients. Hypoproteinemia rarely develops, but total gastrectomy may be required in such cases. The condition merits surveillance in view of the high incidence of gastric cancer reported in some series.

Bezoars

Bezoars are concretions of nondigestible matter that accumulate in the stomach. Trichobezoars, composed of hair, most commonly develop in young women who swallow their hair (trichophagy) (Fig. 24-24). Phytobezoars, composed of vegetable matter, are most often associated with persimmon ingestion. They may also develop after gastric surgery as a result of the loss of trituration. They are associated with symptoms of gastric outlet obstruction but may occasionally produce ulceration and bleeding. They can be diagnosed by upper gastrointestinal series. Treatment consists of proteolytic enzymes, such as papain, and can be facilitated by mechanical fragmentation with the endoscope. Occasionally operation is required for removal.

Dieulafoy's Lesion

Dieulafoy's lesion is a small, easily missed erosion of the gastric mucosa that is believed to involve a mucosal end artery that over time produces pressure necrosis of the stomach and ruptures into the lumen. Bleeding may be massive and recurrent. Endoscopy can miss the diagnosis if the lesion is not actively bleeding. Dieulafoy's lesion can be effectively treated by endoscopic thermal therapy or injection of a sclerosant, although occasionally the lesion will require operation and wedge resection.

Diverticula

True diverticula of the stomach, composed of all three layers of the gastric wall, are rare lesions, most often diagnosed between the ages of 20 and 60 years, although they are a congenital lesion. They are most commonly found in the posterior cardia and body of the stomach (Fig. 24-25) and are usually saccular. Complications are rare, but they may present with hemorrhage or diverticulitis. The majority produce no symptoms. Diagnosis is usually made by upper gastrointestinal contrast study. Operation, diverticulectomy, is reserved for cases in which there are complications, although it may be reasonable to excise a large diverticulum with a narrow neck and inspissated debris to prevent the development of diverticulitis.

Foreign Bodies

As a general rule, ingested foreign bodies that pass through the esophagus and gastroesophageal junction will also traverse the pylorus and pass into the distal gastrointestinal tract. Occasionally, objects such as safety pins will lodge in the stomach and produce symptoms of vague discomfort. Impacted larger objects can obstruct the stomach completely. Often these objects can be removed with a gastroscope and snare; operative removal is occasionally necessary.

Mallory-Weiss Tears

Mallory-Weiss tears are lacerations in the region of the gastroesophageal junction that can follow episodes of vomiting and retching. They are often associated with heavy alcohol consumption and may produce massive bleeding. They may account for as many as 15 percent of upper gastrointestinal bleeds. They stop spontaneously in 80 to 90 percent of patients. Diagnosis is most commonly made endoscopically, and the lacerations can be treated with thermal coagulation. Angiography with vasopressin infusion may also prove useful. Occasionally operation is necessary; usually oversewing the lesion through a long gastrotomy is successful, although care must be taken to identify any esophageal component.

Volvulus

Gastric volvulus, another unusual condition, is a torsion or twist of the stomach. It typically occurs along the long axis of the stomach (organoaxial volvulus) in association with a paraesophageal hernia so that the stomach appears, on contrast studies, to be rotated upside down. Because of this association, most recommend that even asymptomatic paraesophageal hernias be repaired. Gastric volvulus may occur acutely and present as either complete obstruction or strangulation of the stomach, but more commonly it is a chronic process. Acute volvulus typically presents as severe epigastric pain with inability to vomit, and it may be difficult or impossible to pass a nasogastric tube. Chronic conditions produce intermittent epigastric pain and distention. If the stomach is viable, the volvulus is reduced, any accompanying hernia repaired, and the stomach is fixed (gastropexy) to the anterior abdominal wall. A gastrostomy tube is often used to accomplish this. Necrosis may necessitate some form of gastric resection.

SURGERY OF THE STOMACH

The primary goal of surgery of the stomach is to eliminate the pathologic lesion while minimizing the consequent disruption of normal gastroduodenal physiology. Treatment must be individualized depending on the specific lesion and its mode of presentation. In the case of operations for acid-peptic disease, the common theme underlying these procedures is to restore the normal balance between secretion and mucosal resistance. It is impossible to surgically alter mechanisms of mucosal defense, and therefore all previous procedures

have approached these diseases by attempting to reduce acid secretion either by (1) sectioning the vagus, (2) eliminating the hormonal stimulation from the antrum, or (3) decreasing the number of parietal cells. Four general operations have been devised to produce these results: subtotal gastrectomy, vagotomy and drainage, vagotomy and antrectomy, and parietal cell vagotomy. When resection is performed, some form of reconstruction is required.

In contrast, in the case of neoplastic disease, the goals are somewhat different. The foremost consideration is to provide the patient with the best chance for cure, although this must be tempered by concern for the morbidity of more aggressive resection. In some patients, the malignancy is so far advanced that resection for cure is impossible. Most surgeons believe that some form of palliative resection, if it can be performed with acceptable morbidity, is preferable to leaving a patient with an obstructing or bleeding gastric tumor.

Vagotomy

The rationale for vagotomy is the elimination of the direct cholinergic stimulus to acid-pepsin secretion. Because of potentiation, this withdrawal also makes parietal cells less responsive to histamine and gastrin and abolishes the vagal stimulus for release of antral gastrin. Basal and stimulated acid secretion are reduced by 80 percent and 50 percent, respectively. Gastrin levels increase, probably as a result of the reduction in normal acid inhibition of antral G cells.

Three different types of vagotomy are performed: truncal, selective, and highly selective (Fig. 24-26). Truncal vagotomy is the simplest procedure; the anterior and posterior vagus nerves are identified on the distal esophagus, and a small section is removed and sent for tissue analysis to confirm that the nerve has been identified. The second type, selective vagotomy, was devised to prevent some of the undesirable sequelae of truncal vagotomy by preserving the innervation of the rest of the gut. The nerves of Latarjet are sectioned just distal to the hepatic and celiac branches of the vagus. Although the rationale seems logical, this operation is not any better tolerated than truncal vagotomy, and because it is more difficult, it has largely been abandoned.

Truncal vagotomy markedly alters gastric motility patterns. Both receptive relaxation and trituration are impaired. Gastric emptying of liquids is speeded with the impairment in relaxation, whereas solid evacuation is slowed significantly, leading to overt gastric stasis in some patients. As a result, some form of emptying procedure (i.e., pyloroplasty or gastrojejunostomy) should be performed (Fig. 24-27). At least three different types of pyloroplasty have been described; the Heineke-Mikulicz pyloroplasty is probably the simplest technically and is most commonly performed. Although it has never been demonstrated to be clinically superior, pyloroplasty preserves the normal anatomic relationship between the stomach and the duodenum and for that reason is generally preferred. Gastroenterostomy is typically reserved for patients with significant duodenal bulb scarring, which would make pyloroplasty more difficult. These drainage operations also have important physiologic consequences. Pyloroplasty further speeds gastric emptying and, by reducing the period of contact between gastric contents and the antral mucosa, decreases acid secretion mediated through the direct gastric phase of stimulation. Gastroenterostomy may enhance secretion by producing antral gastrin release and eliminating the secretion of inhibitory factors normally generated during the passage of chyme through the duodenum. An increase in gastrin levels may result because of stasis in the partially excluded antrum and because the gastroenterostomy allows acid to empty and be neutralized before it can bathe the antrum. Both procedures eliminate the pyloric sphincter and produce duodenogastric reflux.

The rationale for this highly selective vagotomy is to eliminate vagal stimulation to the acid-secreting portion of the stomach without interrupting motor innervation to the antrum and pylorus. The operation involves severing all branches of the nerves of Latarjet along the lesser curvature that innervate the corpus and fundus. Basal and stimulated acid secretion are reduced by more than 75 percent and 50 percent, respectively. It reduces the secretion of acid in response to gastric distention. Basal serum gastrin is increased, although the response to a meal is reduced. Receptive relaxation is impaired, and emptying of liquids is more rapid than normal; antral peristalsis, trituration, and sphincter function are preserved. The emptying of solids is normal. The normal small bowel innervation preserves intestinal motility patterns and reduces the incidence of postvagotomy complications and dumping.

Resection

Some form of resection is always required for curative treatment of malignant disease. Resections usually take longer and the morbidity and mortality are somewhat higher than those of the various vagotomies. Total gastrectomy is seldom performed for benign disease. It is usually reserved for malignancies of the proximal and midstomach and typically includes omentectomy and some form of lymph node dissection. The procedure completely eliminates the reservoir and digestive functions of the stomach. Although most patients lose considerable weight initially, with smaller, more frequent meals, most stabilize and can survive long term. Vitamin B₁₂ injections are required and deficiencies of iron and calcium can develop.

Subtotal gastrectomy was employed for the treatment of duodenal ulcer disease in the past, but is most commonly used for gastric ulcer and distal gastric malignancies. Typically the distal two-thirds of the stomach, including the pylorus, is removed. No matter what the nature of the lesion, proximal gastrectomy virtually always produces problems with emptying and should not be performed. Distal gastrectomy does reduce acid and pepsin secretion by removing not only a major portion of the parietal and chief cells in the corpus but also the antral G cells, eliminating the gastrin stimulus to secretion. The proximal stomach also atrophies in the absence of gastrin. This operation reduces basal and stimulated secretion by 75 percent and 50 percent, respectively, and eliminates the antral-pyloric mixing mechanism by which food is reduced to chyme before entering the duodenum. The emptying of both liquids and solids is more rapid. With removal of the pylorus, increased reflux of intestinal contents into the stomach may produce significant reflux gastritis.

Until recently, antrectomy combined with truncal vagotomy was one of the most commonly performed procedures for duodenal ulcer disease. This procedure seems to combine some of the advantages of both these operations. The simultaneous effects of vagotomy and antrectomy remove both the cholinergic and gastrin stimulus to acid secretion. Basal acid secretion is virtually abolished, and stimulated acid secretion is reduced by nearly 80 percent. The rationale for this procedure is to preserve some of the reservoir function eliminated by a more subtotal gastrectomy and still reduce acid secretion. Dumping complications, resulting from rapid emptying of hyperosmolar gastric contents into the duodenum, seem to be slightly less frequent than after more complete resection. Improved medical therapy and the availability of highly selective vagotomy have virtually eliminated the need for this procedure.

Reconstruction

After gastrectomy, gastrointestinal continuity may be restored by anastomosing the remnant to the duodenum, i.e., gastroduodenostomy (Billroth I), or, after closing the duodenal stump, anastomosing the remnant to the intestine just distal to the ligament of Treitz, i.e., gastroenterostomy (Billroth II) (Fig. 24-28). The latter has both an afferent limb from the duodenum and an efferent limb extending distally. The choice between the two drainage procedures has important physiologic consequences that should be considered in addition to their respective complications. The Billroth II procedure eliminates the added stimulus to acid secretion from duodenal release of gastrin in response to passing chyme. It also prevents the meal from reaching pH receptors and osmoreceptors in the proximal duodenum, which normally slow gastric emptying. Unlike reconstruction after a subtotal resection where the Billroth II procedure seems preferable in terms of suture line tension and recurrence rates, experience suggests that restoration of normal continuity with a Billroth I procedure is more desirable after vagotomy and antrectomy.

After total gastrectomy, because of the high incidence of alkaline reflux esophagitis, if the esophagus is anastomosed in continuity to the small bowel, a Roux-en-Y reconstruction is usually performed (Fig. 24-29). This interposes a segment of intestine, usually 45 cm in length, between the esophagus and the intestine in continuity with the duodenum and excludes the esophagus from exposure to bile and pancreatic secretion. Attempts to preserve reservoir function by constructing pouches with the small bowel have not been successful; they do not empty appropriately and may limit intake more than the absence of a stomach.

Choice of Operation

The choice among surgical procedures can be complex. The most important single factor is the nature of the disease process itself. Other considerations include the respective operative mortality and postoperative morbidity, the incidence of recurrence, the postgastrectomy side effects of the procedure, the long-term metabolic consequences of the operation (weight loss, bone disease, and anemia), and the potential future risk of gastric carcinoma. The surgeon's own familiarity with the various operative procedures is a crucial consideration.

Reliable data on the results of various operations have only been generated during the past quarter century. Published series in general used different criteria for patient selection and the incidence of side effects. Table 24-9 summarizes the data on the three most commonly performed procedures for duodenal ulcer. Mortality is lowest for highly selective vagotomy and highest for vagotomy with antrectomy. The relative incidence of side effects is about the same for vagotomy with antrectomy or drainage, and lower for the highly selective operations. The incidence of recurrence is significantly lower for vagotomy with antrectomy. Truncal vagotomy with drainage has both the higher recurrence rate of highly selective vagotomy and the unfavorable incidence of side effects of vagotomy with antrectomy. There is still a role for truncal vagotomy and drainage in the emergent treatment of a patient for whom time is a critical factor; this procedure can be performed considerably more quickly than antral resection or highly selective vagotomy. Some surgeons believe that highly selective vagotomy is the only acceptable operation for duodenal ulcer, although resection is still indicated for gastric ulcer. The lumen of the gastrointestinal tract is not entered, reducing the risk of septic complications. Because the normal functional anatomy of the stomach and pylorus are preserved and the distal vagal innervation is not interrupted, the potential for serious long-term side effects is greatly reduced.

The data in Table 24-9 regarding recurrence are probably no longer applicable, because they were generated before the role of *H. pylori* was recognized. This is a time when surgical approaches to peptic ulcer are being completely revised. Elective operation for peptic ulcer is unusual; most patients undergo operation for one of the complications of ulcer and are often elderly and debilitated. In this setting, it is unclear whether anything beyond dealing with the complication is necessary.

Laparoscopic approaches to the stomach, usually for benign disease, have received increasing attention. Although experience with resection has been limited, it is possible to remove portions of the stomach with laparoscopic staplers, and reconstructive techniques are being developed. Laparoscopic vagotomy also is feasible. Conventional highly selective vagotomy is a demanding procedure to perform through the laparoscope. At laparoscopy, surgeons have employed a combination of posterior truncal vagotomy and either anterior seromyotomy (dividing the muscular layers from the gastroesophageal junction to the antrum, which should interrupt the innervation of the fundus and the corpus) or anterior highly selective vagotomy. At open operation these techniques have results comparable to highly selective vagotomy; the results with the laparoscope are still preliminary, and the utility of these minimally invasive procedures cannot yet be judged accurately. Most of these operations have been performed in the elective setting; with eradication of *H. pylori*, few such procedures will be necessary in the future. Serious questions remain about the applicability of laparoscopic techniques in the emergent setting. Neither hemorrhage nor obstruction seem well suited to the laparoscope. It may be most applicable to the treatment of perforations, where patching and highly selective vagotomy seem most appropriate.

Postgastrectomy Syndromes

After operations on the stomach, a variety of chronic undesirable sequelae may develop. Although some of these conditions are related more to vagotomy than they are to resection, they have been referred to as the postgastrectomy syndromes (Table 24-10). Virtually all patients note a change in their digestive habits postoperatively, and about 20 percent are significantly affected. Most are able to adapt with time; only about 5 percent develop lifelong symptoms, and 1 percent are significantly disabled by these syndromes.

These patients are probably the only candidates for remedial operation.

Dumping occurs in about 20 percent of patients after gastrectomy or vagotomy with drainage. Although the exact mechanism has eluded definition, it appears to be related to the rapid emptying of hyperosmolar chyme, particularly carbohydrate, into the intestine. This draws fluid into the intestine and probably releases one or more vasoactive hormones, such as serotonin and vasoactive intestinal polypeptide (VIP). This is associated with epigastric distention, cramps, nausea, vomiting, dizziness, flushing, and palpitations. In most patients the symptoms tend to resolve as the patients learn to avoid foods that aggravate the problem. Frequent small meals, low in carbohydrate, might eliminate the problem. In the most refractory cases, synthetic somatostatin analog (octreotide) may be of benefit.

After truncal vagotomy, approximately 30 percent of patients develop *postvagotomy diarrhea*. The pathogenesis of postvagotomy diarrhea is unclear, but it may be related to the rapid passage of unconjugated bile salts from the denervated biliary tree into the colon, where they can stimulate secretion. Most cases are self-limiting. Long-term oral cholestyramine, which binds the bile salts, is an effective treatment of this complication.

After operations that effect the pyloric sphincter, reflux of bile into the stomach is common. Only about 2 percent of patients develop *alkaline reflux gastritis*, a syndrome of persistent, burning epigastric pain and chronic nausea that are aggravated by meals. Diagnosis is one of exclusion, although endoscopy may reveal gastritis, and technetium biliary scan can demonstrate increased reflux of bile into the stomach. A variety of medical therapies has been reported; none have proved particularly effective, although there is enthusiasm for the use of ursodeoxycholic acid. Conversion of the original drainage operation to a Roux-en-Y anastomosis, diverting the bile 45 to 60 cm from the gastric remnant, has proved effective in selected patients.

Early satiety may develop as a result of postsurgical atony, gastric stasis, as a result of denervation, or from the "small remnant syndrome" related to extensive gastric resection. In both cases, symptoms consist of epigastric fullness with meals, often followed by emesis. Atony can be distinguished with a solid food emptying test and may respond to prokinetic agents that stimulate gastric motility, such as metoclopramide, erythromycin, and cisapride; if these fail, completion gastrectomy is the procedure of choice. The small remnant syndrome usually will improve with time; frequent small feedings are recommended.

Afferent loop syndrome develops after Billroth II reconstruction or gastroenterostomy. It is related to mechanical obstruction of the afferent limb by kinking, anastomotic narrowing, or adhesion and typically is associated with postprandial epigastric pain and nonbilious vomiting that is then relieved by projectile bilious vomiting. The detection of a distended afferent loop on CT scan is diagnostic. Conversion to a Roux-en-Y anastomosis is necessary to treat this problem. *Efferent loop obstruction* is associated with epigastric pain, distention, and bilious vomiting; the cause is always mechanical, and operation is the treatment of choice.

Bezoars, obstructing concretions of undissolved food, develop after gastrectomy as a result of the alterations in gastric motility.

Gastric stump carcinoma develops in approximately 3 percent of patients who have undergone gastrectomy. There has been some controversy about this and it appears to represent a significant increase over the incidence in the general population. Symptoms are usually vague; endoscopic screening has been recommended for this group of patients, although the optimal timing remains unclear. The treatment is resection.

Operations for Morbid Obesity

A variety of criteria have been used for defining morbid obesity; perhaps the most commonly accepted definition is 100 lbs over ideal body weight. The pathogenesis of such obesity is poorly understood, but includes a genetic influence, metabolic abnormalities, disorders of satiety, and psychological abnormalities. Morbidly obese individuals typically develop problems by puberty. In many people, it is related to excessive intake; in others, the problem may be related to a low basal energy expenditure or an inability to burn off excess calories.

Severe obesity is associated with a variety of complications and with excess mortality. Complications include coronary artery disease, hypertension, impaired cardiac function, adult-onset diabetes mellitus, pulmonary dysfunction including hypoventilation and sleep apnea, hypercoagulability, degenerative arthritis, cholelithiasis, and gastroesophageal reflux. Low self-esteem and depression are not unusual.

Criteria for operation vary, but typically complete metabolic and psychological evaluation is necessary; most surgeons require documented evidence of participation in a supervised dietary management program without success. Although 100 lb over ideal weight is required, exceptions may be made for patients with a weight-related medical problem.

Surgery for morbid obesity has undergone considerable evolution. The first operation was the jejunal bypass, in which the proximal jejunum, usually 20 cm distal to the ligament of Treitz, was anastomosed to the terminal ileum, usually 10 cm from the ileocecal valve. Although this produced significant weight loss, it was associated with complications, including significant liver disease in 5 to 10 percent of patients, related both to protein-calorie malnutrition and to factors released by the bacterial overgrowth that invariably develops in the bypassed segment. Arthritis, cholelithiasis, nephrolithiasis, and metabolic abnormalities were common. This procedure has been largely abandoned.

The most commonly performed procedures today are the vertical banded gastroplasty and the Roux-en-Y gastric bypass (Fig. 24-30). Both operations are performed using a combination of surgical staplers to create a proximal gastric pouch with a capacity of approximately 30 mL. This significantly limits intake, producing weight loss. The patients are maintained on a 1000 kcal or less diet with nutritional supplements. The bypass does not allow the meal to enter the distal stomach, adding an element of maldigestion. There may also be a component of dumping syndrome as the gastric contents directly enter the small bowel. Several studies have suggested that the bypass may be superior, and it is becoming the procedure of choice. There now are several surgeons performing the procedure laparoscopically.

Morbidly obese patients are at particular risk for postoperative morbidity. Atelectasis and pulmonary infection are most common. In these patients postoperative mobilization may be particularly difficult, and venous thromboembolism can be a major problem. Wound infections are more common and more difficult to diagnose in the morbidly obese. Intraabdominal complications, such as anastomotic leak, can be difficult to detect.

Weight reduction with these procedures is significant. The average loss is one-half to two-thirds of the excess weight at 1.5 years, at which point weight typically stabilizes. This has shown to be associated with improvements in diabetes, hypertension, mobility, pulmonary problems, and arthritis. In addition, patient self-image is often markedly improved. The operations have not clearly demonstrated a reduction in the excess mortality of the morbidly obese, though it seems likely with longer follow-up.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 25 - Small Intestine

B. Mark Evers
Courtney M. Townsend, Jr.
James C. Thompson

Introduction

Considered teleologically, the small bowel is the *raison d'être* for the entire gut. The esophagus brings food to the stomach, which prepares it for digestion. The exocrine secretions of the liver and pancreas make digestion possible. Digestion is achieved in the lumen of the small bowel, and nutrients are absorbed through the small-bowel mucosa; the colon disposes of whatever is left.

In addition to its vital function in nutrition, the small bowel has other important roles. It is the largest endocrine organ in the body. It has tremendous defenses against infection and is one of the most important, if not the most important, organs in immune defense. It is a marvel of efficiency and works so well that, excepting the proximal 3 cm of the duodenum, it is seldom afflicted with disease. We are supplied with a great excess of small bowel and can exist on less than one-half of the absorptive surface provided.

Until recently, the small intestine had been relatively inaccessible for nonoperative diagnostic procedures, compared to the stomach or colon. Several diagnostic techniques are now available for specific diseases of the small bowel. Mucosal biopsy specimens obtained with the peroral biopsy capsule are often diagnostic in diffuse mucosal diseases. Enteroclysis is a more sensitive radiographic technique than the conventional barium follow-through examination of the small bowel. Selective mesenteric angiography often is helpful in cases of discrete lesions with abnormal vascular patterns, such as neoplasms, vascular malformations, or actively bleeding lesions. Scintigraphy may be helpful in localizing sites of bleeding. Fiberoptic endoscopy of the duodenum, proximal jejunum, and distal ileum are routinely employed.

Some diseases that affect the small intestine are discussed in other chapters: intestinal obstruction (Chap 22), mesenteric vascular disease (Chap 33), diseases of the intestine in infancy and childhood (Chap 37), the intestine in trauma (Chap 6), and duodenal and gastrojejunal peptic ulcer (Chap 24).

Embryology

The primitive gut forms during the fourth week of fetal development. The endoderm gives rise to most of the epithelium and glands of the digestive tract. The splanchnic mesoderm surrounding the endoderm gives rise to muscular connective tissue and other layers comprising the wall of the digestive tract. Early in the fourth week the duodenum begins to develop from the caudal portion of the foregut and the cranial portion of the midgut. The remainder of the small bowel is derived from the midgut. During the fifth week, the fetal gut rapidly elongates and herniates through the umbilicus (Fig. 25-1). This midgut loop has both cranial and caudal limbs; the cephalic limb develops into the distal duodenum to proximal ileum, and the caudal limb becomes the distal ileum to proximal two-thirds of the transverse colon. The juncture of the cephalic and caudal limbs is where the vitelline duct joins to the yolk sac. This duct normally becomes obliterated before birth; however, it occasionally persists as a Meckel's diverticulum. This period of midgut herniation lasts until approximately 10 weeks' gestation, after which the intestines return to the abdomen. During this period of rapid elongation and herniation the midgut rotates 90 degrees counterclockwise around an axis formed by the superior mesenteric artery. As the gut reenters the abdomen at about the tenth week, it undergoes a further 180-degree rotation, thus completing 270 degrees of rotation from the starting point. The proximal jejunum is the first portion to reenter the abdomen and occupies the left side of the abdomen, with subsequent loops lying more to the right. The cecum enters last and is located temporarily in the right upper quadrant; during the third to fifth month of fetal gestation, it descends to its normal position in the right lower quadrant. Congenital anomalies of gut malrotation and fixation can occur during this process.

The small intestine is lined with simple cuboidal epithelium during the sixth and seventh weeks of fetal gestation. The rapid epithelial proliferation may lead to occlusion of the lumen, especially in the proximal gut (particularly in the duodenum). The lumen is reestablished during weeks 9 and 10 with the formation of coalescing vacuoles that restore the patency of the lumen. Villi begin to form first in the proximal intestine at 9 weeks, proceeding caudad and eventually lining the entire gut. Differentiated crypt cells appear between the ninth and twentieth weeks of gestation. Crypt formation begins in the tenth to twelfth weeks of gestation.

This complex and regimented pattern of gut differentiation appears to be mediated by sets of genes that are characteristically turned on and off with temporal and spatial specificity during developmental transitions as the gut evolves into its adult form. This developmental pattern also appears to be dependent on interaction with the mesoderm and lamina propria, which may provide "instructive" signals for developmental progression. The epithelial cells of the gut are generated from a fixed stem-cell population localized in the lower portion of the crypts that give rise to four main cell types: absorptive enterocytes, goblet cells, enteroendocrine cells, and Paneth cells. Until recently the analysis of the factors regulating these complex developmental processes in vivo was not possible; however, elegant studies by Gordon and colleagues, using transgenic mouse techniques, have described marker systems for inferring the biologic properties of intestinal stem cells and provided direct evidence of the transcriptional regulation of spatial patterns of intestinal gene expression. Other investigators have used the intestine-specific sucrase-isomaltase and neurotensin genes as molecular models to better understand the factors regulating gut development and eventual maturation.

Anatomy

The most impressive thing about the small bowel is its immense mucosal surface area, which is responsible for the organ's tremendously efficient digestion of food. Several layers of muscle, combined with actin and myosin components in the microstructures, provide great motility, so that not only is there great surface area, but the interface between the surface and the luminal contents presented for absorption is in constant motion as well.

Gross Anatomy

The small bowel extends from the pylorus to the cecum. The length of the small intestine depends entirely on the state of bowel activity at the time of measurement. Careful estimates provide a duodenal length of 20 cm, a jejunal length of 100 to 110 cm, and an ileal length of 150 to 160 cm. The jejunoileum extends from the peritoneal fold that supports the duodenal-jejunal junction (the ligament of Treitz) downward to the ileocecal valve. The jejunoileum is estimated to constitute 60 percent of the entire length of the gut and to be approximately 160 percent of the body height, so that the small bowel is considerably longer in a 7-ft basketball player than in a 5-ft jockey.

Generally the jejunum occupies the upper abdomen, especially on the left, and is in contact with the pancreas, spleen, colon, and left kidney and adrenal gland. Affliction of these organs may affect the jejunum; pancreatitis, for example, may cause local ileus (the "sentinel loop") of the jejunum.

The jejunum has a larger circumference and is thicker than the ileum, and it may be identified at operation because of this and also because the mesenteric vessels usually form only one or two arcades and send out long, straight vasa recta to the mesenteric border of the jejunum. In contrast, the blood supply to the ileum may have four or five separate arcades; the vasa recta are shorter, and, most important, there is usually much more fat in the mesentery of the ileum than in that of the jejunum (Fig. 25-2). The jejunal mesentery may be transparent, but mesenteric fat usually reaches all the way to the bowel in the ileum. The ileum occupies the lower abdomen, especially on the right, and the pelvis. It is smaller in diameter and somewhat more mobile.

Except for the duodenum the small bowel is entirely covered with visceral peritoneum (the serosa) and is tethered only by its attachment to the mesentery, through which course arteries, veins, and lymphatics. The mesentery is obliquely attached to the posterior body wall, beginning superiorly well to the left of the second lumbar vertebra and ending obliquely downward and to the right to overlie the right sacroiliac joint. The mesentery normally is covered with glistening, nonadherent peritoneum, but after trauma (external, chemical, septic, or operative) it may become adherent to other surfaces (mesenteric, visceral, or parietal) and greatly limit bowel mobility.

Except for the proximal duodenum, which is supplied by branches of the celiac axis, the blood supply of the small bowel is entirely from the superior mesenteric artery, which is the second major branch of the infradiaphragmatic aorta. The superior mesenteric artery also supplies the appendix, cecum, and ascending and proximal transverse colon. There is an abundant collateral blood supply to the small bowel provided by the vascular arcades in the mesentery. In spite of this collateral supply, occlusion of a major branch of the superior mesenteric artery, or of the superior mesenteric artery itself, will lead to bowel death if not quickly corrected. Venous drainage of the segments of the small bowel is in parallel with the arterial supply. The superior mesenteric vein joins the splenic vein behind the neck of the pancreas to form the portal vein. The blood leaving the gut, with its relatively high oxygen content, provides a significant portion of the oxygen supply to the liver.

If the mesentery is not greatly infiltrated by fat and if there are no peritoneal adhesions, the bowel is extraordinarily mobile on its vascular tether. In some individuals jejunal segments may be sufficiently mobilized to allow anastomosis in the neck to replace the cervical esophagus.

The small bowel contains major deposits of lymphatic tissue, particularly in the Peyer's patches of the ileum. There is a rich lymphatic drainage of the entire small bowel, and this plays a major role in fat absorption. Lymphatic drainage proceeds from the mucosa through the wall of the bowel to a set of nodes adjacent to the bowel in the mesentery. Drainage continues to a group of regional nodes adjacent to the mesenteric arterial arcades and then to a group at the base of the superior mesenteric vessels. From there, lymph goes to the cisterna chyli and thence up the thoracic ducts to empty into the venous system in the neck. The lymphatics of the gut play a major role in immune defense and also in the spread of cells arising from neoplasms in the gut.

The small-bowel mucosa is characterized by transverse folds (plicae circulares, or valves or folds of Kerckring), though these are absent in the duodenal bulb and in the distal ileum. They are more prominent in the distal duodenum and the jejunum, where they may reach 1 cm in height and form interlocking transverse ridges (see Fig. 25-2). The small-bowel mucosa has a pink, velvety appearance with a glistening surface. It usually is thicker in the jejunum than in the ileum, where there may be no folds and the surface may be entirely smooth, except for small scattered lymphatic nodules.

The innervation of the small bowel comes from both sympathetic and parasympathetic systems. Parasympathetic fibers come from the vagus and traverse the celiac ganglia. They affect secretion and motility and probably all phases of bowel activity. Vagal afferent fibers are present but apparently do not carry pain impulses. The sympathetic fibers come from the three sets of splanchnic nerves and have their ganglion cells usually in a plexus around the base of the superior mesenteric artery. Their motor impulses affect blood vessel motility and probably gut secretion and motility. Pain from the intestine is mediated through general visceral afferent fibers in the sympathetic system.

Histology

The wall of the small bowel has four layers—the serosa, the muscularis, the submucosa, and the mucosa (Fig. 25-3).

Serosa

The serosa, the outermost layer, consists of visceral peritoneum that encircles the jejunum but covers the duodenum only anteriorly. It consists of a single layer of flattened mesothelial cells.

Muscularis

The muscularis consists of a thin outer longitudinal layer and a thicker inner circular layer of smooth muscle. Specialized gaps in the muscle-cell membranes permit cell-to-cell communication, which facilitates the ability of the muscle layer to function as an electrical syncytium. Ganglion cells from the myenteric (Auerbach's) plexus are interposed between the two muscle layers and send fibers into both layers.

Submucosa

The submucosa is a layer of fibroelastic connective tissue containing blood vessels and nerves. It is the strongest component of the bowel wall and therefore must be included when sutures are placed through the bowel. It contains elaborate networks of lymphatics, arterioles, and venules and an extensive plexus of nerve fibers and ganglion cells (Meissner's plexus). Although frequently subdivided, the nerves from the mucosa, submucosa, and muscle layers are interconnected by small nerve fibers, and cross-connections between adrenergic and cholinergic elements have been described.

Mucosa

Looked on as a device to increase absorptive surface, the small-bowel mucosa is an architectural marvel. The gross transverse folds, the fingerlike villi protruding into the lumen of the bowel, the microvilli (brush border) covering the cells, and the glycocalyx fuzz covering the microvilli each tremendously increase the surface area exposed to luminal contents. Villi protrude 0.5 to 1 mm into the lumen; they are tallest in the distal duodenum and the proximal jejunum and become progressively shorter toward the terminal ileum.

The mucosa can be divided into three layers—the muscularis mucosae, the lamina propria, and the epithelium. The deepest of these, the muscularis mucosae, is a thin sheet of muscle separating mucosa from submucosa. The lamina propria is a continuous layer connective tissue between the epithelium and the muscularis mucosae. It extends into the villi and around the pitlike crypts of Lieberkühn (Fig. 25-4). The lamina propria contains, additionally, a variety of cells—plasma cells, lymphocytes, mast cells, eosinophils, macrophages, fibroblasts, smooth muscle cells—and noncellular connective tissue. The lamina propria is the architectural base on which the epithelium lies, but it also has important functions of its own and apparently serves protectively to combat microorganisms that penetrate the overlying epithelium. The plasma cells are an active site of synthesis of immunoglobulins. In addition, cells in the lamina propria release various mediators (for example, cytokines, arachidonic acid metabolites, and histamine) that can modulate various cellular functions of the overlying epithelium.

The innermost mucosal layer is a continual sheet, one layer thick, of epithelial cells covering the villi and lining the crypts of Lieberkühn (see Fig. 25-4). The main functions of the crypt epithelium are cell renewal and exocrine, endocrine, water, and ion secretion; the main function of the villous epithelium is digestion and absorption. The crypts contain at least four distinct cell types: (1) goblet cells, which secrete mucus; (2) enteroendocrine cells (often called enterochromaffin or argentaffin cells in the older literature), of which there are more than ten distinct populations that produce the gastrointestinal hormones, including gastrin, secretin, cholecystokinin, somatostatin, enteroglucagon, motilin, neurotensin, and gastric inhibitory peptide; (3) Paneth cells, which secrete lysozyme, tumor necrosis factor, and the cryptidins, homologues of leukocyte defensins, and whose function is thought to be related to the host mucosal defense system; and (4) undifferentiated epithelial cells, whose function is to provide for cell renewal. The epithelium of the small intestine is a rapidly proliferating tissue in which old cells are discarded into the lumen and are replaced by newly formed cells that appear to march up from the crypt into the villus in orderly sequence. This trip takes 5 to 7 days in the proximal small bowel, but in the ileum labeled cells may travel from crypt to villous tip in 3 days.

The role of growth factors (e.g., epidermal growth factor, transforming growth factors α and β , insulinlike growth factor, and growth hormone), certain amino acids (glutamine), polyamines, prostaglandins, and gastrointestinal peptides (neurotensin, peptide YY, and glucagonlike peptide-2) in gut epithelial maturation has been the object of extensive recent investigations. All have trophic effects on the small-bowel mucosa; the precise roles, interactions, and importance have not yet been determined.

The epithelium covering the villi consists of scattered endocrine cells, goblet cells, and absorptive cells. The major known functions of the villi are digestion and absorption. These functions are carried out by the absorptive cells, tall columnar cells resting on a thin basement membrane that separates them from the lamina propria. Their luminal surface is covered by microvilli that rest on a terminal web (Fig. 25-5). The microvillar projections multiply the cell surface area exposed to the lumen by 30 times. The microvilli in turn are covered by a fuzzy coat of glycoprotein, the glycocalyx (Figs. 25-6 and 25-7). The microvilli participate actively in absorption and digestion. They contain enzymes for digestion of disaccharides and peptides, and certain cells may contain specific receptors that facilitate absorption (for example, certain ileal cells have receptors for vitamin B₁₂ on their microvilli).

The plasma (or cell) membranes of the epithelial cells consist of three layers and are thicker over the microvilli than over the lateral and basal portions of the cell (see Fig. 25-6). The lateral portion of the plasma membrane also is specialized. There are tight junctions between epithelial cells that prevent communication between intercellular spaces and the lumen. Immediately underneath the tight junction is a narrow space called an intermediate junction, and beneath that is a desmosome, which provides tight attachments of adjacent membrane by binding adjacent cells together. The depths of the tight junctions are greater between adjacent absorptive cells than between adjacent undifferentiated crypt cells. The permeability of the barrier between the intestinal lumen and the space between cells may vary from one location to another.

The cytoplasm immediately beneath the microvilli consists of fine filaments known as the terminal web. This interconnects with filaments forming the core of the microvilli, which contain actin (see Fig. 25-6). Myosin also may be present at the base of the microvilli, and these contractile proteins may allow movement and contraction of the microvilli.

The processes of digestion and absorption within the epithelial cells are carried out by specific organelles (see Fig. 25-5). The mitochondria participate in intracellular oxidation and provide energy for metabolism. The lysosomal sacs contain cytotoxic substances and intracellular waste products. The endoplasmic reticulum is the main synthesizing element within the cell and appears to be responsible for at least two major processes in fat absorption, the resynthesis of triglycerides from absorbed fatty acid fragments, and the synthesis of the lipoprotein coat for chylomicrons. It is also the major synthetic site for intracellular digestive enzymes. The Golgi apparatus segregates, stores, and chemically modifies material that is absorbed and synthesized by the cell.

It may be instructive to review the path of a food element absorbed from the intestinal lumen. Initial contact is with the glycocalyx coating the microvilli, where some digestion may occur. Products of this digestion may go through the microvillous membrane, traversing the terminal web into the cytoplasm. The absorbed material may then move either laterally into the intracellular spaces or into the channels of the endoplasmic reticulum, where it may be biochemically modified and transmitted to the Golgi material, where it may be stored. Eventually the material leaves the cell by crossing either the lateral or the basal plasma (cell) membrane. It penetrates the basal lamina to enter the lamina propria, where it traverses the lymphatic or capillary endothelial cells to gain access to lymph or blood.

Physiology

Motility

Food is propelled through the small bowel by a complex series of muscular contractions. Motility patterns in the small bowel vary greatly between the fed and the fasting states. Pacemaker potentials, probably originating in the duodenum, initiate a series of contractions that propel food through the small bowel. These contractions are of two types—segmentation and peristalsis. Contractions of the circular muscle divide the bowel into segments that are moved to and fro over the column of bowel contents for a short distance. The contents of adjacent segments then combine, and the process is repeated. About 40 percent of contractions are segmental. The circular muscle also initiates peristalsis, circular contractions migrating in an aboral direction, propelling intestinal contents onward. The peristaltic reflex may function independently of extrinsic nerves. Abnormal waves of powerful contractions (*peristaltic rushes*) may rapidly traverse the entire segment of small intestine during episodes of enteritis.

During the interdigestive (fasting) period between meals, the bowel is regularly swept by a series of contractions initiated by the migrating myoelectric complex (MMC). The MMC is under neural and humoral control and initiates a triphasic series of contractions; phase I is resting; in phase II intermittent contractions of moderate amplitude occur; and phase III is the activity front that consists of a brief series of high-pressure waves. The MMC in human beings is stimulated by some (but not all) fluctuating increases in serum concentrations of motilin.

Small-bowel motility is modulated by neural and humoral influences. Extrinsic nerves to the small bowel are vagal and sympathetic. Vagal fibers have two functionally different effects; one is cholinergic and excitatory, and the other is peptidergic and probably inhibitory. Sympathetic fibers from the splanchnics appear chiefly to modulate the activity of intrinsic nerves.

Although gut peptides clearly influence bowel motility, their physiologic function is uncertain, except for the role of motilin in initiating MMC activity. Gastrin, cholecystokinin, and motilin are known to stimulate muscle contraction, whereas muscle activity is inhibited by secretin and often by glucagon (Table 25-1). Cholecystokinin may be physiologically important, since ingestion of a fatty meal may stimulate peristaltic contraction.

Digestion and Absorption

Liters of water and hundreds of grams of food move across the intestinal mucosa from the lumen to the bloodstream each day. The process is remarkably efficient; nearly all food is absorbed unless protected by indigestible cellulose. There is no apparent governor; food absorption is just as efficient in the corpulent as in the starving.

Fat

Most individuals in Western Europe and North America consume 60 to 100 g of fat per day in the form of triglycerides. Fat digestion and absorption occur in the small intestine, where triglycerides are partially hydrolyzed by pancreatic lipase, which splits off the two exposed fatty acids to leave a single central fatty acid still combined with glycerol (beta monoglyceride) plus two fatty acids (Fig. 25-8). The monoglyceride and the fatty acids are poorly soluble in water but combine with bile salts to form micelles. A mixed micelle is composed of bile salts, fatty acids, and the beta monoglyceride, and may also include phospholipids, cholesterol, and fat-soluble vitamins. The micelle must traverse three diffusion barriers in the passive process of entry into the intestinal epithelial cell. These are the unstirred water layer, the mucous coat overlying the brush border, and the lipid bilayer membrane making up the brush border. The micelle may release its fatty acid and monoglyceride component in traversing these barriers. After disaggregation of the micelle, bile salts remain within the intestinal lumen to enter into the formation of other micelles, and the released fatty acids and monoglycerides traverse the plasma membrane into the epithelial cell. The major metabolic pathway within the cell is initiated by reformation of the triglyceride through the interactions of intracellular enzymes (see Fig. 25-8) that are associated with the endoplasmic reticulum.

The triglycerides then combine with cholesterol, phospholipids, and apoproteins to form chylomicrons, which consist of an inner core containing almost all triglycerides with a membranous outer coat of phospholipids and apoproteins. The chylomicron exits the cell from the basolateral region and preferentially enters the central lacteal of the villus, where it

moves to the thoracic duct. Small fatty acids with chain lengths of C₁₀ and less may move directly through the cell into capillaries to flow into the portal vein. The bulk of chylomicron assimilation from the intestinal cell is via lymphatics, but some direct transfer to the portal vein may take place, particularly during periods between meals.

In one of the examples of selective sites of resorption, bile salts are resorbed into the enterohepatic circulation from the distal ileum (Table 25- 2). The total bile salt pool in human beings is about 5 g, and it recirculates about six times every 24 h (the enterohepatic circulation of bile salts). Only about 0.5 g is lost in the stool every day, and this is replaced by resynthesis from cholesterol. All ingested fat is usually absorbed; the small amount appearing in the stool comes from sloughed cells and bacteria.

Protein

Protein digestion is initiated in the stomach, where gastric acid denatures proteins and proteolysis is initiated by activated pepsin. Little actual digestion takes place, however, until protein enters the duodenum and upper jejunum, where they come under the influence of pancreatic proteases. Pancreatic trypsinogen is activated by the duodenal mucosal enzyme enterokinase, and then activated trypsin further activates all proteases. Endopeptidases (trypsin, elastase, chymotrypsin) act on peptide bonds at the interior of the protein molecule, producing peptides that are substrates for exopeptidases (carboxypeptidases), which serially remove a single amino acid at a time from the carboxyl terminal end of the peptide. The final products are amino acids and peptides of two to six amino acid residues. The intraluminal action of pancreatic proteases is efficient and yields 70 percent short-chain peptides and 30 percent amino acids. Short peptides are optimal substrates for the peptide transport mechanism that brings dipeptides and tripeptides into the cell. Of peptides that are assimilated intact, at least 90 percent are hydrolyzed to free amino acids by cytosol peptidases before delivery to the portal venous system.

Carbohydrate

Western men and women take in about 400 g of carbohydrates a day, 60 percent as starch, 30 percent as sucrose, and 10 percent as lactose. Carbohydrates constitute about half the calories ingested in our society, but in underdeveloped countries they typically provide a much higher proportion.

Starch is a polysaccharide consisting of long chains of glucose molecules. Amylose makes up 20 percent of starch in the diet and has an alpha glucose-to-glucose bridge. Amylopectin (80 percent of dietary starch) has branching points every 25 molecules along the straight glucose chains. Both have an alpha₁₋₄ glucose-linked chain. Alpha amylase attacks the alpha₁₋₄ linkage and converts amylose to maltotriose and maltose. Amylase converts amylopectin to shorter dextrans. Intraluminal digestion of starch in the duodenum is rapid, because of the huge amounts of pancreatic amylase, and digestion often is complete by the time the chyme enters the jejunum.

The enzyme responsible for final surface digestion is concentrated in the brush border of the luminal surface (Fig. 25-9). After dietary carbohydrate is reduced to monosaccharides by surface digestion, transport of the released hexoses (glucose, galactose, or fructose) is carried out by a specific process. Glucose and galactose are actively transported across the intestinal membrane, whereas fructose is absorbed by facilitated diffusion. The rate-limiting phenomenon for most carbohydrate absorption occurs during transport through the intestinal cell, but luminal hydrolysis of lactose is slower than its transport capacity, so surface hydrolysis of lactose is rate-limiting.

Water and Electrolytes

In addition to ingested water, salivary, gastric, biliary, pancreatic, and intestinal fluids contribute up to 8 to 10 L of water per day, and all but about 0.5 L per day is resorbed proximal to the ileocecal valve. The small bowel secretes and absorbs huge amounts of water. Net absorption is the algebraic sum of two fluxes going in opposite directions. Water may simply diffuse in or out of the cell or may be drawn through by osmotic or hydrostatic pressures. The osmotic pressures result from active transport of sodium, glucose, or amino acids into cells. Diffusion occurs through pores in plasma cell membranes. Jejunal pores are larger (7 to 9 Å) than those in the ileum (3 to 4 Å). Hypertonic solutions in the duodenum and upper jejunum are rapidly equilibrated to isotonicity by the influx of large amounts of water.

Sodium and chloride are absorbed from the small bowel by active transport, by coupling to organic solutes, and by cotransport on carriers of neutral sodium chloride. A small portion of sodium absorption in the jejunum is by active transport, but the bulk is by coupling to organic solutes. In the ileum sodium is absorbed against steep gradients and is not stimulated by glucose, galactose, or bicarbonate. Bicarbonate is absorbed by a sodium-hydrogen exchange, so that one bicarbonate ion is released into interstitial fluid for every hydrogen ion secreted. Calcium is absorbed by active transport, particularly in the duodenum and jejunum. Absorption appears to be facilitated by an acid environment and is enhanced by vitamin D and parathormone. Potassium appears to be absorbed by passive diffusion.

Endocrine Function

The mucosa of the small bowel is the primary source of regulatory peptides of the gut, and the muscle wall of the small bowel is rich in peptidergic nerves containing neuroendocrine peptides. Although we often call these agents hormones, they do not always function in a truly endocrine fashion; that is, the active peptides are not always discharged into blood vessels to act on some distant site. Sometimes they are discharged and act locally in a paracrine fashion (for example, bombesin [gastrin-releasing peptide] and somatostatin), or they may serve as neurotransmitters, or they may be discharged into blood vessels after nerve stimulation in a true neuroendocrine manner. We will briefly describe some of these agents.

Secretin

The discovery of secretin in 1902 gave birth to the entire field of endocrinology. Secretin is a 27–amino-acid, helical peptide that is present in specialized cells in the small-bowel mucosa and is released by acidification or by contact with bile and perhaps fat. It acts to stimulate release of water and bicarbonate from pancreatic ductal cells, and when this combination flows into the duodenal lumen the bicarbonate neutralizes gastric acid. The amount of pancreatic bicarbonate released after a meal closely approximates the amount of acid secreted by the stomach. Secretin also acts to stimulate the flow of bile and to inhibit gastrin release, gastric acid secretin, and gastrointestinal motility. Secretin has the unique ability to release gastrin from gastrinomas, and intravenous secretin is used as a diagnostic test in patients in whom the Zollinger-Ellison syndrome is suspected.

Cholecystokinin

Cholecystokinin (CCK) is released from small-bowel mucosa by contact with certain amino acids (especially tryptophan and phenylalanine) and medium- to long-chain fatty acids. In addition, the concentration of intraluminal trypsin and bile acids inversely regulates CCK release. CCK has two major actions: it stimulates contractions of the gallbladder and relaxation of the sphincter of Oddi, and it stimulates the secretion of enzymes by pancreatic acinar cells. CCK also stimulates growth of bowel mucosa and pancreas, stimulates bowel motility, and releases insulin. CCK exists in multiple molecular forms (CCK-8, CCK-33, and CCK-39, among others), and the larger forms contain the smaller ones. CCK and gastrin share the identical C-terminal tetrapeptide (Trp-Met-Asp-Phe-NH₂), which explains many of the similarities in their actions.

Other Peptides

Largely through the efforts of Viktor Mutt and colleagues at the Karolinska Institute, several active agents have been isolated from the mucosa of the small bowel that greatly influence physiologic activities of the gut. Of these, only *gastric inhibitory polypeptide* (GIP) has satisfied rigid criteria for hormonal status, but others will be discussed as well. GIP is a 43–amino- acid peptide member of the secretin-glucagon family that is released by fat. Although it was initially studied for its properties of inhibition of gastric secretion, later studies showed that it was a prime incretion candidate, because it greatly stimulated insulin release when levels of glucose were elevated. Glucose-stimulated release of intestinal GIP apparently solves the conundrum posed by the fact that oral ingestion of a fixed amount of glucose releases more insulin than intravenous administration of the same amount. A family of peptides reacting with glucagon antibodies is present in small-bowel mucosa, and it has been given a variety of names. *Enteroglucagon* is a term used to designate all such gut peptides. It inhibits bowel motility. Recent findings suggest that glucagonlike peptide-2 (GLP-2), a 33–amino-acid peptide, is a potent enterotrophic factor. *Vasoactive intestinal peptide* (VIP) is a 28–amino-acid basic peptide of the secretin-glucagon family that appears to function chiefly as a neuropeptide. VIP is a potent vasodilator, and it stimulates pancreatic and intestinal secretion and inhibits gastric acid secretion. It is the chief agent in the watery diarrhea syndrome caused by pancreatic endocrine tumors. *Motilin* is a 22–amino-acid peptide widely distributed through the gut. It causes contraction of intestinal smooth muscle, including the gallbladder, and appears to be involved in the interdigestive pattern of gut motility. *Bombesin* is a 14–amino-acid peptide that was first isolated from frog skin. It has the capacity to release all gut peptides except secretin. Its mammalian equivalent, gastrin-releasing peptide, is present in small-bowel mucosa, where it probably serves as an “on” switch, regulating release of gastrointestinal hormones. *Somatostatin* is a 14–amino-acid peptide that was first isolated from the brain; it is also widely distributed in the gut, where it probably functions in a paracrine fashion as an “off” switch to inhibit gut hormone release and gastrointestinal secretion. *Neurotensin*, a 13–amino-acid peptide produced in large part by endocrine cells in the ileal mucosa, stimulates water and HCO₃⁻ secretion from the pancreas, inhibits gastric secretion, and has trophic effects for small and large bowel mucosa. *Peptide YY* (PYY) is a 36–amino-acid peptide in the distal ileum and colon. It inhibits gastric and pancreatic secretion, has no effect on gallbladder motility, and exhibits trophic effects on small-bowel mucosa. PYY is released by perfusion of the colon with fat and may be involved in the physiologic inhibition of pancreatic secretion.

Immune Function

We ingest thousands of bacteria, parasites, and viruses every day. Only a few of these are pathogenic, but the huge surface of the small-bowel mucosa represents a massive potential portal of entry. An important component of bowel defense is the secretory immune system, which produces a special group of antibodies that resist bacterial proliferation, neutralize viruses, and minimize the penetration of enterotoxins.

The small bowel is a major source of immunoglobulin A (IgA). Cells of the lamina propria of the small intestine contain plasma cells that produce IgA. The population of cells producing IgA (the secretory immunoglobulin) is ten times greater than that producing IgG (the antibody mediating general humoral immunity).

Antigens from the intestinal lumen crossing the mucosal barrier contact M cells overlying lymphoid nodules. These cells are specialized for uptake and transport of antigen, which they convey to the underlying lymphoblasts that produce IgA. After interacting with the antigen, the lymphoblasts migrate to the regional lymph nodes, from which they enter the systemic circulation. They then are returned to the intestine, where they are widely distributed in the lamina propria (Fig. 25-10). In the lamina propria, they differentiate into plasma cells that produce a specific IgA antibody directed to the absorbed antigen. This IgA antibody traverses the epithelial cell to the lumen by means of a protein carrier (the secretory component) that not only transports the IgA but also protects it against the intracellular lysosomes. The antibodies at the free surface of the cell collect in the glycocalyx, where they are in a strategic position to combat new antigens, preventing their attachment to the cell membrane and achieving immune exclusion.

Passage of viable bacteria from the intact gastrointestinal tract to the mesenteric lymph nodes and beyond, which has been termed *bacterial translocation*, has been implicated as a possible mechanism to explain septic complications and multiple organ system failure in burn and trauma patients. Small-intestinal mucosal integrity and function are profoundly affected by systemic injury. Cutaneous burn injury and sepsis cause early mucosal atrophy and subsequent defect in the gut mucosal barrier. This allows escape from local gut immune surveillance. One possible therapeutic strategy to prevent bacterial translocation and its systemic complications may involve stimulation of gut mucosal growth and maintenance of mucosal integrity and function during the early stages after systemic injury. Numerous gut peptides (e.g., gastrin, neurotensin, bombesin, epidermal growth factor) have trophic effects on small-intestinal mucosa; these growth factors may prove to be important in the development of therapeutic strategies for prevention of bacterial translocation. The magnitude of this problem, however, remains to be completely determined. In addition, the effects of polyamines on gastrointestinal growth suggest that these small polycationic substances play a central role of the regulation of intestinal cell growth and differentiation. The activity of ornithine decarboxylase (ODC), the rate-limiting enzyme

in polyamine biosynthesis, is increased dramatically in instances associated with mucosal growth and repair in the duodenum and in the remaining small bowel after injury.

The exact mechanisms that maintain the epithelial integrity of the gastrointestinal tract remain largely unknown. Recent studies suggest that members of the trefoil peptide family, proteins secreted throughout the small intestine and colon in response to injury, are important in stimulating mucosal repair, since transgenic mice lacking intestinal trefoil factor demonstrated poor epithelial regeneration and mucosal healing after injury. Moreover, mice overexpressing the trefoil peptide pS2 have an increased resistance to intestinal damage.

Gut-associated lymphoid tissue (Fig. [25-11 A](#)) consists of the cells of the mucosal immune system located between the external environment (intestinal lumen) and the internal milieu of the host. Follicle-associated epithelial cells are located over Peyer's patch areas (in animals) or lymphoid follicles (in human beings), in which immature germinal center cells and B cells reside. Antigens from the intestinal lumen are transported through specialized epithelial cells called M cells. After being brought across the follicle-associated epithelium, the antigens come into contact with lymphocytes and macrophages. A subcompartment of T suppressor/cytotoxic lymphocytes called *intraepithelial lymphocytes* resides between absorptive epithelial cells. In animals most intraepithelial lymphocytes bear the gamma/delta T cell receptor. Within the lamina propria, mature T cells, B cells, plasma cells, and macrophages carry out cell-mediated immunologic processes, including cytokine secretion, production of IgA, and phagocytosis. Two different types of mast cells (mucosal mast cells and connective tissue mast cells) are found in the mucosa and submucosa.

Antigens within the intestinal lumen (Fig. [25-11 B](#)) are transported across specialized epithelial cells (M cells) into contact with underlying lymphoid cells. Antigen-induced activation of macrophages leads to secretion of interleukin-1 and activation of helper T cells, which secrete interleukin-2. Activation of intraepithelial lymphocytes and subsequent production of γ -interferon leads to enhanced production of major histocompatibility complex (MHC) class II molecules by epithelial cells. Increased numbers of MHC class II molecules on intestinal epithelial cells may lead to enhanced antigen presentation by intestinal epithelial cells, with subsequent augmentation of antigen-specific IgA antibody production. The expansion of IgM-bearing B cells is regulated by switch factors that favor the development of IgA-bearing B cells. T cells secreting interleukin-4, interleukin-5, and interleukin-6 induce the differentiation and proliferation of B cells for subsequent maturation into plasma cells, which synthesize and secrete dimeric IgA with J chain. The dimeric IgA with J chain then interacts with secretory component receptor on the basal membrane of epithelial cells for subsequent secretion into the intestinal lumen.

Inflammatory Diseases

Crohn's Disease

Background

Crohn's disease is a chronic granulomatous disease of the alimentary tract of unknown cause that is characterized by intermittent periods of remission and exacerbation. Although earlier reports described the gross and microscopic pathologic features of this disease, it was not until the paper by Crohn, Ginsburg, and Openheimer in 1932 that the disease process was clearly and completely described. In the introduction to their landmark paper, they stated: "We proposed to describe, with pathologic and clinical details, a disease of the terminal ileum affecting mainly young adults, characterized by a subacute or chronic necrotizing and cicatrizing inflammation. The ulceration of the mucosa is accompanied by a disproportionate connective tissue reaction of the remaining walls of the involved intestine, a process which frequently leads to stenosis of the lumen of the intestine, associated with the formation of multiple fistulas."

Many different terms have been used to describe this disease process, such as regional ileitis, transmural enteritis, and granulomatous enteritis, but because of its multiple clinical appearances and because the disease is not confined to the terminal ileum, Crohn's disease has been universally accepted as its name.

Incidence and Epidemiology

Crohn's disease is the most common surgical disease of the small bowel. The annual incidence of Crohn's disease is approximately 3 to 6 cases per 100,000 of the general population, with the incidence highest in North America and northern Europe. The incidence is higher in patients in the third decade of life; however, a bimodal distribution is apparent, with a second, smaller peak in the sixth decade of life. Crohn's disease is more common in urban dwellers, and there may be a slight female predominance; some studies report an incidence approximately 20 percent higher in females, though others report a similar incidence in males and females. The risk of developing Crohn's disease is about two times higher in smokers than nonsmokers. Although Crohn's is uncommon in African blacks, blacks in the United States have rates similar to whites, suggesting that environmental influences play a role in the disease. There is a familial association, with the risk of developing Crohn's disease increased 30-fold in siblings of patients with the disease and 13-fold for all first-degree relatives.

Etiology

The cause of Crohn's disease remains unknown. Several possibilities have been suggested, including infectious, immunologic, environmental, and genetic causes.

Although controversial, proposed infectious causes that have received more than passing interest include mycobacterial infections, particularly *Mycobacterium paratuberculosis* and *M. avium*. The existence of atypical mycobacteria as a cause for Crohn's was proposed by Dalziel in 1913. Subsequent studies using the polymerase chain reaction (PCR) technique have confirmed the presence of these mycobacteria in intestinal samples of patients with Crohn's disease. Furthermore, the transplantation of intestinal samples from patients with Crohn's disease into other species has resulted in an ileitis. Koch's postulates, however, have yet to be rigorously satisfied since empirical therapy with antimicrobial agents in well-controlled clinical trials failed to ameliorate the disease process. In addition to mycobacteria, other potential sources include persistent measles virus and *Yersinia enterocolitica*.

Immunologic abnormalities that have been identified in patients with Crohn's disease have included humoral as well as cell-mediated immune reactions directed against gut cells, suggesting autoimmunity. There has been no direct correlation between these abnormalities, immunologic reactivity, and the development of Crohn's disease; therefore, these findings may represent an epiphenomenon. Other etiologic factors such as environmental and dietary factors, smoking, and psychosomatic causes have been suggested but never proved.

Genetic factors appear to have a role in the etiology of Crohn's disease, since the single strongest risk factor, by far, is having a relative with the disease. Recent European and American studies have identified a putative susceptibility locus in the centromeric region of chromosome 16 in multiply affected family members, further suggesting a genetic role in Crohn's disease. The advent of molecular biologic techniques has greatly enhanced our understanding of this disease; however, the development of an animal model (e.g., knockout mice) that simulates the findings of Crohn's disease in humans would be a great help to our understanding of this disease.

Pathology

Crohn's disease is an inflammatory disorder that is transmural in nature and associated with varying degrees of fibrosis. It can affect any portion of the gastrointestinal tract. The terminal ileum is involved in approximately 90 percent of patients with Crohn's disease; however, only about one-third of patients have small-bowel involvement alone.

Microscopic Features

The histologic features of Crohn's disease vary, but some occur with sufficient consistency to create a pattern that, though not specific, is at least characteristic of the disease. Some understanding of the pathogenesis may be gained by observing the progression of changes from early to late phases of regional enteritis.

Intense mucosal and submucosal edema may be seen microscopically before any gross changes are apparent. The earliest gross pathologic lesion is a superficial aphthous ulcer. As the disease progresses, the ulceration becomes more pronounced, and complete transmural inflammation results (Fig. [25-12 A](#)). The inflammatory reaction is characterized by extensive edema, hyperemia, lymphangiectasia, an intense infiltration of mononuclear cells and hyperplasia of lymphoid follicles. The ulcers are characteristically linear and may coalesce to produce transverse clefts and sinuses that result in the characteristic cobblestone appearance of the mucosa. Thickening and hypertrophy of the submucosa and the muscularis result in narrowing of the gut lumen; obliterative lymphangitis occurs; the bowel wall becomes thickened and edematous and quite rigid. Granulomas appear later and are found in the bowel wall and in regional lymph nodes in 60 to 70 percent of patients. These are noncaseating granulomas with Langhans' giant cells (Fig. [25-12 B](#)).

Macroscopic Features

At operation, one observes thickened, grayish pink or dull purple-red loops of bowel with areas of thick, gray-white exudate or fibrosis of the serosa. Areas of diseased bowel separated by areas of grossly normal bowel, called skip areas, are commonly encountered. A striking finding of Crohn's disease at operation is extensive fat-wrapping, caused by circumferential growth of the mesenteric fat around the wall of the bowel. The thickened bowel wall is firm, rubbery, and virtually incompressible (Fig. [25-13](#)). The uninvolved proximal bowel often is dilated as a result of the considerable degree of obstruction in the diseased segment. Involved segments often are adherent to adjacent loops or other viscera, or several loops may be matted together into a bulky conglomerate mass. Internal fistulas are common in such adherent areas. The mesentery of the segment is characteristically greatly thickened, dull, and rubbery and contains masses of lymph nodes up to 3 or 4 cm in diameter.

Clinical Manifestations

Although Crohn's disease can occur at any age, the typical patient is a young adult, and the majority present before the age of 40 years. The onset of disease is often insidious, with a slow, protracted course. There are symptomatic periods of abdominal pain and diarrhea interspersed with varying intervals of remission; but over time, symptomatic periods gradually become more frequent, more severe, and longer lasting. Pain (the most common symptom) usually is intermittent and cramping early in the course of the disease but may develop into persistent, dull, aching abdominal pain, most prominent in the lower abdomen. Diarrhea is the next most frequent symptom and is present, at least intermittently, in 85 percent of patients. The frequency of stools is less than in patients with ulcerative colitis, numbering two to five daily, and the stools rarely contain the mucus, pus, or blood seen in ulcerative colitis. Fever is present in about one-third of the patients, and moderate weight loss, loss of strength, and malaise, in over one-half. Frank nutritional disorders and steatorrhea are uncommon before surgical treatment.

The main intestinal complications of Crohn's disease include obstruction and perforation. The obstruction occurs as a result of chronic disease with fibrosis and progressive narrowing of the bowel lumen and can be partial or complete. Perforations can occur into the free peritoneal cavity and present as a generalized peritonitis. More commonly, however, development of fistulas and intraabdominal abscesses occurs. Fistulas can develop between the diseased bowel and any adjacent organ, such as small or large bowel, urinary bladder, vagina, or stomach, and through the skin, usually at the site of a previous laparotomy. Patients with Crohn's colitis may develop toxic megacolon, demonstrated by marked colonic dilatation, abdominal tenderness, fever, and leukocytosis. A potential intestinal complication of long-standing Crohn's disease is cancer of the affected segment of intestine, which occurs with an incidence six times greater than that of the general population. Cancers in patients with Crohn's disease tend to be more advanced when identified and, therefore, have a worse prognosis.

Perianal disease (fissure, fistula, stricture, or abscess) is common, occurring in 25 percent of patients with small-intestinal disease, 41 percent of patients with ileocolitis, and 48 percent of patients with exclusively colonic involvement. Perianal disease is the sole presenting feature in 5 percent of patients and may precede the onset of intestinal disease by

months or years. Crohn's disease should be suspected in any patient with multiple, chronic, recurrent perianal fistulas. Extraintestinal manifestations of Crohn's disease occur in about 30 percent of patients and include arthritis and arthralgia, uveitis and iritis, hepatitis and pericholangitis, and erythema nodosum and pyoderma gangrenosum.

Diagnosis

Patients with chronic, recurring episodes (often of long duration) of abdominal pain, diarrhea, and weight loss should be evaluated for Crohn's disease with endoscopy or radiographic studies.

When the colon is involved, sigmoidoscopy or colonoscopy reveals characteristic aphthous ulcers with granularity and a normal-appearing surrounding mucosa. With more progressive and severe disease, the ulcerations involve more and more of the bowel lumen and may be difficult to distinguish from ulcerative colitis. The presence of discrete ulcers and cobblestoning, however, particularly when discontinuous segments of bowel are involved, favors the diagnosis of Crohn's disease.

Radiographic studies of the small bowel reveal a number of findings, including a nodular contour, diffuse narrowing of the lumen, sinuses, clefts, linear ulcers, separation of bowel loops, and asymmetrical involvement of the bowel wall. A cobblestone appearance of the mucosa, composed of linear ulcers, transverse sinuses, and clefts, may be apparent. Long lengths of narrowed terminal ileum (Kantor's string sign) may be present in long-standing disease (Fig. 25-14). Fistulas between adjacent organs may be apparent (Fig. 25-15). Computed tomography may be useful to resolve diagnostic problems by ascertaining the nature of mass effects such as abscesses, separation, or displacement of segments of small intestine (Fig. 25-16).

Medical Treatment

The medical treatment of Crohn's disease is largely symptomatic and empiric and usually is followed by remissions and exacerbations until a complication that requires surgical intervention supervenes. Goals of medical therapy are relieving abdominal pain, controlling diarrhea, treating infection, and correcting any nutritional deficiencies.

Intraluminal antibiotics, particularly sulfasalazine (Azulfidine), have significant beneficial effects in certain patients. Corticosteroids, such as prednisone or the newer agent budesonide, are also useful in effecting remission of symptoms during acute exacerbations of ileitis, and a combination of sulfasalazine and steroids may be used to maintain patients for a short period after resolution of an acute inflammatory episode. Long-term maintenance therapy, with either agent alone or in combination, however, has not been shown to be of benefit in preventing recurrence of the disease. Although there have been implications of immunologic abnormalities in Crohn's disease, immunosuppressive treatment with azathioprine, 6-mercaptopurine, and cyclosporine have not been shown conclusively to be advantageous in comparison with placebo. Other immune modulators that have been tested in clinical trials with varying success include interleukin-1 receptor antagonists, monoclonal antibodies directed against helper T cells, and a diphtheria toxin–fused protein that is cytotoxic to cells that have receptors for interleukin-2. Therapies reported to be of possible benefit in early clinical trials include zileuton, a potent and specific inhibitor of the enzyme lipoxygenase that metabolizes arachidonic acid to leukotrienes, and superoxide dismutase, which blocks the toxicity of oxygen radicals; more clinical data are required before these agents can be advocated. Systemic antibiotics are valuable in the management of infectious complications but have no effect on the primary disease process and should not be considered therapeutic.

Chemically defined elemental diets and total parenteral nutrition have been touted as beneficial, even curative, in patients with acute Crohn's disease and in patients with complications such as enterocutaneous fistulas. No specific long-term benefit with any of the dietary measures has been shown; however, complete abstinence of oral intake with total parenteral nutrition may lead to temporary remission of symptoms in some patients. The provision of adequate nutrition is much more important in preparing patients for operation than it is for cure.

Surgical Treatment

Indications

Although medical management is certainly indicated during acute exacerbations of disease, the majority of patients with chronic Crohn's disease will require operation sometime during the course of their illness. In patients with more than 20 years of disease, the National Cooperative Crohn's Disease Study (NCCDS) has reported that the cumulative probability of operation was 78 percent. Indications for operation are limited to complications, including obstruction, abscess, fistula, free perforation, urologic complications, hemorrhage, cancer, and perianal disease. Children with Crohn's disease and resulting systemic symptoms such as growth retardation may benefit from resection. The extraintestinal complications of Crohn's disease, though not a primary indication for operation, often subside after resection of involved bowel.

General Tenets

Cure of patients with Crohn's disease is not possible by either medical or surgical therapy; therefore, operative intervention in patients with Crohn's disease should be directed to treating complications, and only the segment of bowel involved in the complicating process should be resected. Even if adjacent areas of bowel are clearly diseased, they should be ignored. Early in the history of surgical treatment of Crohn's disease, many surgeons tended toward wide resection in an attempt to effect cure. Repeated wide resections lead to the short bowel syndrome, a devastating surgical complication. In the past, the most common cause of the short bowel syndrome was repeated resections for Crohn's disease. During the past two decades or more most surgeons have accepted the view that it is not possible to surgically eradicate this disease. We have, for example, given up frozen-section study of the resected margin because we generally limit our resection to areas that are grossly involved adjacent to sites of obstruction or fistulization. We ignore microscopic evidence of involvement.

Management of Specific Problems

Acute Ileitis

There is a subgroup of patients with acute abdominal signs and symptoms in whom the usual preoperative diagnosis is acute appendicitis. At operation, the appendix is found to be normal, but the terminal ileum is edematous and beefy red, having a thickened mesentery with enlarged lymph nodes. The condition is acute ileitis, a self-limited disease that does not lead to subsequent development of Crohn's disease. The precise cause of acute ileitis is unknown; however, some patients have been found to have *Yersinia* or *Campylobacter* infections.

The role of appendectomy in these circumstances is less controversial now than it was in the past. In the absence of acute inflammatory involvement of the appendix or the cecum, appendectomy should be performed. This eliminates the possibility of acute appendicitis in the differential diagnosis of any subsequent abdominal complaints in these patients. In the few patients who have developed enterocutaneous fistulas after appendectomy, the fistulas originated from the ileum in all cases.

Obstruction

Intestinal obstruction is the most frequent indication for surgical therapy in most large series of patients with Crohn's disease. Obstruction in these patients often is partial, and nonoperative management is indicated initially. Operative intervention is required in instances of complete obstruction and in patients with partial obstruction that does not resolve with nonoperative management. The optimum treatment of complications of Crohn's disease is surgical resection of the involved segment with restoration of intestinal continuity (Fig. 25-17). As previously noted, the extent of resection is dictated by the extent of the complication. A short area of profound Crohn's ileitis producing partial obstruction (not involving the ileum immediately adjacent to the ileocecal valve) should be treated by local resection and primary anastomosis. It is common for the cecum and ascending colon to be involved contiguously with the terminal ileum in the disease process. In this case, the diseased bowel is removed and the ileum anastomosed to the transverse colon. No attempt should be made to dissect lymph nodes up to the origin of the ileocolic artery, since this may compromise the blood supply to uninvolved intestine. When two grossly abnormal segments of bowel have a short intervening segment of apparently normal bowel, the question arises of whether to try to save the normal "skip area." The answer depends on the length of the intervening normal segment. If it is 10 cm or longer, it probably should be saved. If it is 5 cm or shorter, it should be taken. Both areas would be excised in this instance, but each case must be individualized, and the answer will depend also on the configuration of the blood supply, the mobility of the mesentery, and other local factors. If previous resections have left the bowel critically short, we may go to great lengths to preserve small segments.

In selected patients with obstruction caused by strictures (either single or multiple), one option is to perform a stricturoplasty that effectively widens the lumen but avoids excision. Stricturoplasty is performed by making a longitudinal incision through the narrowed area of intestine followed by closure in a transverse fashion (Fig. 25-18). This technique preserves intestine and is associated with complication and recurrence rates comparable to resection. In the past, bypass procedures were commonly used, but today, bypass with exclusion is used only in elderly, poor-risk patients, in patients who have had several prior resections and can ill afford to lose any more bowel, and in patients in whom resection would necessitate entering an abscess or endangering normal structures. The reasons for this are twofold: the disease often persists in the bypassed segment with development of intraabdominal sepsis, and cancer may occur in the bypassed segment.

Fistula

Fistulas in patients with Crohn's disease usually are enteroenteral or enterocutaneous. The presence of a radiographically demonstrable enteroenteral fistula without any signs of sepsis or other complication is not in itself an indication for operation. The vast majority of patients who have radiographic demonstration of enteroenteral fistulas, however, will develop later complications, most commonly sepsis, and will require resection of the fistula. Enterocutaneous fistulas are rarely spontaneous, but follow resections or external drainage of intraabdominal abscesses. Enterocutaneous fistulas should be managed by excising the area of grossly involved small intestine and debriding the fistulous tract through the abdominal wall from the inside. After the bowel is reanastomosed, it should be separated, if possible, from the abdominal wall with omental fat. The formation of enteroenteric or enterovesical fistulas usually involves either a previous intestinal perforation with formation of a walled-off abscess that finally perforates or the adherence of one diseased loop of bowel to another diseased loop of bowel or an adjacent normal organ. If the fistula forms between two or more adjacent loops of diseased bowel, the entire involved segment should be excised. On the other hand, if the fistula involves an adjacent normal organ, such as the bladder or the colon, only the segment of the diseased small bowel and fistulous tract should be resected, and the defect that is left in the normal organ is simply closed. The majority of patients with ileosigmoid fistulas do not require resection of the sigmoid because the disease is confined to the small bowel. If the segment of sigmoid involved in an ileosigmoid fistula is found to have Crohn's disease, it should be resected along with the segment of diseased small bowel. The extent of surgical resection in patients with Crohn's disease depends on the extent of critical involvement (Fig. 25-19).

Free Perforation

Perforation into the free peritoneal cavity is not common but occurs occasionally. Free perforation usually occurs in a diseased segment of bowel but can also occur in bowel proximal to an obstruction. The segment should be resected back to relatively good bowel on each side of the perforation, and no attempt should be made to eradicate all diseased bowel. If generalized peritonitis is present, an anastomosis should not be done, because there is a high rate of anastomotic dehiscence under these circumstances; an ileostomy should be performed until intraabdominal sepsis is controlled. Later restoration of intestinal continuity can be carried out safely.

Urologic Complications

The most common urologic complication is ureteral obstruction, which is usually due to ileocolic disease with retroperitoneal abscess. Surgical treatment of the primary intestinal disease is adequate in most patients; however, in a few instances of long-standing inflammatory disease, periureteral fibrosis may be present and require extensive ureterolysis.

Hemorrhage

Although anemia from chronic blood loss is common, life-threatening hemorrhage in patients with Crohn's disease is rare. The incidence of clinical hemorrhage is more common when Crohn's disease involves the colon rather than the small bowel, but both occur much less frequently than bleeding in patients with ulcerative colitis. When life-threatening hemorrhage occurs, the segment involved should be resected and intestinal continuity restored; arteriography may be useful to localize the bleeding before surgery.

Growth Retardation

Up to one-fourth of children with Crohn's disease have significant growth retardation, which may precede the appearance of active bowel disease. The major reason for this growth retardation appears to be nutritional sequelae that accompany Crohn's disease. In these children, in the absence of other common indications, bowel resection is indicated; resection often results in restoration of normal growth when it is performed before epiphyseal closure has occurred. As in adults, children with Crohn's disease tend to have disease of the terminal ileum and colon and can be treated with limited resection and reanastomosis. Involvement of the colon and rectum appears to be more virulent and may require proctocolectomy before restoration of growth is accomplished.

Cancer

Until recently, a causal relation between Crohn's disease and the development of cancer was not accepted. Although cancer does not occur at the rate it does in patients with ulcerative colitis, a number of reports have clearly demonstrated that patients with long-standing Crohn's disease have an increased incidence of cancer. The management of these patients is the same as in any patient with cancer—that is, resection of the cancer with appropriate margins and regional lymph nodes. Patients with cancer associated with Crohn's disease tend to have a worse prognosis than those who do not have Crohn's. This can be explained in large part by the fact that many of these patients are initially managed medically and therefore present at a later time.

Anorectal Disease

Treatment of perianal disease should be conservative; wide excision of abscesses or fistulas is not indicated. Definitive fistulotomy is indicated in the majority of patients, although one must recognize that some degree of anal stenosis may occur as a result of the chronic inflammation. Fissures usually are lateral, relatively painless, large, and indolent, and they usually respond to conservative management. Abscesses should be drained, but large excisions of tissue should not be performed.

Colorectal Disease

The same principles apply to the management of Crohn's disease that is limited to the colon and rectum as to disease in the small intestine. Operative indications include a lack of response to medical management or complications of Crohn's disease. As much bowel as possible should be conserved. Procedures include segmental colectomy with colocolic anastomosis, subtotal colectomy with ileoproctostomy in a patient with mild or minimal rectal disease, and total proctocolectomy with Brooke ileostomy. Patients with a toxic megacolon should undergo colectomy, closure of the proximal rectum, and end ileostomy.

A potential problem after proctocolectomy is delayed healing of the perineal wound. Several series have reported that 25 to 60 percent of perineal wounds are open 6 months after operation. Persistent nonhealing wounds require excision with secondary closure or closure with the use of musculocutaneous flaps (most commonly gracilis or gluteus maximus flaps). Continence-preserving operations such as ileoanal-pouch anastomoses or continent ileostomies (Kock pouch) that have been used successfully in patients with ulcerative colitis are not recommended for patients with Crohn's colitis because of the high rate of recurrence of Crohn's disease in the pouch, fistulas to the anastomosis, and peripouch abscesses.

Duodenal Disease

Crohn's disease of the duodenum requiring operative intervention is rare. The primary indication for surgery in these patients is duodenal obstruction that does not respond to medical therapy. Gastrojejunostomy to bypass the disease, rather than duodenal resection, is the procedure of choice; however, stricturoplasties have been performed with success in selected patients.

Prognosis

There is no medical or surgical cure for Crohn's disease. A substantial proportion of patients who have had one operation will require another operation for complications of Crohn's disease. In one study the cumulative recurrence rates were reported to be 29 percent at 5 years, 52 percent at 10 years, 64 percent at 15 years, and 84 percent at 25 years. No important differences were noted between disease location and the type of operation. Which patients will develop recurrent disease cannot be predicted. Sex, age, duration of disease, granulomas, enteral or perirectal fistulas, length of resection, length of diseased gut, and proximal resection margin have no significant influence on the rate of disease recurrence or on functional outcome. The most common site of recurrence is the small bowel proximal to the site of previous resection. Recurrence is five times more likely to involve the adjacent or remote colon in patients with ileocolitis than in patients with ileitis. Few patients with ileitis eventually require ileostomy, whereas one-third of patients with ileocolitis or colitis require permanent ileostomy.

Crohn's disease gradually recedes with advancing age. Active disease is unusual between the ages of 50 and 55 years, but gradual cicatrization on healing may cause bowel obstruction in older patients.

Tuberculous Enteritis

A resurgence in the incidence of tuberculosis has occurred in Western countries as a result of the growing AIDS epidemic, an influx of Asian immigrants, and widespread use of immunosuppressive agents after organ transplantation. Tuberculosis continues to be a significant and relatively common disease in developing countries with incidences as high as 400 to 500 people per 100,000 in some countries. Tuberculosis of the gastrointestinal tract occurs in two forms. Primary infection usually is caused by the bovine strain of *Mycobacterium tuberculosis* and results from ingesting infected milk; this now accounts for less than 10 percent of reported cases. The majority of infections occur secondarily from the swallowing of bacilli by patients with active pulmonary tuberculosis. Chemotherapy has greatly decreased the incidence of secondary tuberculous enteritis, so that it now occurs in only about 1 percent of patients with pulmonary tuberculosis.

The ileocecal region is the site of involvement in about 85 percent of patients with tuberculous enteritis, presumably because of the abundance of lymphoid tissue in this area. Three patterns of involvement are seen: hypertrophic, ulcerative, and ulcerohypertrophic.

Occasionally a hypertrophic reaction produces contracture and stenosis of the lumen of the distal ileum, which requires resection. The diagnosis is not often made preoperatively; only 50 percent of patients have evidence of active pulmonary tuberculosis; 45 percent of patients have negative skin tests. Cultures, when positive, provide a definitive answer, but results may take longer than 4 weeks. The polymerase chain reaction technique has been used successfully to differentiate tuberculosis enteritis from Crohn's disease in biopsy specimens; results may be obtained within 2 days. If the diagnosis is suspected preoperatively, chemotherapy, best given as a combination of isoniazid (100 mg three times a day) with either rifampin (600 mg daily) or ethambutol (15 mg/kg daily), should be administered for about 2 weeks before operation. If unsuspected tuberculosis is found at operation, chemotherapy should be started. In either case drugs should be continued for 1 year after the patient has become asymptomatic. Surgical therapy of hypertrophic tuberculous enteritis usually is inadvisable unless bowel obstruction from high-grade stenosis requires relief. Resection and anastomosis should be performed unless extensive disease is present; otherwise ileocolic bypass is the safer procedure.

Symptoms of the more common ulcerative form of tuberculous enteritis include alternating constipation and diarrhea associated with nonspecific crampy abdominal pain. The diagnosis sometimes is made only by radiographic intestinal examination of patients with pulmonary tuberculosis. In severe cases diarrhea is persistent, and anemia, hypoalbuminemia, and inanition are progressive. Clinical confirmation of a presumptive diagnosis may be obtained by a prompt response to antituberculous chemotherapy. Surgical therapy is contraindicated except for the rare complications of perforation, obstruction, or hemorrhage.

Typhoid Enteritis

Typhoid enteritis is an acute systemic infection of several weeks' duration caused by *Salmonella typhosa*. Hyperplasia and ulceration of Peyer's patches of the intestine occur along with mesenteric lymphadenopathy, splenomegaly, and parenchymatous changes in the liver. Confirmation of diagnosis is obtained by culturing *S. typhosa* from blood or feces or by finding a high titer of agglutinins against the O and H antigens.

Typhoid fever is still a major disease in areas of the world that have not yet attained high public health standards, but it is now rare in Western countries. The death rate, formerly about 10 percent, is now about 2 percent, in large part because of the specific antimicrobial chloramphenicol, introduced in 1948.

There is some disagreement on whether chloramphenicol remains the drug of choice, however, because of the emergence of resistant strains, high relapse rate, and risk of marrow toxicity. Trimethoprim-sulfamethoxazole has emerged as the best successor to chloramphenicol. Trimethoprim 160 mg and sulfamethoxazole 800 mg are given, either orally or parenterally, twice daily for 2 weeks. Amoxicillin also is effective and should be given intravenously or intramuscularly in doses of 1 g every 6h for 2 weeks.

Gross hemorrhage occurs in 10 to 20 percent of hospitalized patients, even while on adequate therapy. Transfusion is indicated and usually suffices. Every effort should be made to avoid operation, since the bleeding typically is from multiple ulcers and the bowel is exceedingly friable. Rarely, laparotomy must be done for uncontrollable, life-threatening hemorrhage.

Perforation through an ulcerated Peyer's patch occurs in about 2 percent of cases. Typically it is a single perforation in the terminal ileum. Operative treatment is indicated unless the patient is moribund, since localization or walling-off of the perforation is uncommon. Simple closure of the perforation is successful in the majority of patients. With multiple perforations, which occur in about one-fourth of the patients, resection with primary anastomosis or a temporary ileostomy is preferred. The mortality rate of those with free perforation is about 10 percent.

Neoplasms

General Considerations

Primary neoplasms of the small bowel are rare, despite the much greater mucosal surface area and the rapidity of cell turnover of the small bowel compared to the stomach and colon. Neoplasms are 40 times more common in the colon; the reasons for this difference are unknown. Several theories have been proposed as possible explanations: rapid transit time decreases the time for contact of carcinogens with the mucosa; the local immune system of the small-bowel mucosa; the alkaline pH of the succus entericus; the absence of bacteria that might convert certain ingested products into carcinogens; and the presence of mucosal enzymes that destroy certain carcinogens.

Primary small-bowel neoplasms are either benign or malignant and may arise from any of the cells of the small bowel. They often are asymptomatic; when symptoms occur, they usually are those of anemia or of obstruction (e.g., enteric cramping, diarrhea, distention, and obstipation). Any tumor involving the surface mucosa of the gut will bleed from time to time, and so patients present with symptoms of anemia, as do patients with colonic tumors. Benign tumors may be hamartomas, adenomas, hemangiomas, fibromas, lipomas, or leiomyomas, among others. Benign tumors usually can be excised locally. Whether or not to excise a complete segment of bowel depends on whether the tumor can be removed easily without compromising the lumen. Malignant tumors usually are adenocarcinomas, lymphomas, or leiomyosarcomas. Any kind of soft-tissue malignancy may involve the small bowel. Malignant tumors should be resected with a good margin (about 10 cm) of normal bowel on each side, along with the subtended wedge of mesentery, so as to remove local and regional lymphatics and lymph nodes.

Adenocarcinoma of the small bowel (Fig. 25-20 A) is similar to colonic adenocarcinoma. Both often present as apple-core lesions that may cause blood loss and anemia; after producing constriction of the lumen, it causes cramping and intermittent diarrhea and, when the constricted lumen is ultimately plugged by enteric material, obstruction. The material causing the obturator obstruction may later be spontaneously passed, so that intermittent symptoms of obstruction and diarrhea may occur. Malignant soft-tissue tumors may appear in any form (Fig. 25-20 B). Sarcomas may rarely invade the mucosa and bleed. Hamartomas or adenomas may be single or multiple (Fig. 25-20 C). Benign tumors of the wall of the bowel may grow to sufficient size to cause narrowing of the lumen and even obstruction (Fig. 25-20 D). Occasionally these tumors may become sufficiently large and elongated to assume a polypoid shape and may be carried along by peristalsis; sometimes they infarct and bleed. Malignant tumors should be resected with a wide margin of normal bowel and a subtended wedge of mesentery (Fig. 25-20 E) in an attempt to capture any tumor in lymphatic transit. After resection an end-to-end anastomosis of the small bowel and repair of the mesentery are performed (Fig. 25-20 F).

The frequencies of benign and of malignant neoplasms are reported to be equal in surgical series, and benign neoplasms far exceed malignant neoplasms in autopsy series. Malignant neoplasms (75 to 80 percent) more often produce symptoms than benign neoplasms.

Symptoms associated with small-bowel neoplasms often are vague; they include epigastric discomfort, nausea, vomiting, abdominal pain (often intermittent and colicky), diarrhea, and bleeding (often manifested as symptoms of anemia). Symptoms may be present for months or years before operation. Bleeding usually is occult; hematochezia or hematemesis may occur, although life-threatening hemorrhage is uncommon. The most common indications for operation in patients with neoplasms of the small bowel are obstruction, bleeding, and pain. The mechanism of obstruction varies, depending on whether it is caused by a benign or a malignant neoplasm. Benign neoplasms are the most common cause of intussusception in the adult, whereas malignant neoplasms commonly cause obstruction by circumferential growth or kinking of the bowel as a result of longitudinal, intramural growth.

Because of their relative infrequency and the vague symptoms they produce, the diagnosis of small-bowel neoplasms requires informed suspicion. Endoscopy is useful in evaluating the duodenum and possibly the most proximal jejunum, just beyond the ligament of Treitz. However, since most of the small bowel is not accessible to endoscopic scrutiny, barium contrast radiography usually is required. Upper gastrointestinal series with small-bowel follow-through is the single most useful diagnostic test. Enteroclysis (intubation infusion for small-bowel examination) has been promoted as a sensitive technique to demonstrate small-bowel disease. In performing enteroclysis, the tip of the nasal–small bowel tube is placed just beyond the ligament of Treitz, and dilute barium is instilled at a constant rate. The examination is complete when the barium reaches the cecum. Enteroclysis (Fig. 25-21) is superior to small-bowel follow-through for detection of small filling defects and for determination of changes in mucosal pattern; the diagnostic accuracy of enteroclysis approaches 90 percent.

Benign Neoplasms

Pathology

Benign tumors of the small bowel may be of epithelial or connective tissue origin. The incidence of these tumors varies, depending on whether they are reported in autopsy or clinical series. The most common lesions are adenomas, leiomyomas, and lipomas. Other benign lesions include hamartomas, fibromas, angiomas, lymphangiomas, neurofibromas, and hemangiomas. Adenomas are the most common benign tumors reported in autopsy series, but leiomyomas are the most common benign small-bowel lesions that produce symptoms.

Clinical Manifestations

An appreciable number of small-bowel benign tumors apparently cause no serious symptoms during life and are incidental findings at autopsy. The diagnosis is delayed or missed in many patients because symptoms may be absent or vague or nonspecific until significant complications have developed. Physical examination rarely provides any clue unless intestinal obstruction is present, and radiographic studies of the small bowel and selective angiography, the only specific diagnostic aids, may fail to demonstrate an existing tumor, even if it is suspected clinically. In only about one-half of small-bowel tumors found at operation has the correct diagnosis been made preoperatively.

The two most common clinical manifestations of small-bowel tumors are bleeding and obstruction (Fig. 25-22). Rarely, perforation of the bowel wall occurs, resulting in abscess or internal fistula formation, peritonitis, or pneumatosis cystoides intestinalis. Bleeding occurs in about one-third of patients but is rarely gross. More commonly, bleeding is occult and intermittent, producing guaiac-positive stools and iron-deficiency anemia. Leiomyomas and hemangiomas are the lesions that most often cause bleeding.

Treatment

Surgical treatment of benign tumors is nearly always indicated because of the risk of subsequent complications and because the diagnosis of benign disease cannot be made without microscopic evaluation. The complications of benign neoplasms that most often require treatment are bleeding and obstruction. Segmental resection and primary reanastomosis is most commonly used except for very small lesions that may be excised by enterotomy. The entire small bowel should be searched for other lesions, since they are often multiple.

Adenoma

Adenomas, which account for approximately 35 percent of all benign tumors, are of three primary types: true adenomas, villous adenomas, and Brunner's gland adenomas. Twenty percent of adenomas are found in the duodenum, 30 percent in the jejunum, and 50 percent in the ileum. The majority of adenomas are asymptomatic; most occur singly, and most are found incidentally at autopsy. It is rare for duodenal polyps to be sufficiently large to cause symptoms. If polyps do cause symptoms, they usually are associated with bleeding or obstruction. If symptomatic, duodenal polyps can be removed by endoscopic snare, or they may be sufficiently large to be difficult to retrieve through the pylorus. In such cases, particularly if malignancy is suspected, they may best be removed in an open operation. As a rule, the question of malignancy can be settled by endoscopic biopsy; usually the polyp can be removed endoscopically.

Villous adenomas of the small bowel are rare but do occur and are most commonly found in the duodenum. Their presence may be suspected by the characteristic “soap bubble” appearance on contrast radiography. They may attain large size (> 5 cm in diameter) and usually are found because of pain or bleeding. Obstruction also may occur. There have been no reports of secretory diarrhea associated with villous tumors of the small bowel; however, the malignant potential of these lesions is reportedly between 35 and 55 percent.

Brunner's gland adenomas are hyperplastic proliferations of normal exocrine glands located in the submucosa of the duodenum. Brunner's gland adenomas may produce symptoms that mimic those of peptic ulcer disease, or they may cause obstruction. Diagnosis can be made by endoscopy and biopsy, and symptomatic lesions in an accessible region should be resected. There is no malignant potential for Brunner's gland adenomas, and a radical resection should not be employed.

Leiomyoma

The most common symptomatic benign lesions of the small bowel are leiomyomas. Leiomyomas are benign tumors of smooth muscle that are most common in the jejunum. They usually occur singly, although multiple tumors may occur. The incidence is equal in men and women. Two growth patterns are noted: the tumor may grow primarily intramurally and cause obstruction, or both intramural and extramural growth may occur and produce a dumbbell-shaped mass. These tumors may attain considerable size, outgrowing their blood supply, and tumor necrosis with bleeding may occur. The most common indication for operation on leiomyomas is bleeding. Angiography may provide the correct preoperative diagnosis.

Lipoma

Lipomas, most common in the ileum, are single intramural lesions, submucosal in location, and usually small. Fewer than one-third of lipomas are found at operation and, when found, usually they are the cause of obstruction, most commonly as the lead point of an intussusception. Bleeding may occur from ulceration of the overlying mucosa. Lipomas do not have malignant potential, and therefore when found incidentally, they should be removed only if the resection is simple. Pedunculated lipomas should be excised.

Hamartoma (Peutz-Jeghers Syndrome)

Hamartomas of the small bowel occur only as part of the Peutz-Jeghers syndrome, which is an inherited syndrome of mucocutaneous melanotic pigmentation and gastrointestinal polyps. The pattern of inheritance is simple mendelian dominant, with a high degree of penetrance. A single pleiotropic gene is responsible for both polyps and melanin spots. The classic pigmented lesions are small, 1- to 2-mm brown or black spots located in the circumoral region of the face, buccal mucosa, forearms, palms, soles, digits, and perianal area. The syndrome was first reported in 1921 by Peutz, and Jeghers and colleagues described it anew in 1949. Multiple pigmented lesions may be noted, or only a single buccal lesion may be present. Pigmentation appears in childhood. All cutaneous lesions may fade, leaving only buccal lesions. Polyposis with and without pigmentation has been reported. The entire jejunum and ileum are the most frequent portions of the gastrointestinal tract to be involved with multiple polyps. Fifty percent of patients may also have rectal and colonic polyps, and 25 percent may have gastric polyps. The chief point to note is that if a patient with multiple rectal, colonic, or gastric polyps is found to have hamartomas rather than adenomas, a search for small-bowel polyposis and pigmented lesions should be carried out.

The lesions are not true polyps but are hamartomas and hence are not premalignant (Fig. 25-23). However, there have been a few reported cases of malignant tumors of the gastrointestinal tract associated with Peutz- Jeghers syndrome. Some of these adenomatous and carcinomatous changes were noted in the hamartomatous polyps. It is not clear, however, whether this represents a coincidence or a true malignant transformation of this syndrome.

The most common symptom is recurrent colicky abdominal pain, caused by intermittent intussusception. Lower abdominal pain associated with palpable mass has been reported to occur in one-third of patients. Hemorrhage occurs less frequently and is most commonly manifested by insidious involvement of anemia. Acute life-threatening hemorrhage is uncommon but may occur.

Surgical therapy is required only for obstruction or persistent bleeding. The resection should be limited to the segment of bowel that is producing complications, that is, polypectomy or limited resection. Because of the widespread nature of intestinal involvement, cure is not possible and extensive resections are not indicated.

Hemangioma

Hemangiomas are developmental malformations consisting of submucosal proliferation of blood vessels. They can occur at any level of the gastrointestinal tract, and the jejunum is the most commonly affected small-bowel segment. Hemangiomas account for 3 to 4 percent of all benign tumors of the small bowel and are multiple in 60 percent of patients. Hemangiomas of the small bowel may occur as part of an inherited disorder known as Rendu-Osler-Weber disease. Besides the small bowel, hemangiomas may be present in the lung, liver, and mucous membranes. Patients with Turner's syndrome are likely also to have cavernous hemangiomas of the intestine. Most patients are symptomatic, and the most common symptom is diffuse intestinal bleeding. Diagnosis is often difficult; angiography and ^{99m}Tc—red cell scanning are the most useful diagnostic tools. If a hemangioma is localized preoperatively, resection of the involved segment of intestine is warranted. If the site is not localized preoperatively, intraoperative transillumination and palpation can be helpful.

Malignant Neoplasms

The most common malignant neoplasms of the small bowel, in approximate order of frequency, are adenocarcinomas, carcinoids, sarcomas, and lymphomas.

Clinical Manifestations

Rochlin and Longmire called attention to three distinct clinical presentations of patients with malignant small-bowel neoplasms: diarrhea, with large amounts of mucus, and tenesmus; obstruction with nausea, vomiting, and cramping abdominal pain; and chronic blood loss with anemia, weakness, guaiac-positive stools, and occasionally melena or hematochezia. As with benign neoplasms, symptoms of malignant neoplasms often are present for many months before the diagnosis is made, emphasizing their insidious nature.

Treatment

The treatment for malignant neoplasms of the small bowel is wide resection, including regional lymph nodes. This may require a radical pancreaticoduodenectomy (Whipple operation) for duodenal lesions. Because of the extent of the disease at the time of operation, curative resection may not be possible. Palliative resection should be performed when possible to prevent further complications of bleeding, obstruction, and perforation. If that is not possible, however, bypass of the involved segment is an alternative that may provide worthwhile relief of symptoms. If bypass is used, the proximal end of the bypassed segment should be brought out as a mucous fistula to prevent development of a closed loop. Recent reports of an increased association of small-bowel cancers in patients with celiac sprue suggest that new intestinal complaints in these patients should be evaluated to rule out small-bowel cancer.

Prognosis

The overall survival rate for patients with malignant neoplasms of the small bowel is not good. The highest survival rates are reported for duodenal periampullary carcinomas (about 30 to 40 percent), whereas adenocarcinomas occurring elsewhere in the small bowel have a 5-year survival rate of 20 percent or less. Leiomyosarcomas of the small bowel have a 5-year survival rate of 30 to 40 percent. Radiation and chemotherapy play little role in the treatment of patients with adenocarcinomas of the small bowel. There may be some improvement in survival when radiotherapy is used in patients with sarcomas. Determinants of survival for patients with lymphomas are the cell type and the extent of disease. Radiotherapy and chemotherapy combined with surgical excision provide the best survival rates for patients with lymphomas. Five-year survival rates have been reported to range from 10 to 50 percent, with an average of about 30 percent.

Carcinoma

Carcinomas constitute about 50 percent of the malignant tumors of the small bowel in most reported series. They are twice as common in men as in women, and the average age at diagnosis is 50 years. For unknown reasons, adenocarcinomas are more common in the duodenum and the proximal jejunum than in the remainder of the small bowel. About half of duodenal carcinomas involve the ampulla of Vater.

The carcinoma's location in the small bowel often determines the presenting symptoms. For example, periampullary adenocarcinomas are associated with intermittent jaundice, whereas carcinomas of the jejunum usually produce symptoms of mechanical small-bowel obstruction (Fig. [25-24](#)). The presence of jaundice, often intermittent, and a guaiac-positive stool should immediately call to mind the possibility of a periampullary carcinoma.

As with carcinomas arising in other organs, survival of patients with small-bowel carcinomas is related to the stage of disease at the time of diagnosis. Diagnosis often is delayed and disease far advanced at the time of operation. The delay in diagnosis is due to a combination of factors, including lack of suspicion because of the relative rarity of the lesions, vagueness of symptoms, and absence of physical findings.

Sarcoma

Sarcomas, which arise from mesodermal tissue, constitute about 20 percent of malignant neoplasms of the small bowel. The most common sarcoma by far is the leiomyosarcoma. Leiomyosarcomas are evenly distributed throughout the small bowel. The incidence is the same for men and women, and diagnosis usually is made in the sixth decade. The most common indications for operation are bleeding and obstruction, although free perforation as a result of hemorrhagic necrosis in large tumor masses may occur. Leiomyosarcomas are spread by direct invasion of adjacent structures, by hematogenous dissemination, or by transperitoneal seeding, producing sarcomatosis. Other rare forms of sarcoma that may affect the small bowel include fibrosarcoma, angiosarcoma, liposarcoma, and Kaposi's sarcoma.

Lymphoma

Lymphomas constitute 10 to 15 percent of small-bowel malignant tumors in the adult; in children under 10 years of age, they are the most common intestinal neoplasm. Lymphomas most commonly are found in the ileum, where the greatest concentration of gut-associated lymphoid tissue is present, and they may be primary or part of a generalized disease. Dawson and colleagues have devised criteria to determine whether lymphoma of the small bowel is primary. These include (1) absence of peripheral lymphadenopathy; (2) normal chest x-ray without evidence of mediastinal lymph node enlargement; (3) normal white blood cell count, total and differential; (4) at operation, the bowel lesion must predominate, and the only involved nodes are associated with the bowel lesion; and (5) absence of disease in the liver and spleen. Even when these criteria are used, one-third or more patients with lymphomatous involvement of the small bowel are found to have generalized lymphoma.

There are three syndromes of small-bowel lymphoma. *Western lymphoma* is a disease predominantly of adults, typically found in the Western Hemisphere and associated with severe malabsorption in 5 to 10 percent of patients. Another form, known as *Mediterranean lymphoma*, is a malignant form that was first noted in non-Ashkenazi Jews and Arabs in Israel and was subsequently reported in other countries and in other ethnic groups, including Hispanic Americans. Since this disease is not confined to the Mediterranean basin, the term *immunoproliferative small intestinal disease* has been used. One-third of the patients may be found to have an abnormal fragment of IgA heavy chain in their serum, which is produced by plasma cells infiltrating the small bowel. This variant is known as *heavy-chain disease*.

The third intestinal manifestation of lymphoma is childhood abdominal lymphoma. This is a group of lymphomas including American (nonendemic) Burkitt's lymphoma, undifferentiated non-Burkitt's lymphoma, and diffuse histiocytic lymphoma.

Carcinoid

Carcinoids of the small bowel arise from the enterochromaffin cells, or Kulchitsky cells, found in the crypts of Lieberkühn. These cells are also known as argentaffin cells because of their staining by silver compounds. Carcinoids may be classified by embryologic site of origin and secretory product (Table [25-3](#)). Carcinoids of the small bowel occur with almost the same frequency as adenocarcinoma, and together they make up the preponderance of malignant neoplasms of the small bowel. Carcinoids have variable malignant potential and are composed of multipotential cells with the ability to secrete numerous humoral agents, the most prominent of which are serotonin and substance P. Although the carcinoid syndrome, characterized by episodic attacks of cutaneous flushing, bronchospasm, diarrhea, and vasomotor collapse, occurs in fewer than 5 percent of the patients with malignant carcinoids, it is quite dramatic and has been extensively described and discussed with fascination by many authors.

The primary importance of carcinoid tumors, however, is not the carcinoid syndrome but the malignant potential of the tumors themselves. Oberndorfer coined the term *karzinoide* to indicate that this tumor is carcinomalike and to emphasize the assumed lack of malignant potential. By 1930 this concept was no longer supported, since many patients with metastatic carcinoid tumors had been reported. In 1953 and 1954 the carcinoid syndrome was described.

Pathology

Carcinoids may arise in organs derived from the foregut, midgut, and hindgut. In the gastrointestinal tract over 90 percent of carcinoids are found in three sites: the appendix is most frequently involved (46 percent), followed by the ileum (28 percent) and the rectum (17 percent), but all organs may give rise to carcinoids. Other gastrointestinal locations are listed in Table [25-4](#). The malignant potential (ability to metastasize) is related to location, size, depth of invasion, and growth pattern. Only about 3 percent of appendiceal carcinoids metastasize, but about 35 percent of ileal carcinoids are associated with metastasis. Seventy-five percent of gastrointestinal carcinoids are less than 1 cm in diameter; only 2 percent of this group metastasize. About 20 percent of primary tumors are 1 to 2 cm in diameter, and 50 percent of this group metastasize. Only about 5 percent are over 2 cm in diameter; 80 to 90 percent of these metastasize. Carcinoids limited to the submucosa do not metastasize, whereas 69 percent of carcinoids involving the serosa have metastases. The histologic growth patterns appear to be related to survival and therefore have prognostic importance.

Multiple carcinoids of the small bowel occur in 30 percent of patients with carcinoid but are rare in the appendix. This tendency to multicentricity exceeds that of any other malignant neoplasm of the gastrointestinal tract. An unusual observation that is yet unexplained is the frequent coexistence of a second primary malignant neoplasm of a different histologic type; a second primary neoplasm was reported in 25 percent of patients in one large series.

Carcinoids present grossly as a slightly elevated, smooth, rounded, hard nodule, covered with normal mucosa. On cut section, they have a characteristic yellow-gray or tan appearance. Extensive fibrosis of the mesentery and of the bowel wall, caused by an intense desmoplastic reaction, may be present (Fig. [25-25](#)). This fibrosis may produce mechanical bowel obstruction from kinking or matting of loops of small bowel together. Obstruction rarely is a result of direct tumor encroachment on the lumen of the bowel. In addition to the desmoplastic reaction, apparently produced by humoral agents elaborated by the tumor, metastases to mesenteric nodes also result in kinking and fixation by large metastatic tumor deposits. A small primary tumor associated with massive mesenteric metastases is often noted (Fig. [25-26](#)).

Clinical Manifestations

In the absence of the malignant carcinoid syndrome, symptoms of patients with carcinoid tumors of the small bowel are similar to those of patients with small-bowel tumors of other histologic types. The most common symptoms are abdominal pain, bowel obstruction, diarrhea, and weight loss. In the majority of patients (in the absence of the malignant carcinoid syndrome), the diarrhea is a result of partial bowel obstruction rather than being secretory. Rarely, malignant carcinoid tumors of the ampulla of Vater are found in patients with disseminated neurofibromatosis (von Recklinghausen's disease).

Diagnostic Findings

Radiographic studies of the small bowel may exhibit multiple filling defects, sometimes caused by tumors but more often a result of kinking and fibrosis of the bowel, mesenteric calcifications, and fixed, rigid loops of intestine. Mesenteric vascular angiography may reveal abnormal arrangement of mesenteric arteries and narrowing of peripheral branches together with poor accumulation of contrast and poor venous drainage of the tumor area. Tumor staining during angiography may be enhanced by administration of norepinephrine. Angiography is the most sensitive diagnostic test to detect hepatic metastasis, particularly diffuse metastatic disease with fine nodular distribution. Hepatic metastases are hypervascular and intensely stained during arteriography.

Treatment

The treatment of patients with small-bowel carcinoid tumors is based on the size and the site of the tumor and the presence or absence of metastatic disease. For primary tumors less than 1 cm in diameter without evidence of regional lymph node metastasis, a segmental intestinal resection is adequate. For patients with lesions greater than 1 cm, with multiple tumors, or with regional lymph node metastasis regardless of the size of the primary tumor, wide excision of bowel and mesentery is required. Since the majority of small-bowel carcinoids are found in the ileum, wide excision usually entails a right hemicolectomy. Malignant carcinoid tumors of the duodenum may require radical pancreatoduodenectomy. When carcinoids are localized, almost all are cured by surgical resection; patients with the malignant carcinoid syndrome caused by metastatic disease may receive significant palliation but are not cured with surgical resection.

Treatment of Carcinoid Tumors of the Appendix

Appendiceal carcinoids typically are diagnosed in patients 20 to 40 years of age, in contrast to the mean age at diagnosis of small-bowel carcinoids of 60 years. The majority of appendiceal carcinoids are located at the tip. Simple appendectomy is curative for patients with tumors less than 1 cm in diameter without gross evidence of metastasis. Because of the potential for metastasis, right hemicolectomy should be performed for tumors greater than 2 cm. Intramural lymphatic invasion, serosal involvement, or microscopic involvement of the mesoappendix associated with tumors less than 2 cm in diameter is not an indication for extensive resection.

Prognosis

The survival rate for patients with small-bowel carcinoid tumors has been reported to be 75 percent for those with tumors staged as local, 59 percent for regional tumors, and 19 percent for tumors with distant spread. The overall survival rate is 54 percent. The growth pattern of carcinoid tumors is an independent predictor of outcome. The median survival times of the various growth patterns, in decreasing order, are as follows: mixed insular plus glandular, 4.4 years; insular, 2.9 years; trabecular, 2.5 years; mixed insular plus trabecular, 2.3 years; glandular growth pattern, 0.9 years; and undifferentiated, 0.5 years.

When widespread metastatic disease precludes cure, extensive resection for palliation is indicated. Since these tumors are often indolent and slow-growing, long-term palliation often is obtained. Bypass procedures may be used in poor-risk patients with extensive disease. The overall 5-year survival rate after resection of intestinal carcinoids is about 50 percent. When "curative resection" is done, the 5-year survival rate is 70 percent. Chemotherapy has not been entirely successful; however, treatment with streptozotocin and 5-fluorouracil (5-FU) may provide significant palliation. In patients with palliative resections the 5-year survival rate is 25 percent.

Malignant Carcinoid Syndrome

This rare syndrome is widely described but rarely seen. Thirty to 70 percent of carcinoid tumors of the gut are metastatic at the time of diagnosis, but only 6 to 9 percent of patients with metastatic disease develop manifestations of the malignant carcinoid syndrome. By far the most commonly associated primary tumor is located in the small bowel, and massive hepatic replacement by metastatic tumor is usually found.

Clinical Manifestations

The syndrome is characterized by hepatomegaly, diarrhea, and flushing in 80 percent of patients, right heart valvular disease in 50 percent, and asthma in 25 percent. Malabsorption and pellagra (dementia, dermatitis, and diarrhea) are occasionally present and are thought to be caused by excessive diversion of dietary tryptophan to meet the metabolic requirements of the tumor.

Diarrhea is episodic, often occurring after meals, and is a result of elevated circulating levels of serotonin, which stimulates secretion of small-bowel fluid and electrolytes and increases intestinal motility. Some patients may present with acute abdominal symptoms, characterized by severe abdominal cramping without mechanical bowel obstruction, which has been called *carcinoid abdominal crisis*. The mechanism of this crisis is thought to be intestinal ischemia, caused by the vasoactive substances elaborated by the tumor, combined with decreased mesenteric blood supply from perivascular fibrosis.

Flushing is not temporally related to diarrhea, and, although both may be present, either may be present without the other. The lack of relationship between flushing and diarrhea suggests that these two manifestations of the syndrome are caused by different mediators. Although substance P produces all of the vasomotor phenomena associated with the flush, it has been questioned as the primary mediator. Besides serotonin and substance P, other substances that have been implicated include bradykinin and prostaglandins E and F.

Valvular heart disease is a result of irreversible endocardial fibrosis, which is similar in genesis to the fibrosis noted in the gut wall, in the retroperitoneum, and around the mesenteric blood vessels. It occurs in patients with hepatic metastases and is limited to tricuspid and pulmonary valves. The reason the lesions are restricted to the valves on the right side is that they are exposed to high levels of serotonin; the pulmonary filter deactivates serotonin, thereby preventing left-sided valvular lesions.

Asthma is due to bronchoconstriction, which may be produced by serotonin, bradykinin, or substance P. Treatment of asthma associated with carcinoid syndrome must be carried out very carefully, since the use of adrenergic drugs may cause the release of humoral agents that may cause status asthmaticus.

Although the syndrome is seen in patients with high circulating levels of serotonin and often substance P, these are probably not the only mediators of all components of the syndrome. The malignant enterochromaffin cells produce 5-hydroxytryptamine (5-HT), which is serotonin. Circulating serotonin is metabolized in the liver and in the lung to 5-hydroxyindoleacetic acid (5-HIAA), which is pharmacologically inactive. Elevated levels of 5-HIAA are seen only in patients with metastasis. However, not all patients with metastasis have increased levels of 5-HIAA. The majority of patients who exhibit malignant carcinoid syndrome have massive hepatic replacement by their metastatic disease. Tumors that bypass the hepatic filter, specifically ovarian and retroperitoneal carcinoids, may produce the syndrome in the absence of liver metastasis.

Diagnostic Findings

The diagnosis is most reliably established by repeated determinations of urinary 5-HIAA. A single determination may be normal in the presence of metastatic disease. Provocative testing to reproduce symptoms has employed injection of pentagastrin, calcium, or epinephrine. The pentagastrin test is by far the most reliable and safest provocative test. During times of testing for increased levels of 5-HIAA, the patient must avoid foods rich in serotonin, such as bananas, tomatoes, walnuts, and pineapples, and certain drugs, including phenothiazines, glycerol guaiacolate, and reserpine.

Treatment

Ideally, treatment of the carcinoid syndrome would require removal of all tumor, but this rarely is possible. Hepatic resection, however, even when known tumor is left behind, may result in significant relief of symptoms because of the removal of the mass of tumor. When resection is not possible, hepatic dearterialization or embolization of the hepatic arterial branches with chemotherapy (doxorubicin, 5-fluorouracil, cisplatin) may provide some relief. The duration of response after resection in one study was 6 months, compared to 4.8 months for hepatic artery ligation. In certain selected patients, liver transplantation may offer the hope of a longer disease-free interval; however, the long-term outcome of this radical approach is not known.

Drug therapy for prevention or relief of symptoms is directed at blockade of the effects of humoral agents elaborated by the tumor. Interferon has provided some symptomatic improvement in a small group of patients. Somatostatin-14 and its long-acting analog octreotide (Sandostatin) relieve symptoms (diarrhea and flushing) of the carcinoid syndrome in the majority of patients. Furthermore, Kvols and colleagues have found that survival is significantly prolonged in patients with carcinoid tumors treated with Sandostatin. In addition to preventing the devastating sequelae of the carcinoid syndrome, Sandostatin also may possess a direct antitumor effect. Chemotherapeutic attacks on the tumor itself have been disappointing, although streptozotocin, alone or combined with 5-fluorouracil, appears to be the most effective.

Treatment of carcinoid tumors is wide surgical resection of the small bowel and regional lymph nodes. In addition, significant palliation may be achieved with aggressive hepatic resection in patients with malignant carcinoid syndrome.

Diverticular Disease

Diverticula of the small bowel may be congenital or acquired. A congenital diverticulum is a true diverticulum; it is composed of all layers of the bowel wall. An acquired diverticulum is a false diverticulum; only mucosa and submucosa protrude through a defect in the muscle coat of the bowel wall. Diverticula may occur in any portion of small intestine. Duodenal diverticula are the most common acquired diverticula of the small bowel, and Meckel's diverticulum is the most common true diverticulum of the small bowel.

Duodenal Diverticula

Duodenal diverticula are common; the duodenum is the second most common site of diverticulum formation, after the colon. The true incidence of duodenal diverticula is unknown, and reported incidences vary, depending on whether they are found clinically (by x-ray, endoscopy, or operation) or at autopsy. It has been reported that 1 to 5 percent of upper gastrointestinal x-ray examinations reveal duodenal diverticula; 9 to 20 percent of upper gastrointestinal endoscopic examinations show them. In autopsy series, incidences of 10 to 20 percent are reported. More than 90 percent of these diverticula are clearly asymptomatic, and less than 5 percent will require operation because of a complication of the diverticulum itself. Duodenal diverticula occur twice as often in women as in men. They are rare in patients under 40 years of age, and the incidence increases with increasing age. Two-thirds to three-fourths of duodenal diverticula are found in a periampullary region and project from the medial wall of the duodenum (Fig. 25-27). Duodenal diverticula are clinically important for two reasons. First, they occasionally produce symptoms related to the diverticulum, including obstruction, perforation, or bleeding, and second, the presence of the diverticulum may cause recurrent pancreatitis, cholangitis, or recurrent common-duct stones after cholecystectomy.

Only those diverticula associated with the ampulla of Vater have a significant relationship to the complications of cholangitis, pancreatitis, and stone disease. In patients with these diverticula, the ampulla most often enters the duodenum at the superior margin of the diverticulum, rather than through the diverticulum itself. The mechanism proposed for the increased incidence of complications of the biliary tract is the location of the perivaterian diverticula, which may produce mechanical distortion of the common bile duct as it enters the duodenum, resulting in partial obstruction and stasis. Bile stasis allows proliferation of bacteria and subsequent formation of stones. The incidence of bacteriobilia is significantly higher in patients with perivaterian diverticula than in those with diverticula located in other parts of the duodenum. The bacteria isolated from the bile duct and from the diverticula

are identical. There is also evidence that dysfunction of the choledochal sphincter is produced by the presence of diverticula. In one study of 101 patients who had undergone cholecystectomy more than 2 years earlier, a significantly higher incidence of recurrent calculi was noted in patients with diverticula than in patients without. A causal relationship of duodenal diverticula to biliary tract stones, however, has not been demonstrated. It is important not to extend indications for operation to include people with unexplained abdominal pain. Nor should retention of barium within diverticula serve by itself as an indication for operation. The important point is that the great majority of duodenal diverticula cause no trouble; they should be left alone unless they can be conclusively related to disease.

Fewer than 60 patients have been reported with intraluminal duodenal diverticula. These diverticula, which probably originate from incomplete duodenal webs, are lined both inside and out with duodenal mucosa, have a characteristic picture of a barium-filled wind sock on contrast radiography, and most often required operation because of duodenal obstruction or recurrent pancreatitis. In these diverticula the common bile duct and pancreatic duct usually enter the diverticulum, and a second orifice is present that allows drainage of the biliary-pancreatic secretions into the lumen of the gut.

Symptoms related to duodenal diverticula, in the absence of any other demonstrable disease, usually are nonspecific epigastric complaints. Bleeding, perforation, and diverticulitis all are rare. The morbidity and mortality caused by complications of diverticula are nonetheless high as a result of delay in diagnosis because of the lack of suspicion of the underlying condition. Diagnosis is seldom made preoperatively.

Treatment

Treatment of complications of the diverticulum are directed toward the control of the complication. In patients who have bleeding or symptoms that are related to the duodenal diverticulum, several operative procedures have been described. The most common and most effective treatment in this situation is diverticulectomy. This is most easily accomplished by performing a wide Kocher maneuver that exposes the diverticulum (Fig. 25-28). The diverticulum then is excised, and the duodenum is closed in a transverse or a longitudinal manner, whichever produces the least amount of luminal obstruction. For diverticula embedded deep within the head of the pancreas, lateral duodenotomy is performed; the diverticulum is invaginated into the lumen and excised, and the wall is closed. An alternative method that has been described for duodenal diverticula associated with the ampulla of Vater is an extended sphincteroplasty through the common wall of the ampulla and the diverticulum.

A perforated diverticulum may cause a great deal of trouble. When found, the perforated diverticulum should be excised and the duodenum closed with a serosal patch from the jejunal loop. If the inflammation is severe, it may be necessary to divert enteric flow away from the site of the perforation, with a gastrojejunostomy or, preferably, a duodenojejunostomy if possible. It may be possible to interrupt duodenal continuity proximal to the perforated diverticulum with a row of staples. Great care should be taken if the perforation is adjacent to the papilla. In one early perforation, we were able simply to invert the diverticulum, close the duodenal wall, and reinforce this with a serosal patch.

Because of the relationship of the common bile duct and pancreatic duct in patients with intraluminal diverticula associated with the ampulla of Vater, subtotal resection of the diverticulum should be carried out to protect the entry of the biliary-pancreatic ducts. If an intraluminal diverticulum arises at a site distant from the ampulla, complete excision may be possible.

The vast majority of duodenal diverticula are asymptomatic and benign, and when they found incidentally they should not be resected.

Jejunal and Ileal Diverticula

Diverticula of the jejunum and the ileum are much less common than duodenal diverticula, with an incidence between 0.5 and 1 percent on small- bowel x-ray examination. Jejunal diverticula are more common and are larger than those in the ileum (Fig. 25-29). Multiple diverticula are more common in the jejunum and the ileum than in the duodenum. These are false diverticula that usually protrude from the mesenteric border of the bowel, and they may be overlooked at operation because they are embedded within the small-bowel mesentery. When jejunal and ileal diverticula cause trouble, the symptoms usually are from incomplete bowel obstruction, acute diverticulitis, hemorrhage, or malabsorption due to bacterial overgrowth within the diverticulum. A specific syndrome of intestinal pseudoobstruction or jejunal dyskinesia is characterized by symptoms of intermittent partial bowel obstruction. On enteroclysis, barium may be seen to pass back and forth from the intestinal lumen into the diverticulum rather than moving normally through the bowel. This condition may be associated with hypertrophy and dilatation of the bowel proximal to the diverticulum. A recent study found that the condition may be associated with one of three syndromes: systemic sclerosis, visceral myopathy, or visceral neuropathy. If this condition is the only finding, resection of a large segment of jejunum containing diverticula should be avoided. Treatment of complications of obstruction, bleeding, and perforation is by intestinal resection and end-to- end anastomosis.

Obstruction also may be caused by enteroliths that form in a jejunal diverticulum and are subsequently dislodged and obstruct the distal intestine. This condition mainly occurs in the elderly and may be treated by enterotomy and removal of the enterolith. Patients with malabsorption from production of the blind loop syndrome by bacterial overgrowth within the diverticulum usually can be treated with antibiotics. Asymptomatic diverticula require no treatment.

Meckel's Diverticulum

Meckel's diverticulum, the most common true diverticulum of the gastrointestinal tract, is a congenital diverticulum that results from incomplete closure of the omphalomesenteric or vitelline duct. The usual manifestation is a relatively wide projection arising from the antimesenteric side of the ileum, generally 45 to 60 cm proximal to the ileocecal valve. Occasionally the diverticulum is merely a small bump or hillock, or it may be long and narrow (1–12 cm), but usually it is short and relatively wide (Fig. 25-30 A). Various forms exist, depending on the presence or absence of communication with the umbilicus or with various stages of atrophy of the omphalomesenteric ductal connection between the ileum and the umbilical skin (variations ranging from a persistent fibrous cord to a patent fistula) (Fig. 25-30 B–H). Heterotopic gastric or pancreatic tissue often is found in Meckel's diverticula. The vast majority of Meckel's diverticula are entirely benign. Trouble arises from bleeding or inflammation, which gives rise to a symptom complex usually indistinguishable from appendicitis; from perforation, usually caused by peptic ulcer from acid and pepsin secretion from ectopic gastric fundic mucosa; and from problems associated with persistence of part or all of the omphalomesenteric duct. The great majority of symptomatic Meckel's diverticula are found in childhood, and the most common symptom in childhood is bleeding. On the basis of several autopsy series, Meckel's diverticula has a reported incidence in the general population of 2 percent. In adults most Meckel's diverticula are found incidentally by radiographic examination of the small bowel. Enteroclysis usually provides an accurate diagnosis. Another technique for detection of Meckel's diverticulum in symptomatic patients is radionuclide scanning with ^{99m}Tc-pertechnetate. The basis for this is the uptake of radioisotope by the heterotopic gastric mucosa within the diverticulum. The diagnostic accuracy of this scanning technique may be increased by administration of pentagastrin (6 mg/kg subcutaneously) 15 min before the scan, which enhances the uptake of the radioisotope by the gastric mucosa. It has been shown, however, that the parietal cells do not specifically accumulate pertechnetate and are not essential for detection purposes. The scan far more useful in children than in adults.

Complications of Meckel's diverticulum in adults include intestinal obstruction, bleeding, acute diverticulitis, or the presence of a diverticulum in a hernia sac (Littre's hernia). Obstruction may be produced by one of two mechanisms. The most common is volvulus or kinking around a band running from the tip of the diverticulum to the umbilicus, abdominal wall, or mesentery. The diverticulum also may cause obstruction by intussusception. Bleeding is the second most common complication and usually is found only when there is heterotopic gastric mucosa within the diverticulum. The bleeding ulcer is found not in the diverticulum but in the ileum adjacent to the diverticulum.

Meckel's diverticulitis, which is clinically indistinguishable from appendicitis, is the third most common complication in adults. The incidence of perforation or peritonitis with Meckel's diverticulitis is about 50 percent.

A Meckel's diverticulum should be considered in the differential diagnosis of patients who present with a mechanical bowel obstruction, with low small- bowel hemorrhage, or with signs and symptoms of inflammation or peritonitis. Treatment is prompt surgical intervention, with resection of the diverticulum or resection of the segment of ileum bearing the diverticulum (Fig. 25-30 I–J). Segmental intestinal resection is required for treatment of patients with bleeding, because the bleeding site usually is in the ileum adjacent to the diverticulum. Recent reports have demonstrated the feasibility and safety of laparoscopic diverticulectomy; long-term outcome with this procedure, however, are lacking.

It is generally recommended that asymptomatic diverticula found in children during laparotomy be resected; however, the treatment of an asymptomatic Meckel's diverticulum found incidentally at laparotomy in the adult patient is controversial. In a landmark paper that has greatly influenced the surgical management of the asymptomatic diverticulum, Soltero and Bill estimated that the likelihood of a Meckel's diverticulum becoming symptomatic in an adult is 2 percent or less and that morbidity from incidental removal (reported to be as high as 12 percent in some studies) far exceeds the potential for prevention of disease. From this study, the authors concluded that an asymptomatic Meckel's diverticulum in the adult should not be removed. This study has been criticized, however, since it was not a population-based analysis. An epidemiologic, population-based study by Cullen and associates has challenged the practice of ignoring an incidentally found Meckel's diverticulum. A 6.4 percent rate of developing complications from the Meckel's diverticulum was calculated to occur over a lifetime; this incidence of complications does not appear to peak during childhood as originally thought. These authors therefore recommended that an incidentally discovered Meckel's diverticulum should be removed at any age up to 80 years, as long as no additional condition (e.g., peritonitis) makes removal hazardous. The rates of short-term and long-term postoperative complications from prophylactic removal are low (approximately 2 percent), and in the Cullen series mortality was related to the primary operation or the general health of the patient and not to the diverticulectomy. Therefore, this study as well as other recent studies suggest that prophylactic diverticulectomy in selected patients may be beneficial and safer than originally reported.

Miscellaneous Problems

Congenital Abnormalities

Although primarily limited to infants and children, developmental anomalies of the small intestine also may present in adulthood. In addition to Meckel's diverticulum, discussed above, intestinal duplications and anomalies of gut rotation and fixation may be manifested in the adult.

The term *duplication* actually is a misnomer, since approximately 75 percent are enteric cysts and not true duplications. The remainder are true tubular duplications that are intimately attached to the adjacent portion of the gastrointestinal tract, possess a gastrointestinal type of mucosal lining, and may or may not communicate with the visceral lumen. Duplications occur on the mesenteric side of the intestine between the leaves of the mesentery and are most often found in the ileum. In rare instances, a duplication from the abdomen extends into the chest. Cystic lesions may present with symptoms of obstruction from extrinsic compression. Communicating tubular duplications often are lined by gastric mucosa, which may ulcerate and present with gastrointestinal hemorrhage or perforation. In addition, cancers have been reported to arise in the mucosal lining of these duplications. Treatment is surgical resection. Small tubular duplications and all cysts are excised along with overlying normal intestine; intestinal continuity is restored by end-to-end anastomosis. Long tubular duplications that would require massive small-bowel resection for complete removal are more problematic and may be treated by mucosal stripping of the duplication.

Various anomalies may occur from defects in gut rotation. Failure of the midgut to rotate results in the duodenum's descending to the right of the midline with the entire small bowel lying in the right half of the abdomen and the colon lying to the left. Patients usually are asymptomatic. Symptoms from malrotation result from the midgut's failing to rotate completely and usually occur in patients 2 months of age or less; however, symptoms of midgut volvulus occur in 10 percent of patients at a later age and present either acutely or

with chronic symptoms of intermittent partial obstruction. Partial obstruction of the duodenum by peritoneal bands (Ladd's bands) may cause bilious vomiting and intermittent cramping abdominal pain. Mesocolic hernias (also referred to as paraduodenal hernias) occur when failure of proper gut fixation results in the small bowel's being trapped within the mesentery of the colon. The most common presentation is that of obstruction. Treatment for anomalies of rotation and fixation include lysis of Ladd's bands (if present), reduction of the volvulus, and repositioning of the intestine to prevent twisting.

Small-Bowel Ulcerations

The majority of ulcerations of the small bowel have definable causes, which include the following: drug-induced (enteric-coated potassium chloride tablets or corticosteroids), vascular (occlusion or vasculitis), Crohn's disease, syphilis, typhoid fever, tuberculosis, lymphoma, heterotopic gastric mucosa (Meckel's diverticulum), and ulcers associated with gastrinoma. This discussion is limited to ulcerations of the small intestine in which no causative agent can be identified.

Patients with discrete isolated ulcers of the small bowel and without any identifiable underlying disease have nonspecific ulcers. Ulcers are more common in the terminal ileum. These appear to be self-limited and do not recur after bowel resection.

These patients usually present with a single ulceration, although multiple ulcers may be present. Indications for operation are for complications of the ulcers, including perforation, bleeding, or stricture. Recurrence of ulceration in the small bowel distal to the duodenum is rare. A review of 59 patients studied from 1956 to 1979 reports the overall incidence of small-bowel ulcerations to be 4 per 100,000 new patients seen at the Mayo Clinic. These investigators noted that the yearly rate fell from 3.6 new cases per year from 1960 to 1969 to 1.2 cases per year from 1970 to 1979. They believe, despite the absence of a documented underlying cause, that the decrease in incidence was directly related to the removal of enteric-coated potassium chloride tablets from the market.

The ulcers are discrete, varying from 0.3 to 5 cm in diameter; they are sharply demarcated, and the surrounding mucosa is perfectly normal. They occur more frequently on the antimesenteric border and may be associated with fibrous scar formation, which produces the obstruction. The characteristic microscopic findings include acute granulation tissue and inflammatory cells at the base of the ulcer, local hyperplasia of the muscularis mucosae, and pyloric metaplasia of the adjacent mucosa. Varying degrees of edema and fibrosis, depending on the chronicity of the ulcer, are noted. There is no evidence of vascular disease associated with the ulcer. With time, the ulcers may increase in size, may become annular, and, with healing, may produce fibrous strictures of the intestine, which may produce obstruction.

Diagnosis is rarely if ever made before operation; the majority of patients present with complications. In the Mayo Clinic series 63 percent of patients had intermittent small-bowel obstruction, 25 percent had bleeding, and 12 percent had symptoms of acute abdominal inflammation caused by perforation. With the increasing use of enteroclysis, the diagnosis of small-bowel ulcer may be made more frequently. An asymptomatic solitary ulcer does not require treatment.

Treatment

The treatment of small-bowel ulcers depends on the complications encountered at the time of diagnosis. Mechanical bowel obstruction and bleeding should be treated by segmental resection. Although excision and primary closure of perforated ulcers has been advocated by some, high recurrence rates have been associated with this technique. The ulcerated segment of bowel should be resected.

Ingested Foreign Bodies

A great variety of objects that are capable of penetrating the wall of the gut are swallowed, usually accidentally, but sometimes intentionally by the mentally deranged. These include glass and metal fragments, pins, needles, cocktail toothpicks, fish bones, coins, whistles, toys, and broken razor blades.

Treatment is observation, since the vast majority pass without difficulty. If the object is radiopaque, progress can be followed by serial plain films. Catharsis is contraindicated.

Sharp, pointed objects such as sewing needles may penetrate the bowel wall. If abdominal pain, tenderness, fever, or leukocytosis occurs, immediate surgical removal of the offending object is indicated. Abscess or granuloma formation is the usual outcome without surgical therapy.

Small-Bowel Fistulas

The vast majority of small-bowel fistulas are a result of operation; less than 2 percent are associated with granulomatous disease of the bowel (Crohn's disease) or trauma. In some patients there are contributing factors, such as preoperative radiotherapy, intestinal obstruction, inflammatory bowel disease, mesenteric vascular disease, and intraabdominal sepsis. But in the majority, surgical misadventures are the primary cause. These include anastomotic leak, injury of bowel or blood supply at operation, laceration of bowel by wire mesh or retention sutures, and retained sponges. In addition, fistulas may result from erosion by suction catheters and by abscesses.

The major complications associated with small-bowel fistulas include sepsis, fluid and electrolyte depletion, necrosis of the skin at the site of external drainage, and malnutrition. Successful management of patients with small-bowel fistulas requires meticulous attention to detail and a logical, stepwise plan of management. One must establish controlled drainage, manage sepsis, prevent fluid and electrolyte depletion, protect the skin, and provide for adequate nutrition.

Mortality for patients with intestinal fistulas remains high (20 percent or greater), even with the use of total parenteral nutrition (TPN). Although TPN has not been shown to reduce mortality significantly, it is the single most important advance in the management of patients with enterocutaneous fistulas. Fluid, electrolyte, and nutritional status may be maintained from the time the fistula becomes apparent and throughout the time required for control of sepsis. The key to successful management of intestinal fistulas is control of sepsis and prevention of malnutrition.

Diagnosis of small-bowel fistula usually is not difficult. When the damaged area of the small bowel breaks down and discharges its contents, dissemination may occur widely in the peritoneal cavity, producing generalized peritonitis. More commonly, however, the process is more or less walled-off in the immediate area of the leak, with formation of an abscess. This usually underlies the operative incision, so that when a few skin sutures are removed to ascertain why the incision is becoming red and tender, contents of the abscess are discharged and the fistula established. The discharge initially may be purulent or bloody, but this is followed—sometimes immediately, sometimes within a day or two—by drainage of obvious small-bowel contents. If the diagnosis is in doubt, confirmation can be obtained by oral administration of a nonabsorbable marker such as charcoal or Congo red.

Small-bowel fistulas are classified according to their location and volume of daily output, since these factors dictate treatment as well as morbidity and mortality rates. In general, the more proximal the fistula in the intestine, the more serious the problem. Proximal fistulas have a greater fluid and electrolyte loss, the drainage has a greater digestive capacity, and an important (distal) segment is not available for food absorption. High-output fistulas are those that discharge 500 mL or more per 24 h. It is important, therefore, as soon as the patient's condition is stabilized, to identify the site of the fistula, to determine the extent of the associated abscess cavity by fistulogram, and to ascertain whether there is distal obstruction, since fistulas will not close in the presence of distal obstruction (Fig. [25-31](#)). Upper gastrointestinal series with small-bowel follow-through and barium enema studies also may provide useful information.

Treatment

Control of sepsis is aided by sump suction, which provides drainage of the associated intraabdominal abscess cavity and prevents accumulation of intestinal contents. Control of fistula output is most easily accomplished by percutaneous intubation of the fistula tract. Protection of the skin around the fistulous opening is important. In the past, frequent applications of zinc oxide, aluminum paste ointment, or karaya powder were required; excoriation and destruction of skin still occurred. The advent of Stomahesive appliances used for colostomy and ileostomy bags greatly improved and facilitated protection of the skin at the site of the fistula. The Stomahesive appliance should be cut so that the opening just fits over the fistulous opening and no unprotected skin remains. The suction catheter can be brought out through the end of the bag, which is fixed firmly about the tube. This allows for collection of all the drainage and accurate quantitation of the lost volume.

The volume depletion that occurs from a proximal small-bowel fistula may present a formidable problem. Volume losses exceeding 5 L/day are not uncommon. Agents that inhibit gut motility (codeine, Lomotil, or loperamide) generally are not helpful. The hormone somatostatin is considered a general "off switch" and inhibits both intestinal secretion and motility. We and others have used a long-acting octapeptide analog of somatostatin (octreotide) in combination with TPN in the management of enterocutaneous fistulas. In most series, the overall fistula closure rate was not altered, but the analog ameliorated the problems associated with massive volume loss. Systemic antibiotics should be administered until sepsis is controlled. TPN should be instituted at the same time, because a prolonged course of inability to use the gut for nutrition is likely.

Several factors may prevent spontaneous closure of fistulas. Fistulas will not close spontaneously if there is high output (> 500 mL/24 h) or severe disruption of intestinal continuity (> 50 percent of the circumference of the bowel involved in the fistula). A fistula will not close spontaneously if it arises from a segment of bowel involved with active granulomatous disease, cancer, or radiation enteritis, if there is distal obstruction, or if there is an undrained abscess cavity. If a foreign body is in the tract, if the fistula tract is less than 2.5 cm in length, or if there is epithelialization of the tract, spontaneous closure will not occur. Radiographic investigation of the fistula by injection of water-soluble contrast material through the fistula tract should be carried out early to delineate the presence and extent of any abscess cavities and to obtain information about the length of the tract and the extent of bowel wall disruption. A diligent search by means of contrast studies for distal obstruction should be performed. CT will often reveal undrained collections of fluid.

When any of the conditions noted above are present, spontaneous closure is unlikely, and therefore management should be directed toward obtaining prompt control of sepsis, maintaining positive nitrogen balance, and early operation.

Conservative treatment for up to 3 months with TPN has been advocated by some to allow spontaneous closure of the fistula. We do not believe that the results support this approach. Fewer than 30 percent of all small-bowel fistulas will close spontaneously. In patients with low-output fistulas, particularly those located in the distal small bowel without any of the conditions that prevent spontaneous closure, a wait of up to 6 weeks may be indicated. The patient usually can wait at home. When we reviewed reported series that advocate conservative therapy for longer than 6 weeks, we found that the majority of fistulas that close spontaneously do so within 3 weeks after their appearance. After 3 weeks, if sepsis has been controlled and adequate nutritional status has been achieved, operative control of the fistula should be carried out promptly. Delay only produces delay. TPN simplifies management of patients, but it does not cure fistulas. The single most important determinant in successful treatment of fistulas is sepsis. If sepsis is not controlled, the patient will die. After sepsis has been controlled, one should not wait endlessly for a fistula to close spontaneously simply because malnutrition can be avoided by use of TPN. The proper role of TPN is prevention or treatment of malnutrition before operative closure of fistulas.

Operation is most easily accomplished by entering the previous abdominal wound. The wound should be reopened with great care to avoid needless reinjury to the bowel. The fistula tract is excised, the bowel should be completely mobilized, and the portion of bowel involved in the fistula is resected. The technique of excision and fistula closure must be precise and accurate, and all rigid or diseased bowel must be resected. Simple closure of the fistula after removing the fistula tract and minimal mobilization of the bowel almost always results in recurrence of the fistula.

If an unexpected abscess is encountered or if the bowel wall is rigid and distended over a large distance, a proximal enterostomy should be performed. Later, resection of the bowel involved in the fistula will be required for successful closure. Side-to-side bypass should not be done.

The overall mortality rate in enterocutaneous fistulas of the small bowel is greater than 20 percent. It is higher in jejunal fistulas and significantly lower in ileal fistulas. Successful treatment of the majority of patients with small-bowel fistulas requires control of sepsis, provision of adequate nutrition, and operative closure.

Pneumatosis Cystoides Intestinalis

This is an uncommon condition manifested by multiple gas-filled cysts of the gastrointestinal tract. The cysts are either submucosal or subserosal and vary in size from microscopic to several centimeters in diameter. The jejunum is most frequently involved, followed by the ileocecal region and the colon; extraintestinal structures such as mesentery, peritoneum, and the falciform ligament also may be involved. Gas cysts are associated with other lesions of the gastrointestinal tract in about 85 percent of cases. These “secondary” causes include inflammatory, obstructive, or infectious conditions of the intestine, iatrogenic conditions such as endoscopy and jejunostomy placement, ischemia, and extraintestinal diseases such as chronic obstructive pulmonary disease, diabetic enteropathy, and collagen vascular diseases. Pneumatosis not associated with other lesions (15 percent of cases) is referred to as primary pneumatosis.

Grossly, the cysts resemble cystic lymphangiomas or hydatid cysts. On section, the involved portion has a honeycomb appearance. The cysts are thin-walled and break easily. Spontaneous rupture gives rise to pneumoperitoneum.

Symptoms are nonspecific and in secondary pneumatosis may be those of the associated disease. Symptoms in primary pneumatosis, when present, resemble those of irritable bowel syndrome. The diagnosis usually is made radiographically by plain abdominal or barium studies (Fig. 25-32). Diagnosis also may be made by CT. No treatment is necessary unless one of the very rare complications supervenes, such as rectal bleeding, cyst- induced volvulus, or tension pneumoperitoneum. Prognosis in most patients is that of the underlying disease. When pneumatosis occurs in infants with necrotizing enterocolitis, it does not make the outlook any worse. The cysts may disappear spontaneously or may persist for prolonged periods without serious symptoms.

Blind Loop Syndrome

This is a rare clinical syndrome manifested by diarrhea, steatorrhea, anemia, weight loss, abdominal pain, multiple vitamin deficiencies, and neurologic disorders. The underlying cause is not a blind loop per se, but bacterial overgrowth in stagnant areas of small bowel produced by stricture, stenosis, fistulas, blind pouch, or diverticula (Table 25-5). The bacterial flora are altered in the stagnant area, both in number and in kind. Bacteria compete successfully for vitamin B 12, producing a systemic deficiency of B 12 and megaloblastic anemia. Steatorrhea also occurs; bacteria in the stagnant area deconjugate bile salts, causing disruption of micellar solubilization of fats. There also may be absorptive defects of other macronutrients and micronutrients, probably caused by direct injury of the mucosal cells.

The syndrome can be confirmed by a series of laboratory investigations. First, a Schilling test (60Co-labeled vitamin B12absorption) is performed; this should reveal a pattern of urinary excretion of vitamin B 12resembling that of pernicious anemia (that is, a urinary loss of 0 to 6 percent of vitamin B 12, compared to the normal of 7 to 25 percent). The test is then repeated with the addition of intrinsic factor. In true pernicious anemia, the excretion should rise to normal; in the blind loop syndrome, the addition of intrinsic factor will not increase the excretion of vitamin B 12. Next, the patient is given a course of tetracycline for 3 to 5 days, and the Schilling test is repeated. With blind loop syndrome, absorption of 60Co-labeled vitamin B 12 returns to normal; this does not occur in the macrocytic anemia due to steatorrhea. Patients with the blind loop syndrome respond to tetracycline and parenteral vitamin B 12therapy. Medical treatment is not definitive but should be employed to prepare patients for operation. Surgical correction of the condition producing stagnation and blind loop syndrome effects a permanent cure and is indicated.

Radiation Enteritis

Radiation is a common adjuvant therapy in the multimodality approach to the treatment of various abdominal and pelvic cancers. Along with the beneficial effects gained by killing tumor cells, surrounding normal tissues such as the small intestine may sustain severe acute and chronic deleterious effects. The intestinal epithelium is the most radiosensitive tissue of the gut. The amount of radiation required to produce clinical signs of enteritis varies with the treatment regimen as well as the patient. Serious morbidity is unusual if the total radiation dosage is less than 4,000 cGy; however, the complication rate increases when doses exceed 5,000 cGy. In addition, previous abdominal operations, preexisting vascular disease, hypertension, diabetes, and treatment with certain chemotherapeutic agents (e.g., 5-fluorouracil, doxorubicin, actinomycin D, and methotrexate) increase the likelihood of enteritis after radiation treatments.

The mucosa and the submucosal vasculature of the intestine are the most sensitive to radiation damage. With the relatively rapid turnover of the intestinal mucosa, radiation damage to this layer tends to be acute and self-limiting, with symptoms consisting mainly of diarrhea, abdominal pain, and malabsorption. Late effects of radiation are the result of damage to the small submucosal vessels, with a progressive obliterative arteritis and submucosal fibrosis, resulting eventually in thrombosis and vascular insufficiency. This injury may produce necrosis and perforation of the involved intestine, but more commonly it leads to strictures, with symptoms of obstruction, or fistula formation.

Radiation enteritis may be minimized by adjusting ports and dosages of radiation to deliver optimal treatment specifically to the tumor and not to surrounding tissues; this could be facilitated by placement of radiopaque markers, if laparotomy is performed, to delineate the area to be irradiated. Prophylactic measures to help decrease the incidence of enteritis include methods to exclude small bowel from the pelvis such as reperitonealization, omental transposition, or placement of absorbable mesh slings.

Treatment of the acute sequelae of radiation injury may be obtained by conservative measures such as the administration of antispasmodics, analgesics, and antidiarrheal agents. Agents that may prove useful in the treatment of acute sequelae include glutamine-enriched enteral formulas and the gut hormone bombesin, which ameliorates methotrexate-mediated enterocolitis in experimental animals.

Operative intervention may be required for symptomatic strictures or fistulas. Caution must be exercised when operating on patients with previously irradiated bowel, since the vascular injury may be widespread and not readily recognizable by gross inspection of the intestine; if resection and reanastomosis are performed, dehiscence of the anastomosis can occur. Extensive adhesiolysis should be avoided if possible. Perforation of the intestine should be treated with resection and anastomosis; when reanastomosis is thought to be unsafe, ostomies should be created. Frozen sections and laser Doppler flowmetry have been used to assist resection and anastomosis; however, the clinical usefulness of these techniques has been debated. Obstruction caused by a rigid, fixed pelvis incorporating small intestinal loops is best treated by bypass.

Radiation enteritis is a relentless disease process reflecting widespread bowel involvement. Almost one-half of patients who survive their first laparotomy for radiation bowel injury require further surgery for ongoing bowel damage. Up to 25 percent of these patients die from radiation enteritis and complications from its management.

Short Bowel Syndrome

In the adult, emergency massive resection of the small bowel must sometimes be done when extensive gangrene precludes revascularization. Mesenteric occlusion, midgut volvulus, and traumatic disruption of the superior mesenteric vessels are the most frequent causes. Short bowel syndrome may result from such massive resections; it also may be produced by several bowel resections in patients with severe recurrent Crohn's disease. In neonates the most common cause of short bowel syndrome is necrotizing enterocolitis.

The short bowel syndrome is a group of signs and symptoms that result from a total small bowel length that is inadequate to support nutrition. The clinical hallmarks of the short bowel syndrome include diarrhea, fluid and electrolyte deficiency, and malnutrition. The small bowel has two primary functions, digestion and absorption of nutrients. Problems that result from extensive resection of the small bowel can be divided into two types: those related to the extent of small-bowel loss, and those related to the specific area of bowel removed.

Although there is considerable individual variation, resection of up to 70 percent of the small bowel usually can be tolerated if the terminal ileum and ileocecal valve are preserved. Length alone, however, is not the only determining factor of complications of small-bowel resection. For example, if the distal two-thirds of the ileum, including the ileocecal valve, is resected, significant abnormalities of absorption of bile salts and vitamin B 12 may occur, although only 25 percent of the total length of the small bowel has been removed. Proximal bowel resection is tolerated much better than distal resection, since the ileum can adapt and increase its absorptive capacity more efficiently than the jejunum.

Digestion and absorption in the small bowel depend on the presence of brush border enzymes, an adequate number of enterocytes for absorption, and normal intestinal motility. With massive resection of the small bowel, there is reduced absorption of all nutrients, including electrolytes, water, carbohydrates, protein, fat, trace elements, and vitamins. The proximal small bowel is the primary site of absorption of iron, folate, and calcium, whereas the distal small bowel is the site of absorption of bile salts and vitamin B 12.

The bowel has an intrinsic capacity to adapt after small-bowel resection, and in many instances this process of intestinal adaptation effectively prevents severe complications resulting from the reduced surface area available for absorption and digestion. Any adaptive mechanism can be overwhelmed; maximum adaptation will be inadequate if too much small bowel is lost. Intestinal adaptation is characterized by an increased absorptive surface area as a result of hyperplasia of the remaining enterocytes. The villi lengthen (but do not increase in number), more cells are produced, and there are increased cell renewal and migration to the villous tip. All this serves to produce an increase in total absorptive surface area. Although there are more cells, individual cells do not increase their life span (they must migrate farther) or their capacity to synthesize digestive enzymes or to increase absorptive processes, so the overall net increase in digestive efficacy is not great (Fig. 25-33).

The mechanisms responsible for intestinal adaptation have been studied widely in animals. Multiple factors are responsible and are required for development of successful intestinal adaptation. For unknown reasons, the ileum exhibits a much greater adaptive response than the jejunum. Luminal nutrients, trophic gut hormones, specific amino acids (e.g., glutamine), and pancreatic and biliary secretions are all required for complete adaptation to occur. In animals maintained on TPN after extensive small-bowel resection, nutrition may be maintained but intestinal adaptation does not occur. Enteral diets containing glutamine, the principal fuel used by the small intestine, have been shown to increase mucosal proliferation and gut regeneration in experimental models. In addition, dietary fiber (e.g., pectin) has been shown to enhance mucosal adaptation, increase stool acidity, and improve colonic water absorption. The trophic gut peptides studied most intensively in small-bowel adaptation are gastrin, CCK, secretin, neurotensin, peptide YY, and bombesin. We now know that, although hypergastrinemia is associated with the short bowel syndrome, gastrin does not play a major role in the adaptive response after resection. CCK and secretin may have direct effects on enterocyte replication; however, their primary importance in intestinal adaptation may be in stimulation of pancreaticobiliary secretions rather than in directly stimulating enterocyte proliferation. The small-bowel mucosa is a complex and highly proliferative tissue with multiple factors regulating growth. The gut hormone neurotensin appears to be one of the most potent enterotrophic factors identified to date. Administration of neurotensin stimulates small-bowel mucosal growth in rats fed a normal diet, prevents gut mucosal hypoplasia induced by feeding rats an elemental diet, and augments intestinal regeneration after small-bowel resection. Recent studies demonstrate a marked mucosal growth response by the gut-derived peptide glucagonlike peptide-2 (GLP-2), which may hold promise for patients with short bowel syndrome. Two other hormones, not derived from the gut, that have been evaluated extensively in various experimental and in limited clinical trials include growth hormone and insulinlike growth factor-1 (IGF-1).

Growth hormone enhances mucosal hyperplasia in both the remnant small bowel and colon. IGF-1, which is regulated by growth hormone, increases the weight and length of the small and large intestines. Results reported recently by Byrne and colleagues using a combination of growth hormone, glutamine, and a modified diet appear promising in allowing for a reduction or elimination of TPN requirements in some refractory patients with severe short bowel syndrome and TPN dependence.

In human beings, the adaptive responses to massive resection have been found to be increased caliber of the remaining small bowel, hypertrophy of the gut wall, increased villous height, and increase numbers of enterocytes. This process often takes weeks or months to complete after small-bowel resection. With time, absorptive function increases. This increase is characterized by decreasing stool losses of water and electrolytes and increased absorption of glucose and vitamin B₁₂.

Hypergastrinemia and gastric hypersecretion occur after massive small-bowel resection and have been widely studied in experimental animals; some information has come from human subjects. Diarrhea associated with gastric hypersecretion is caused by (1) delivery of a massive volume of fluid and electrolytes to the shortened small bowel, (2) steatorrhea from failure of lipolysis by pancreatic lipase, which requires an intraluminal pH greater than 5.0 for activity (pancreatic secretion of lipase is not affected), and (3) acid enteritis. Although acid hypersecretion was at one time thought to be of prime importance in producing diarrhea after extensive small-bowel resection, both hypergastrinemia and hypersecretion of acid are transient. Acid hypersecretion is now easily managed by H₂-receptor antagonists or proton-pump blockers. Control of acid secretion controls diarrhea to a great extent during the early phase. Several operations, including vagotomy with pyloroplasty, or vagotomy with antrectomy, have been employed in treatment of the short bowel syndrome to control acid hypersecretion. Since the problem is self-limited, however, these procedures are not indicated and should not be done.

Resection of specific segments of the small bowel leads to specific problems. Resection of the distal small bowel results in diarrhea, steatorrhea, and malabsorption. Conjugated bile salts, essential for normal fat absorption, are almost totally absorbed in the distal ileum by active transport mechanisms. Resection of the ileum results in disruption of the enterohepatic circulation of bile salts and may lead to two types of diarrhea. When less than 100 cm of small bowel has been resected, excessive amounts of bile salts enter the colon and produce a chemical enteritis; this type of diarrhea has been termed *cholerheic*. The toxic effects of bile acids on colonic epithelial cells are twofold: bile salts inhibit absorption of water and electrolytes in the colon, and the injured colonic cells secrete excessive amounts of water and electrolytes. The response to decreased absorption of bile salts is increased hepatic production of bile salts, and this leads to perpetuation of diarrhea.

When more than 100 cm of ileum has been resected, the loss of bile salts is so great that hepatic synthesis cannot compensate. Furthermore, decreased intraluminal bile acids stimulate CCK release, and CCK stimulates gut motility. In addition to the direct toxic effects of bile salts on the colonic epithelium, fat malabsorption (steatorrhea) occurs. Differentiation of the two types of diarrhea associated with distal resection is important, because treatment is different. Measurement of stool fat content, vitamin B₁₂ absorptive capacity, and fecal bile salt concentrations is important for accurate determination of the deficits produced. For patients who have cholerheic diarrhea, agents that bind bile acids (e.g., cholestyramine) may alleviate diarrhea. If steatorrhea is present, then medium-chain triglycerides that do not require micelle formation for absorption also should be used.

Another factor contributing to diarrhea after ileal resection is loss of the ileocecal valve. The ileocecal valve has two important actions. It prolongs intestinal transit time, and it prevents retrograde passage of colonic bacteria into the small bowel, which, if not prevented, causes bacterial enteritis.

Other complications associated with alteration of enterohepatic circulation of bile acids include gallstones and anemia. The changes in the bile salt pool produce lithogenic bile; the incidence of gallstones in patients with ileal resection is three to four times greater than in the normal population. The ileum is the specific site for transport mechanisms for intrinsic-factor-mediated vitamin B₁₂ absorption, and with total ileal resection, stores of vitamin B₁₂ are depleted and anemia results.

Treatment

The most important principle in the treatment of short bowel syndrome is prevention. This means that at operation, when intestinal viability is questionable, the smallest possible resections should be performed, and “second-look” operations 24 to 48 h later should be carried out to allow the ischemic bowel to demarcate. Delay may prevent unnecessary, extensive resection of bowel. In patients with Crohn's disease, the devastating complications of the short bowel syndrome have led all students of the disease to recognize that only limited resections should be performed.

After massive small-bowel resection, the program for treatment may be properly divided into early and late phases. Early on, treatment is primarily directed at the control of diarrhea, replacement of fluid and electrolytes, and the prompt institution of total parenteral nutrition. Volume losses may exceed 5 L/day, and vigorous monitoring of intake and output with adequate replacement must be carried out. Depletion of fluid volume caused by diarrhea, especially in the early phase, is often a formidable problem. Judicious use of agents that inhibit gut motility (codeine, Lomotil, loperamide) may be helpful. These drugs may cause profound ileus if used excessively, and volume lost through the gut is simply traded for volume lost through nasogastric suction. In addition, prolonged ileus with dilatation and edema of the bowel wall may result.

As intestinal adaptation progresses and gut absorption increases, the stool volume gradually decreases. Once patients have completely adapted to an oral diet, semiformal stools may appear, but these patients will likely never have the normal number or consistency of stools.

As soon as the patient has recovered from the acute phase, enteral nutrition should begin, so that intestinal adaptation may begin early and proceed successfully. The most common types of enteral diets are elemental (Vivonex, Flexical) or polymeric (Isocal, Ensure). Each presents problems with increased osmolality and may contain foodstuffs that may not be absorbed because of enzyme deficiency (for example, lactase deficiency). Milk products should be avoided, and diets should be begun at isoosmolar concentrations and with small volumes (50 mL/h), even though the full nutrient value may not be obtained. As the gut adapts, the osmolality, volume, and caloric content may be increased. As previously noted, the combination of glutamine, growth hormone, and a modified oral diet may prove efficacious at this stage. Gut hormones in combination with dietary modifications may prove to be useful in the future.

Reduction of dietary fat has long been considered to be important in the treatment of patients with the short bowel syndrome. High-carbohydrate, high-protein, low-fat diets have been prescribed. Fat has more than twice as many calories per gram as protein and carbohydrate and is important for maintenance of proper nutrition. Supplementation of the diet with 100 g or more of fat should be carried out. Often this requires the use of medium-chain triglycerides, which may be absorbed in the proximal bowel without micelle formation. Vitamins, especially fat-soluble vitamins, as well as calcium, magnesium, and zinc supplementation, also must be provided. H₂-receptor antagonists or proton-pump blockers may greatly diminish the diarrhea that is largely caused by the early, transient acid hypersecretion. Measurement of intragastric pH can be used to guide the dose of drugs required. Antacids are not useful, because they may aggravate diarrhea or bind essential ions. In no case should gastric resection or vagotomy be used for the treatment of the short bowel syndrome.

Since the dysfunction of massive small-bowel resection is caused by decreased absorptive surface and rapid transit time, most attempts at surgical treatment have been directed toward increasing the absorptive surface or slowing the intestinal transit time. The intestinal tapering and lengthening procedure, originally described by Bianchi, improves intestinal function by correcting the dilatation and ineffective peristalsis of the intestine as well as by prolonging intestinal transit time by doubling the intestinal length while preserving the mucosal surface area. Dissection is performed longitudinally on the mesenteric edge of the bowel to create a space that permits longitudinal division of the bowel with a stapling device. This procedure may be beneficial in selected patients; however, potential complications include necrosis of divided segments and anastomotic leak. Procedures to slow intestinal transit time include reversed intestinal segments, circular loops, colon interposition, and construction of valves (artificial sphincters). These procedures have received limited clinical evaluation; most reports are only anecdotal, and the efficacy of these techniques remains questionable. In fact, some reports actually show a decreased effective functional length and an increased risk of bowel infarction with these procedures. Therapy for the short bowel syndrome may be revolutionized when small-intestine transplantation becomes a less morbid procedure. The greatest obstacles to successful bowel transplantation have been rejection and sepsis.

In summary, the present treatment of the short bowel syndrome is palliative and is directed toward control of diarrhea and prevention of dehydration and malnutrition. In the future, use of glutamine, growth factors, and gastrointestinal hormones may be useful adjuncts in certain patients; however, only with the development of successful allotransplantation of the gut will the short bowel syndrome be cured.

Intestinal Bypass

Morbid Obesity

Surgical procedures to treat morbid obesity (defined as more than 100 lb over ideal weight) have become popular because the long-term success rate of nonsurgical treatment of this condition is only 1 percent. There are two basic approaches: (1) restrict the amount of food entering the gastrointestinal tract (gastric reduction) or (2) reduce the amount of food absorbed (gut shortening). The original procedure designed to create a short-gut malabsorption syndrome, jejunocolostomy, had to be abandoned because of an unacceptable rate of complications. It was succeeded by jejunoileostomy, either end-to-side (Payne procedure) or end-to-end (Scott procedure). Many thousands of these procedures have been done in the United States. These too have been abandoned because of long-term complications, which included persistent diarrhea, profound electrolyte abnormalities, arthralgias, neurologic symptoms, enteropathies, cholelithiasis, renal disease, and hepatic failure. In addition, the long-term weight control was poor. An alternative method of operative therapy of obesity is the gastric bypass or partition. This method limits food intake by reducing the reservoir capacity of the stomach to 5 to 10 percent of normal. After a Consensus Development Conference at the National Institutes of Health in March 1991, the only procedures currently recommended for evaluation are gastric reductions.

Hyperlipidemia

Surgical bypass of a portion of the small intestine is a useful method of treating hypercholesterolemia and hypertriglyceridemia. The operation, designed by Buchwald and Varco, bypasses either the distal 200 cm or the distal one-third of the small intestine, whichever is greater (Fig. 25-34). This operation, though occasionally associated with diarrhea, does not cause significant weight loss and is not associated with the undesirable side effects of the jejunoleal bypass.

This procedure lowers serum cholesterol level through two mechanisms: the absorption of cholesterol is reduced by bypassing the usual site of absorption, and cholesterol and bile acid excretion is increased, which accelerates cholesterol turnover.

Clinical metabolic studies have demonstrated a 60 percent decrease in cholesterol absorption, a 40 percent reduction in serum cholesterol levels, and a reduction of more than 50 percent in plasma triglyceride levels. About 70 percent of patients with angina have had improvement or total remission of symptoms after this operation.

Results have been reported from the long-term follow-up of the first group of patients studied as well as from the Program on the Surgical Control of the Hyperlipidemias (POSCH), a prospective, randomized, multicenter trial. Of the original 57 patients, partial ileal bypass produced significant sustained (longer than 20 years) reduction in total plasma cholesterol levels. Only 3.5 percent of patients required reversal of the bypass; however, 40 percent developed nephrolithiasis. These findings were confirmed in the POSCH trial. The POSCH trial data also have shown an apparent reversal of coronary artery atherosclerosis and reduction of mortality rates in some subgroups of patients. The response of the 78 women in the POSCH trial was similar to that of the group as a whole, although small numbers made statistical significance difficult to achieve. Thus, partial ileal bypass, when employed for the correction of hyperlipidemia, appears to be an effective method of lipid reduction and may result in a decrease of mortality in certain subgroups of patients.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 26 - Colon, Rectum, and Anus

Ira J. Kodner
Robert D. Fry
James W. Fleshman
Elisa H. Birnbaum
Thomas E. Read

ANATOMIC CONSIDERATIONS

Embryology

The primitive gut, which is divided into the foregut, the midgut, and the hindgut, develops during the fourth week of gestation. The midgut develops into the small intestine (beginning at the entrance of the common bile duct) and the large intestine proximal to the midtransverse colon. This intestinal segment receives blood from the *superior mesenteric artery*. The hindgut develops into the large bowel distal to the midtransverse colon as well as the proximal anus and the lower urogenital tract and receives its main blood supply from the *inferior mesenteric artery*. The distal anal canal receives blood from branches of the internal pudendal artery and is ectodermal in origin (Fig. 26-1).

The developing midgut migrates out of the abdominal cavity during the sixth week of pregnancy. During the ensuing 4 weeks, the midgut rotates 270° in a counterclockwise direction about the superior mesenteric artery before assuming its final anatomic position in the abdominal cavity.

The developing hindgut is terminated by the *cloaca*, which during the sixth week of development is partitioned by the anorectal septum into the ventral urogenital sinus and the dorsal rectum. The anal canal is completely formed by the end of the eighth week when the thin anal membrane ruptures. The *dentate line* in the lower anal canal marks the transition from endodermal hindgut to ectodermal tissue.

Anatomy

Colon

The colon, approximately 3 to 5 feet in length, extends from the ileum to the rectum. The terminal ileum joins the cecum on its posteromedial border at the ileocecal valve. The fold of Treves (the only antimesenteric fatty appendage on the small bowel) is located on the distal ileum just proximal to the ileocecal valve. The *cecum* projects from the antimesenteric side of the ascending colon and is a large blind pouch with no mesentery. The cecum is approximately 7.5 to 8.5 cm in diameter and is the widest portion of the colon. The colon progressively diminishes in size to the *sigmoid* colon, its narrowest portion, which is approximately 2.5 cm in diameter. This size discrepancy accounts for the frequent observation that cecal tumors can grow to be large and bulky before the onset of symptoms, whereas sigmoid tumors are symptomatic at smaller sizes. The cecum, because of its relatively large diameter, is also the most common site of colonic rupture caused by distal obstruction. This is explained by Laplace's law: Tension is directly proportional to the radius of a sphere ($T = PR$, where T is tension in the wall of the bowel, P is internal pressure, and R is the radius of the bowel).

The layers of the colon wall include mucosa, submucosa, inner circular muscle, outer longitudinal muscle, and serosa (Fig. 26-2). The longitudinal muscle is separated into three distinct bands called *teniae coli* and positioned 120° apart about the circumference of the colon. The teniae converge proximally at the appendix and disappear as distinct bands at the proximal rectum at the level of the sacral promontory. *Haustra coli* are sacculations between the teniae and are separated by crescent-shaped folds called *plicae semilunares*. *Appendices epiploicae* are fatty appendages attached to teniae.

The *omentum* is attached to the transverse colon on its anterosuperior edge. The ascending colon, descending colon, and posterior surface of the hepatic and splenic flexures are usually retroperitoneal, whereas the cecum, transverse colon, and sigmoid colon are intraperitoneal in location. Although volvulus is most common in the sigmoid colon, the cecum and, rarely, the transverse colon also might twist about their mesenteries because of their intraperitoneal location and relative lack of fixation.

Arterial Supply

The *superior mesenteric artery* arises from the ventral surface of the aorta just below the celiac axis (see Fig. 26-1). It passes downward behind the pancreas and crosses in front of the third portion of the duodenum. It supplies the cecum, ascending colon, and transverse colon via its ileocolic and middle colic branches. Cadaver studies have shown that the right colic artery arises from the superior mesenteric artery in fewer than 15 percent of patients. It arises from the ileocolic in 85 percent of patients. *The inferior mesenteric artery* arises from the infrarenal aorta and supplies the descending colon, sigmoid colon, and upper rectum via its left colic, sigmoidal, and superior rectal branches.

Collaterals exist between the superior and inferior mesenteric arteries in the region of the splenic flexure. The arcades of the ileocolic, right, middle, and left colic arteries are peripherally connected by a series of anastomosing vessels commonly referred to as the *marginal artery of Drummond*. The marginal artery of Drummond runs along the mesenteric border of the colon and provides the vasa recta to the colon. The *arc of Riolan* is located closer to the mesenteric root. It is an inconstant vessel that exists between the left colic branch of the inferior mesenteric artery and the middle colic branch of the superior mesenteric artery. It is frequently referred to as the *meandering mesenteric artery* because of its tortuous course, and it is best visualized when there is an occlusion of either the inferior or superior mesenteric artery.

Venous Drainage

Except for the inferior mesenteric vein, which lies adjacent to the ascending branch of the left colic artery, the veins draining the colon follow the same course as the corresponding arteries. The *inferior mesenteric vein* drains the descending colon, sigmoid colon, and proximal rectum. It runs in a retroperitoneal location to the left of the ligament of Treitz, continues behind the body of the pancreas, and enters the splenic vein. The *superior mesenteric vein* drains the cecum, ascending colon, and transverse colon and joins the splenic vein to form the portal vein.

Lymphatic Drainage

The colon is encircled by lymphatic channels located in the submucosa and the muscularis mucosae. The mucosa has rich vascular plexi but no lymphatics. For this reason, superficial cancers that do not penetrate the muscularis mucosae cannot metastasize via the lymphatic route. This segmental architecture limits longitudinal intramural extension of tumors. Circumferential extension into submucosal and serosal zones results in annular lesions. Lymphatic vessels follow the arterial supply of the colon. Lymph nodes are located on the bowel wall (epicolic), along the inner margin of the bowel (paracolic), around the named mesenteric arteries (intermediate), and about the origin of the superior and inferior arteries (main).

Nerve Supply

Sympathetic nerves inhibit and *parasympathetic* nerves stimulate peristalsis. Sympathetic fibers to the right colon originate in the lower six thoracic segments of the spinal cord. They travel in the thoracic splanchnic nerves to the celiac plexus and then to the superior mesenteric plexus. The parasympathetic supply to the right side of the transverse colon is presumed to come from the right vagus nerve.

Sympathetic innervation of the left colon and rectum originates in the first three lumbar segments. These nerves join the preaortic plexus and become the inferior mesenteric plexus below the bifurcation of the aorta. The parasympathetic nerves to the left colon arise from the sacral nerves to form the nervi erigentes on either side of the rectum. Extensions of the sacral parasympathetics ascend through the hypogastric plexus to the area of the splenic flexure.

Rectum and Anal Canal General Anatomy

The rectal wall consists of mucosa, submucosa, and two complete muscular layers: inner circular and outer longitudinal (Fig. 26-3). The rectum is approximately 12 to 15 cm in length and extends from the sigmoid colon to the anal canal following the curve of the sacrum (Fig. 26-4). The anterior *peritoneal reflection* is about 5 to 7.5 cm above the anus in females and 7 to 9cm above the anus in males. The posterior peritoneal reflection is usually 12 to 15 cm above the anus. The upper third of the rectum is covered by peritoneum on its anterior and lateral surfaces. The middle third of the rectum is covered by peritoneum only on its anterior surface, and the lower third of the rectum is below the peritoneal reflection. The proximal rectum is identified as the level at which the teniae coli of the colon coalesce to form a complete layer of longitudinal muscle at approximately the level of the sacral promontory.

The rectum contains three distinct curves: The proximal and distal curves are convex to the right, whereas the middle curve is convex to the left. These folds project into the lumen as the *valves of Houston*. These mucosal infoldings present some difficulty for proctoscopic examination, but they are excellent targets for mucosal biopsy because they do not contain all layers of the muscular rectal wall and the risk of perforation is therefore diminished. The middle valve of Houston roughly correlates with the level of the anterior peritoneal reflection.

Waldeyer's fascia is a dense rectosacral fascia that begins at the level of the fourth sacral body and extends anteriorly to the rectum, covering the sacrum and overlying the vessels and nerves (see Fig. 26-86 E). Anterior to the extraperitoneal rectum is *Denonvilliers' fascia*, which is the rectovesical septum in men and the rectovaginal septum in women. The *lateral ligaments* of endopelvic fascia support the lower rectum but do not usually contain major blood vessels, as previously believed. Division of the lateral ligaments is thus possible without impairing the blood supply to the rectum or encountering a significant bleeding. Accessory middle hemorrhoidal arteries can be located in the lateral ligaments but are not critical to the blood supply of the rectum.

The *pelvic floor* is a musculotendinous sheet formed by the levator ani muscle and is innervated by the fourth sacral nerve (Fig. 26-5). The pubococcygeus, iliococcygeus, and puborectalis muscles make up the levator ani muscle. These are paired muscles that intertwine and act as a single unit. The line of decussation is called the *anococcygeal raphe*. The rectum, vagina, urethra, and the dorsal vein of the penis pass through the levator hiatus in the pubococcygeal portion of the levator ani. During defecation, the puborectalis relaxes and the levator ani contracts, widening the levator hiatus.

The *anal canal* starts at the pelvic diaphragm and ends at the anal verge (see Fig. 26-4). It is approximately 4 cm long and normally exists as a collapsed anteroposterior slit. The *anatomic anal canal* extends from the anal verge to the dentate line. For practical purposes, however, surgeons usually define the *surgical anal canal* as extending from the anal verge to the anorectal ring, which is the circular lower (see Fig. 26-3) border of the puborectalis that is palpable by digital rectal examination. The anorectal ring is 1 to 1.5 cm above the dentate line.

The *anal verge* is the junction between anoderm and perianal skin. The anoderm is a specialized epithelium rich in nerves but devoid of secondary skin appendages (hair follicles, sebaceous glands, or sweat glands). The *dentate line* is a true mucocutaneous junction located 1 to 1.5 cm above the anal verge (see Fig. [26-3](#)). A 6- to 12-mm transitional zone exists above the dentate line over which the squamous epithelium of the anoderm becomes cuboidal and then columnar epithelium.

The anal canal is surrounded by an internal and external sphincter, which together constitute the *anal sphincter mechanism* (see Fig. [26-5](#)). The *internal sphincter* is a specialized continuation of the inner circular smooth muscle of the rectum. It is an involuntary muscle and is normally contracted at rest. The intersphincteric plane represents the fibrous continuation of the longitudinal smooth muscle layer of the rectum.

The *external sphincter* is a voluntary, striated muscle divided into three U- shaped loops (subcutaneous, superficial, and deep) acting as a single functional unit. It is a specialized continuation of the levator muscles of the pelvic floor, specifically of the puborectalis muscle. The *puborectalis* originates at the pubis and joins posterior to the rectum. It is normally contracted, causing an 80° angulation of the anorectal junction.

The *columns of Morgagni* consist of 8 to 14 longitudinal mucosal folds located just above the dentate line and forming the anal crypts at their distal end (see Fig. [26-3](#)). Small rudimentary glands open into some of these crypts. The ducts of these glands penetrate the internal sphincter, and the body of the gland resides in the intersphincteric plane.

Arterial Supply

The terminal branch of the inferior mesenteric artery becomes the *superior rectal artery* as it crosses the left common iliac artery (Fig. [26-6A](#)). It descends in the sigmoid mesocolon and bifurcates at the level of the third sacral body. The left and right branches of the superior rectal artery supply the upper and middle rectum.

The middle and inferior rectal arteries supply the lower third of the rectum. The *middle rectal arteries* arise from the internal iliac arteries, run through Denonvilliers' fascia, and enter the anterolateral aspect of the rectal wall at the level of the anorectal ring. Collaterals exist between the middle and superior rectal arteries. Preservation of the middle rectal arteries is necessary to maintain viability of the remaining rectum after proximal ligation of the inferior mesenteric artery.

The *inferior rectal arteries* are branches of the internal pudendal arteries. They traverse Alcock's canal and enter the posterolateral aspect of the ischiorectal fossa. They supply the internal and external sphincters and the lining of the anal canal and do not form collaterals with the other rectal arteries. The *middle sacral artery* arises just proximal to the aortic bifurcation and provides very little blood supply to the rectum.

Venous Drainage

The venous drainage of the rectum parallels the arterial supply and empties into both the portal and the systemic (caval) systems. The upper and middle rectum are drained by the *superior rectal vein*, which enters the portal system via the inferior mesenteric vein (see Fig. [26-6 B](#)). The lower rectum and upper anal canal are drained by the *middle rectal veins*, which empty into the internal iliac veins and then into the caval system. The *inferior rectal veins* drain the lower anal canal and empty into the pudendal veins, which drain into the caval system via the internal iliac veins. Low rectal tumors can thus metastasize through venous channels into both the portal and systemic venous systems.

There are three submucosal *internal hemorrhoidal complexes* located above the dentate line (see Fig. [26-3](#)). The left lateral, right posterolateral, and right anterolateral internal hemorrhoidal veins drain into the superior rectal vein. Below the dentate line the *external hemorrhoidal veins* drain into the pudendal veins. There is communication between the internal and external plexi.

Lymphatic Drainage

The rectal lymphatic flow is segmental and circumferential and follows the same distribution as the arterial blood supply (Fig. [26-7](#)). Lymph from the upper and middle rectum drains into the inferior mesenteric nodes. The lower rectum is drained primarily by lymphatics that follow the superior rectal artery and enter the inferior mesenteric nodes. Lymph from the lower rectum also can flow laterally along the middle and inferior rectal arteries, posteriorly along the middle sacral artery, or anteriorly through channels in the rectovesical or rectovaginal septum. These channels drain to the iliac nodes and subsequently to periaortic lymph nodes.

Lymphatics from the anal canal above the dentate line drain via the superior rectal lymphatics to the inferior mesenteric lymph nodes or laterally to the internal iliac lymph nodes. Below the dentate line, the lymphatics drain primarily to the inguinal nodes but can drain to the inferior or superior rectal lymph nodes as well.

Nerve Supply

The innervation of the rectum is shared with the urogenital organs of the pelvis and consists of both sympathetic and parasympathetic nerves (Fig. [26-8](#)). *Sympathetic* nerves from thoracolumbar segments unite below the inferior mesenteric artery to form the *inferior mesenteric plexus*. These purely sympathetic nerves descend to the *superior hypogastric plexus* located below the aortic bifurcation. They then bifurcate and descend in the pelvis as the *hypogastric nerves*. The lower rectum, bladder, and sexual organs in both men and women receive sympathetic innervation via the hypogastric nerve. Injury to the inferior mesenteric plexus can result during ligation of the inferior mesenteric artery at its origin.

Parasympathetic fibers from the second, third, and fourth sacral roots (the *nervi erigentes*) unite with the hypogastric nerves anterior and lateral to the rectum forming the *pelvic plexus*, which runs laterally in the pelvis. The periprostatic plexus arises from the pelvic plexus. Mixed fibers from these plexi innervate the rectum, internal anal sphincter, prostate, bladder, and penis. The *pudendal nerve* (S2, S3, S4) mediates sensory stimuli from the penis and clitoris via the dorsal nerve.

Both sympathetic and parasympathetic fibers are essential for penile erection. The parasympathetic fibers cause vasodilation and increased blood flow in the corpus cavernosum, resulting in an erection. The sympathetic fibers cause vasoconstriction of the penile veins and thus sustain the erection. Sympathetic nerves cause contraction of the ejaculatory ducts, seminal vesicles, and prostate and are necessary for ejaculation. Damage to the periprostatic plexus might occur during surgical dissection of the rectum. Injury to the pelvic autonomic nerves may result in bladder dysfunction, impotence, or both.

The *internal anal sphincter* is innervated by both sympathetic and parasympathetic nerves, and both are inhibitory to the sphincter. The internal sphincter has a continuous tone that decreases as rectal pressure increases. Once the rectum empties, the internal sphincter tone rises again. The *external anal sphincter* and *levator ani muscles* are innervated by the inferior rectal branch of the internal pudendal nerve (S2, S3, S4) and the perineal branch of the fourth sacral nerve. Any distention of the rectum results in relaxation of the internal sphincter. The external sphincter can be contracted voluntarily and kept in that state for approximately 1 minute.

Below the dentate line, cutaneous sensations of heat, cold, pain, and touch are conveyed by afferent fibers of the inferior rectal and perineal branches of the pudendal nerve. Above the dentate line, a poorly defined dull sensation, experienced when the mucosa is pinched or when internal hemorrhoids are ligated, is probably mediated by parasympathetic fibers.

Resection of the sacrum with sacrifice of *sacral nerves* occasionally may be required for total resection of pelvic tumors. Sacrifice of the lower sacral nerves will lead to saddle anesthesia and possible motor weakness in the lower extremities. Preservation of at least one of the third sacral nerves is required to maintain acceptable anal continence. Near-normal continence will be maintained if the upper three sacral roots on one side are preserved along with the upper two sacral roots on the contralateral side. If all the sacral roots are destroyed unilaterally but the contralateral nerves are preserved, the patient should maintain continence. If both S3 roots are destroyed, the patient will be incontinent. The upper half of S1 is needed for stability of the spine and pelvis.

DISTURBANCES OF PHYSIOLOGY

Normal Colonic Function

The colon is much more than a receptacle and conduit for the end products of digestion. This organ absorbs water, sodium, and chloride and secretes potassium, bicarbonate, and mucus; it is the site of digestion of certain carbohydrates and proteins and provides the environment for the bacterial production of vitamin K.

Water and Electrolyte Exchanges

The major absorptive function of the colon is the final regulation of water and electrolyte balance in the intestine. The colon reduces the volume of enteric contents by absorbing greater than 90 percent of the water and electrolytes presented to it. This accounts for 1 to 2 liters of fluid and 200 meq of sodium and chloride per day on average. During a 24-h period, 8 liters of fluid enters the jejunum. In healthy individuals, the small bowel absorbs about 6.5 liters and the colon 1.4 liters, leaving 0.1 liter of normal fecal water content. Under maximum conditions, the colon can absorb 5 to 6 liters of fluid a day. Only if small bowel absorption is reduced to less than 2 liters a day is colonic salvage overwhelmed, and the resulting increase in fecal water content is manifested as diarrhea.

Sodium and Potassium

In healthy persons, the colon absorbs water, sodium, and chloride while secreting potassium and bicarbonate. The colon is able to absorb sodium against very high concentration gradients, especially in the distal colon, which shares many basic cellular mechanisms of sodium and water transport with the distal convoluted tubule of the kidney (Fig. [26-9](#)). The colon has the capacity to absorb up to 400 meq of sodium per day, and the colonic response to aldosterone stimulation may be an important compensatory mechanism during dehydration. A patient with an ileostomy loses this absorptive capacity and cannot tolerate increased sodium losses or decreased sodium intake.

Potassium transport in the colon is mainly passive, along an electrochemical gradient generated by the active transport of sodium. Potassium continues to be secreted into the lumen as long as the luminal concentration is less than 15 meq/L; however, at concentrations below this level, potassium secretion ceases and potassium absorption occurs.

Chloride and Bicarbonate

Chloride, like sodium, is actively absorbed across the colonic mucosa against a concentration gradient. Chloride and bicarbonate are exchanged at the luminal border. Chloride absorption is facilitated by an acidotic environment, and the secretion of bicarbonate is enhanced by increased concentration of luminal chloride. This becomes clinically important in patients with ureterosigmoidostomy, who may develop hyperchloremia and metabolic acidosis due to the absorption of urinary chloride accompanied by the secretion of bicarbonate into the colon.

Short-Chain Fatty Acids

Although active absorption of nutrients is minimal, the colon can passively absorb short-chain fatty acids (SCFAs) formed by intraluminal bacterial fermentation of unabsorbed

carbohydrates, particularly fiber. This can account for up to 540 kcal/day of assimilated calories. The absorbed SCFAs butyrate, acetate, and propionate are the major fuel sources of the colonic epithelium. They provide the energy required for active sodium transport, and altered SCFA metabolism or SCFA deficiency may result in impaired colonic sodium absorption.

There is evidence that SCFA metabolism is impaired in patients with ulcerative colitis and that intraluminal infusion of SCFAs can be of benefit in patients with colitis. SCFAs also have been shown to be effective in treating diversion colitis (which can occur after a diverting ileostomy or colostomy), implicating colonocyte nutritional deficiency as the cause of this disorder. Elemental diets, which lack fiber, are associated with colonic mucosal atrophy.

Ammonia

Colonic bacteria degrade protein and urea to produce ammonia. Ammonium ions react with bicarbonate to form nonionized ammonia and carbon dioxide. The nonionized ammonia diffuses across the colonic mucosa and is carried to the liver through the portal vein. This absorptive process depends on bacterial degradation of protein and intraluminal pH. As the pH in the colonic lumen falls, ammonia absorption decreases.

Ammonia absorption becomes a problem in patients with severe liver disease, who are unable to metabolize ammonium. The administration of poorly absorbed broad-spectrum antibiotics, coupled with cathartics and enemas, diminishes the colonic bacterial population and decreases ammonia production. Lactulose can be administered with the rationale that its metabolism to organic acids lowers the colonic luminal pH and decreases the absorption of ammonia, and it also reduces bacterial numbers by its purgative action because it causes an osmotic diarrhea.

Diarrhea

A number of agents can stimulate fluid and electrolyte secretion in the colon, including bacteria, enterotoxins, hormones, neurotransmitters, and laxatives. The diarrhea associated with *Shigella* and *Salmonella* infection is caused by diminished absorption or increased secretion of water, sodium, and chloride. Gut hormones, particularly vasoactive intestinal polypeptide, have been shown to have significant effects on colonic absorption and secretion. Prostaglandins play a role in the pathogenesis of diarrhea associated with ulcerative colitis and several laxatives. Any sort of irritation to the colon also can cause increased secretion, such as bile salt malabsorption after terminal ileal resection and long-chain fatty acid malabsorption in steatorrhea. The induced colonic mucus and fluid are high in potassium and may result in potassium depletion in chronic cases.

Motility

Colonic motility is more difficult to study than small bowel motility because of the great regional heterogeneity in the colon and because of the intermittent and unpredictable nature of colonic contractile waves. Even the normal pattern of colonic motility remains the subject of debate. In the first decade of the twentieth century, radiologists observed three patterns of colonic contractions: retrograde movements, segmental contractions, and mass movements. More recently, electrophysiologic and manometric measurements appear to correlate with observed colonic motor contractions; however, much more research is required before the complexities of colonic motility are satisfactorily understood.

Retrograde movements were identified initially as contractile waves originating in the transverse colon and traveling toward the cecum. These contractions appear to delay the transit of material from the right colon, prolonging the exposure of luminal contents to the mucosa and thus increasing the absorption of fluid and electrolytes.

Segmental contractions are the most commonly observed motility pattern. These are localized simultaneous contractions of longitudinal and circular colonic musculature that isolate short segments of the colon. These contractions occur more frequently on the right side of the colon, but their occurrence in the sigmoid colon has been hypothesized to play a role in the formation of diverticula. This contractile activity is decreased by anticholinergic drugs and increased by food and cholinergic drugs.

The third and most infrequent type of colonic activity is the *mass movement*. This pattern is characterized by an antegrade propulsive contractile wave involving a long segment of colon. The contraction propels the colonic contents forward at the rate of 0.5 to 1.0 cm/s. These contractions are 20 to 30 s in duration and generate intracolonic pressures of 100 to 200 mmHg. They occur three or four times a day, often after arising in the morning or in the late postprandial period. Mass movements in the right colon move colonic contents to the more distal colon. A mass movement in the sigmoid colon may occur at the time of defecation.

Many factors affect colonic motility. Emotions such as hostility, anger, and resentment are associated with hypermotility, whereas anxiety and fear are associated with hypomotility. Exercise has been shown to increase segmental and peristaltic colonic activity; sleep is a colonic motility depressant. Mechanical colonic distention stimulates motility and is the basis for the effect of bulk laxatives. Nondigestible polysaccharides and cellulose derivatives absorb water and increase fecal mass, thus stimulating colonic propulsion. Fatty meals appear to have a greater stimulatory effect on colonic motility than carbohydrate or protein meals. Colonic motility is influenced by certain hormones. Glucagon and somatostatin have inhibitory effects. Cholecystokinin increases colonic motor activity and is thought to be responsible for the gastrocolic reflex (increased colonic motility during and after meals).

Neurogenic Control

Colonic motility is regulated by complex and poorly understood extrinsic and intrinsic neuronal systems. The extrinsic system consists of preganglionic parasympathetic neurons from the vagus nerve or the pelvic plexus (S2, S3, S4) and the postganglionic sympathetic neurons (T11–L2). The intrinsic (enteric) nervous system is composed of neurons residing within the colon wall. These intramural neurons communicate through extensive interconnecting plexuses. The enteric plexuses are identified by their location within the colon wall: subserosal, myenteric (Auerbach's plexus), submucosal (Meissner's plexus), and mucosal.

The enteric neurons also may be classified by function: sensory, associative, or motor. The complex interaction between these neurons can be demonstrated by a localized radial distention of the colon that activates sensory neurons. These neurons communicate through associative neurons to proximal and distal motor neurons, resulting in contraction of the colon proximal to the stimulus and simultaneous distal relaxation.

Motor neurons of the enteric nervous system may be excitatory or inhibitory. Excitatory neurons are mainly postganglionic parasympathetic cells that are located in the myenteric or submucosal plexus. Acetylcholine is the most important excitatory neurotransmitter, and its action can be blocked by atropine or scopolamine. Noncholinergic agents, such as substance P, may be released by enteric neurons and appear to have an excitatory action.

Inhibitory motor neurons of the enteric nervous system reside mainly in the myenteric plexus. These cells are noncholinergic and nonadrenergic. There is evidence that purine nucleotides such as adenosine triphosphate may be the neurotransmitters of these cells. Vasoactive inhibitory peptide (VIP) is also a very important inhibitory substance. The inhibitory neurons are involved in the coordinated propagation of colonic propulsive movements such as the mass movement.

Colonic Microflora

The human colon is sterile at birth, but within a matter of hours the intestine is colonized from the environment in an oral to anal direction. *Bacteroides*, destined to be the dominant bacteria in the colon, is first noted at about 10 days after birth. By 3 to 4 weeks after birth, the characteristic stool flora is established and persists into adult life. The large intestine harbors a dense microbial population, with bacteria accounting for approximately one-third of the dry weight of feces. Each gram of feces contains 10^{11} to 10^{12} bacteria, with anaerobic bacteria outnumbering aerobic organisms by a factor of 10^2 to 10^4 . *Bacteroides* species are the most common colonic organisms, present in concentrations of 10^{11} to 10^{12} organisms per milliliter of feces. *Escherichia coli*, by comparison, is present in concentrations of 10^8 to 10^{10} organisms per milliliter.

The complex symbiotic relationship between human beings and colonic bacteria is poorly understood, but it is recognized that endogenous colonic bacteria suppress the emergence of pathogenic microorganisms, play an important role in the breakdown of carbohydrates and proteins that escape digestion in the small bowel, participate in the metabolism of numerous substances that are salvaged by the enterohepatic circulation (including bilirubin, bile acids, estrogen, and cholesterol), and produce certain beneficial elements such as vitamin K.

Colonic Gas

Nitrogen, oxygen, carbon dioxide, hydrogen, and methane make up 99 percent of all the gas in the gut. Nitrogen and oxygen are found in the atmosphere and appear in the colon via swallowing air. Hydrogen, methane, and carbon dioxide are produced by bacterial fermentation of carbohydrates and proteins in the colon. An eminent flatologist, Levitt, has shown that most patients who complain of excessive flatus have high concentrations of hydrogen and carbon dioxide in their intestinal gas. Because carbon dioxide is an end product of bacterial fermentation, therapy consists of diet manipulation to decrease the amount of ingested carbohydrate, especially lactose, wheat, and potatoes. In addition, various vegetables (particularly beans) contain undigestible carbohydrates that serve as a substrate for colonic bacterial fermentation and should be avoided by these patients.

Normal humans pass approximately 600 mL of flatus daily (ranging from 200 to 2000 mL/day). There is normally about 200 mL of gas in the intestinal tract, and the amount of detectable gas is seldom increased in patients who complain of bloating and distention. Most complaints of increased intestinal gas probably represent motility disorders instead of actual excessive production of gas.

The five predominant colonic gases (N_2 , O_2 , CO_2 , H_2 , CH_4) are odorless. Trace substances such as dimethyl sulfide and methanethiol appear to be responsible for the odor of human feces. Hydrogen and methane are combustible gases and can explode if present in the colon when electrocautery is used for polypectomy or biopsy. Adequate bowel cleansing is necessary before the use of intracolonic electrocautery. Mannitol is an inappropriate purgative under such circumstances because it is a substrate for colonic bacterial fermentation. Polyethylene glycol purgatives are not metabolized by intestinal bacteria and have not been associated with intracolonic explosions.

Disorders of Colonic Physiology

Irritable Bowel Syndrome

Irritable bowel syndrome is a disorder manifested by altered bowel habits and abdominal pain in the absence of other pathology. Patients with the diagnosis of irritable bowel syndrome have been shown to have increased slow-wave activity (3 cycles per minute) in the rectosigmoid, corresponding to increased contractile activity at the same frequency. A similar motility pattern has been noted in patients with diverticular disease, and some authors have suggested that uncoordinated smooth muscle activity followed by increased intraluminal pressure may be the underlying mechanism producing the irritable bowel syndrome and colonic diverticuli. Irritable bowel syndrome is a diagnosis of exclusion and has many different presentations. It is important to exclude organic colonic disease by the appropriate diagnostic tests. Irritable bowel syndrome is not amenable to surgical treatment.

Delayed Transit

Constipation is a symptom with numerous and diverse causes. It is difficult to formulate a precise definition of constipation, but one that is generally accepted is fewer than three stools per week while following a high-residue diet (30 g dietary fiber daily). The causes of constipation are multiple and include metabolic and endocrine disorders (hypothyroidism and hypercalcemia), faulty diet (principally inadequate intake of fiber and fluid), neurologic abnormalities (Parkinson's disease, multiple sclerosis), Hirschsprung's disease,

medications (codeine, antidepressants, anticholinergics), colonic inertia, and lack of exercise. Abnormalities of pelvic floor function, such as nonrelaxation of the puborectalis muscle and intussusception of the rectum, have been implicated as important causes of constipation.

A change in bowel habits is a common presenting symptom for colorectal tumors; thus the evaluation of a patient with new-onset constipation should include colonoscopy or contrast enema plus proctoscopy. If no obstructing lesion is present, a trial of fiber supplementation and increased fluid intake relieves constipation symptoms in the majority of patients. If unsuccessful, colonic transit time can be measured. A standardized test has not been generally accepted, but the techniques used by most institutions rely on observation of the movement of ingested radiopaque markers through the intestinal tract. The patient ingests a high-fiber diet (30 g/day) and abstains from all laxatives and enemas for 48 h before and during the investigation. The markers, which are commercially available, are ingested before breakfast on the first day and then followed by radiographs, usually at days 3 and 5. If all markers are still present on the radiograph by day 3, intestinal transit time is slower than normal. Normally, at least 80 percent of markers should have passed by day 5 after ingestion, and all markers should pass by day 7. The abdominal location of the markers indicates the segment of the colon involved with the motility disorder. If a significant number of markers remain in the proximal colon for several days, total colonic inertia should be suspected; if the markers accumulate in the rectosigmoid, anorectal outlet obstruction is the likely cause of the delayed transit.

Anorectal outlet obstruction is evaluated with a variety of techniques. Defecography can demonstrate rectal intussusception and nonrelaxation of the puborectalis muscle. Cinefluoroscopy is used to obtain video images of the rectum from the lateral view during attempts to evacuate barium paste. Rectocele (anterior displacement of the posterior wall of the vagina) and enterocele (inferior displacement of the small bowel on the vagina) also can be demonstrated. Anal manometry can demonstrate absence of the normal rectoanal inhibitory reflex (relaxation of the internal anal sphincter and brief contraction of the external anal sphincter with distention of the rectum), which is suggestive of Hirschsprung's disease, and may be used as an adjunct to rectal biopsy to make this diagnosis. Inability to expel a saline- filled balloon from the rectum suggests failure of the puborectalis muscle to relax during defecation.

Slow-transit constipation (colonic inertia) is a poorly understood disorder of chronic intestinal stasis that afflicts predominantly women. This disorder is sometimes referred to as *Arbuthnot Lane's disease*, after the British surgeon who, at the beginning of the century, recommended colectomy with ileorectal anastomosis as the treatment for chronic constipation. An extreme form of colonic inertia presents as acquired megacolon, which has been seen in association with neurologic disorders and in institutionalized and psychiatric patients. These patients present with frequent fecal impaction, volvulus, and episodes of abdominal distention mimicking large bowel obstruction. The colon can attain immense proportions, filled with gas or stool. Sigmoidoscopy is often performed in these patients in an effort to reduce what appears to be a sigmoid volvulus. The instrument slips easily into the rectum, decompressing it somewhat, but significant decompression is not achieved. A water-soluble contrast enema is often diagnostic and therapeutic. Patients with chronic megacolon do not have evidence of obstruction, and the impaction may be relieved after the study. Complete evaluation includes colonic transit studies to rule out colonic inertia and defecography and anal manometry to rule out pelvic floor abnormalities.

The medical treatment of colonic inertia consists of eliminating obvious sources of constipation, encouraging proper exercise, and ensuring appropriate intake of liquids and dietary fiber. A diet containing 30 g of dietary fiber daily should be prescribed (the average American adult consumes 13 to 19g of fiber per day). Laxatives and enemas should be avoided if possible. While laxatives are useful for occasional relief of constipation, their chronic use is counterproductive and leads to increased motility abnormalities. Medications known to cause constipation should be reduced or eliminated whenever possible.

The exact role of surgery in the treatment of colonic inertia is not clearly defined. If transit studies demonstrate total colonic inertia and appropriate medical treatment is not successful, abdominal colectomy with ileorectal anastomosis has been beneficial. Such patients are relatively rare and must be carefully selected. If the motility disorder is caused by anorectal outlet obstruction, abdominal colectomy is inappropriate; treatment should be directed toward correction of the primary problem, as described in a later section of this chapter. In rare cases where the individual is incapacitated by chronic megacolon, the only relief of symptoms may be total colectomy with construction of an ileostomy.

Colonic Pseudo-Obstruction (Ogilvie's Syndrome)

Intestinal pseudo-obstruction, a profound ileus without evidence of mechanical obstruction, was first described by Ingelfinger in 1943. The first description of the colonic variant of pseudo-obstruction is thought to be Sir Heneage Ogilvie's 1948 report of two cases associated with malignant infiltration of the celiac plexus. Colonic pseudo-obstruction is associated with neuroleptic medications, opiates, malignancy, severe metabolic illness, or any acute severe illness. Excess sympathetic tone has been postulated as the common mechanism. This concept is supported by anecdotal reports of success with epidural anesthesia, which paralyzes the sympathetic afferent and efferent nerve fibers to the colon and with neostigmine, which increases parasympathetic tone by its anticholinesterase effect.

The diagnosis of colonic pseudo-obstruction is one of exclusion; a water- soluble contrast enema or colonoscopy should be performed to exclude mechanical obstruction. Initial treatment maneuvers include nasogastric decompression, rectal tube placement, correction of metabolic disorders, and discontinuation of medications that decrease colonic motility (e.g., narcotics and anticholinergics). Evidence of rapid cecal dilatation or a cecal diameter greater than 11 to 12 cm on abdominal x-ray should prompt more aggressive intervention because of the risk of cecal ischemia, necrosis, and perforation. Colonoscopic decompression is successful initially in 70 to 90 percent of patients, with a recurrence rate of 10 to 30 percent. The majority of patients will respond to a second colonoscopic decompression. The use of long colonic tubes has not been uniformly successful. Operation is reserved for patients with persistent colonic dilatation despite colonoscopic decompression and patients with peritonitis. Cecostomy or loop colostomy are effective in relieving distention; laparotomy should be performed in patients with peritonitis, and nonviable bowel should be resected.

Volvulus

Volvulus is a twisting of an air-filled segment of bowel about its narrow mesentery, causing an obstruction that can lead to strangulation and gangrene (Fig. [26-10](#)). Volvulus almost never occurs when the colon is filled with solid stool. The condition is classified and treated according to its location in the colon. Sigmoid, cecal, and transverse colon are the most common sites. Volvulus causes less than 10 percent of intestinal obstructions in the United States, but it has been reported as the cause of 55 percent of intestinal obstructions in parts of Iran and Russia.

Sigmoid Volvulus

Sigmoid volvulus accounts for about 90 percent of all volvulus seen in the United States. It usually occurs in older or institutionalized patients and in patients with a variety of neurologic disorders. A redundant sigmoid colon with a narrow-based mesocolon is the factor necessary for sigmoid volvulus.

Diagnosis

Patients present with complaints of abdominal pain, cramping, distention, and obstipation. Many patients give a history of previous episodes of abdominal pain and distention. Elevated white blood cell count, fever, or evidence of peritoneal irritation on physical examination suggests the presence of gangrenous bowel.

Plain abdominal radiographs may be diagnostic (Fig. [26-11](#)). The inflated sigmoid appears as an inverted U-shaped, sausage-like loop. A dense line is often seen running downward to the point of torsion. Water-soluble contrast enema may be diagnostic but should not be performed on patients with suspected gangrene. The column of barium stops at the obstruction and tapers to a point ("bird's beak" deformity) (Fig. [26-12](#)). Reduction of the volvulus may occur during the examination.

Treatment

If gangrene is suspected, or if the patient has signs of peritoneal irritation, reduction should not be attempted, and the patient should undergo emergency operation. If peritonitis is not present, rigid sigmoidoscopy should be performed. An attempt to reduce the volvulus by inserting a soft rectal tube through the sigmoidoscope past the point of volvulus usually results in dramatic decompression. The tube is left in place to further decompression. Flexible sigmoidoscopy or colonoscopy has been used to reduce volvulus beyond the limits of the rigid sigmoidoscope. Unsuccessful detorsion, bloody discharge, or evidence of mucosal ischemia indicates strangulation or gangrene. If this is found, sigmoidoscopy should be terminated, and the patient should undergo an emergency operation.

Recurrence rates following nonoperative reduction can be as high as 40 percent. If the volvulus is reduced successfully, the patient should be stabilized and undergo mechanical bowel preparation, followed by elective sigmoid resection. Resection with primary anastomosis is the recommended surgical treatment for sigmoid volvulus in a patient who is adequately prepared. Resection of the sigmoid colon with construction of a colostomy and Hartmann pouch is usually necessary if emergent operation is performed on unprepared bowel. All nonviable bowel must be resected, and mortality is related to the presence of gangrenous bowel.

Cecal Volvulus

Cecal volvulus is generally seen in younger patients and accounts for less than 20 percent of colonic volvulus. It is thought to be due to anomalous fixation of the right colon to the retroperitoneum, leading to a freely mobile cecum.

Diagnosis

Abdominal pain is the most common symptom of cecal volvulus. Nausea and vomiting, obstipation, and diarrhea are also common symptoms. Clinically, the patient appears to have a small bowel obstruction, and many patients give a history of chronic intermittent symptoms.

Ninety percent of patients with cecal volvulus have a full axial volvulus; 10 percent have a cecal bascule (cecum folded on itself in an anterior cephalad direction). A plain abdominal radiograph may be diagnostic. Cecal volvulus is seen as a kidney-shaped, air-filled structure in the left upper quadrant. The convexity of the loop faces the left upper quadrant. Water-soluble contrast enema may show obstruction of the column of contrast material at the level of the volvulus. The tapered edge of the contrast material points toward the site of the torsion.

Treatment

Resection, with ileostomy and mucous fistula, is indicated if there is gangrenous bowel. The operative treatment of nongangrenous cecal volvulus is controversial. Right hemicolectomy, cecopexy, tube cecostomy, and simple detorsion have been recommended. Resection has the lowest recurrence rates (essentially zero), and primary anastomosis often can be performed safely. Suture of the right colon to the lateral peritoneal surface (cecopexy) risks the sutures pulling through the thin-walled, distended cecum and has had variable recurrence rates. Cecostomy has been associated with abdominal wall and wound complications and persistent fecal fistula in up to 50 percent of patients and recurrence of volvulus in up to 15 percent of patients. Logic dictates that detorsion alone without fixation should lead to high recurrence rates; however, only 7 to 15 percent of patients have been reported to have recurrent volvulus. Regardless, few authors have recommended simple detorsion as adequate treatment for cecal volvulus.

Transverse Colon Volvulus

Volvulus of the transverse colon is rare because the transverse colon is normally prevented from rotation by a broad-based, short mesentery. Failure of the mesenteries to fix

normally or narrowing of the mesenteric attachments, allowing a short, fixed point of attachment, may predispose the transverse colon to volvulus. Patients present with a clinical picture similar to sigmoid volvulus. Barium enema is diagnostic and reveals the point of obstruction. Successful attempts at colonoscopic detorsion have been reported. Most patients require operative detorsion with resection of the redundant transverse colon.

Diverticular Disease General Considerations

The word *diverticulum* is used to describe an abnormal sac or pouch protruding from the wall of a hollow organ such as the colon. The term *true diverticulum* implies that the pouch is composed of all layers of the intestinal wall (true diverticula are rare), whereas a *false diverticulum* lacks a portion of the normal bowel wall. Specifically, the commonly observed diverticula of the colon consist of protrusions of mucosa through the muscular layers of the wall and are devoid of a muscular coat and thus are false diverticula.

Diverticular disease and *diverticulosis* are terms used to indicate the presence of colonic diverticula (Fig. 26-13), a condition that is rare before age 30, becomes more common with increasing age, and is present in approximately 75 percent of Americans over age 80. The variation in incidence by sex varies among reports, and a review of large series of patients fails to indicate that either men or women are more susceptible to the development of diverticula.

Diverticular disease seems to be a phenomenon of the industrial revolution and Western society. The marked increase in the prevalence of diverticulosis in England after 1910 has been attributed to the process introduced approximately 30 years earlier of roller-milling wheat flour. This advance resulted in a decreased consumption of unprocessed cereals and was accompanied by an increased consumption of sugar and meat. In the past half century, the amount of fiber consumed by individuals in North America and western Europe has declined, while the prevalence of diverticular disease has increased dramatically. Further evidence of reduced intake of fiber as a cause of diverticulosis is provided by the observation that the disease is almost unknown in sub-Saharan African blacks who consume a high level of dietary fiber; however, blacks in Johannesburg who consume a low-fiber diet have the same high incidence of diverticular disease as South African whites.

Etiology

The precise cause of diverticulosis is not known. Anatomic studies have demonstrated that diverticula are actually herniations of mucosa through the colon at sites where arterioles penetrate the muscular wall. These sites are located predominantly on the mesenteric side of the antimesenteric teniae. A lesser number of protrusions can occur between the antimesenteric teniae (Fig. 26-14). In some cases the artery divides, with one branch penetrating the wall at the site of a diverticulum and the other branch passing external to the muscular coat and sometimes actually being displaced over the dome of the diverticulum. This intimate association between the artery and the diverticulum is responsible for the massive hemorrhage that occasionally can complicate diverticulosis.

Another striking anatomic abnormality associated with diverticulosis is marked thickening of the affected colonic wall. This muscular abnormality seems to precede the appearance of diverticula and occurs predominantly in the sigmoid colon. Diverticula occur most often in the sigmoid, with the abnormalities being confined to the sigmoid in over half the patients with diverticular disease; other colonic segments (usually the descending colon) are also involved in about another 40 percent of affected patients, and the entire colon is involved in approximately 5 to 10 percent of patients.

A low-fiber diet appears to be associated with a narrow sigmoid colon. The small lumen permits segmental contractions to isolate small regions of the colon into compartments that are subjected to high pressures. Normally, the colonic pressure is only slightly above atmospheric pressure, but pressures as high as 90 mmHg can be generated in the narrow sigmoid by segmental contraction. It is presumed that repeated high-pressure contractions produce the herniation of mucosa through the colonic wall resulting in diverticula.

Diverticulitis denotes infection associated with diverticula. This term is somewhat a misnomer because the infectious process commonly associated with diverticular disease is pericolic in nature and predominantly involves the surrounding soft tissues such as mesentery or adjacent organs, in addition to the serosal surface of the colon.

Peridiverticulitis would be a more accurate descriptive term for this disease. It should be recognized that the infection is the result of a perforation of a diverticulum, leading to extravasation of feces from the lumen to the serosal surface of the colon. In most instances, the amount of fecal contamination is minute, and the body's natural defenses confine the infection. If the contamination is significant, or if the patient's response to infection is compromised, the local peritoneal defenses may be overwhelmed, and abscess formation or even generalized peritonitis can occur. Whether the perforation of the diverticulum is the result of increased colonic pressure, obstruction of the neck of the diverticulum by inspissated feces, or ischemia of the protruded mucosal sac, the end result, pericolic infection, is the same. Recognition that the cause of the abdominal infection is a perforation of a diverticulum provides a basis for understanding the signs and symptoms of the disease as well as the rationale for determining appropriate diagnostic tests and treatment.

Patients with *acute diverticulitis* usually complain of pain in the left lower abdomen. The pain may radiate to the suprapubic area, left groin, or back. An alteration in bowel habit, usually constipation or occasionally diarrhea, may be noted. Fever and chills are not unusual, and urgency or frequency of urination may be present if the inflammation is adjacent to the bladder. Rectal bleeding is distinctly unusual in the presence of diverticulitis.

The physical findings depend on the severity of the infection. The most common finding is tenderness over the left lower abdomen. If considerable inflammation is present, abdominal palpation may reveal a poorly circumscribed area of induration. A tender mass in the left lower quadrant is suggestive of a localized phlegmon or, more likely, an abscess. Occasionally, abdominal distention is caused by ileus or by partial obstruction of small bowel involved in the inflammatory process. Rectal examination may demonstrate pelvic tenderness, and a tender pelvic abscess may be palpated by rectal or vaginal examination.

A limited sigmoidoscopic examination occasionally is indicated to rule out a rectal cancer. If diverticulitis is suspected, air should not be insufflated through the sigmoidoscope because this will only distend the colon and may increase the inflammation by forcing more feces through a perforated diverticulum. Usually the sigmoidoscope cannot be advanced beyond 12 cm without undue discomfort to the patient. Fiberoptic endoscopy offers little information in the evaluation of suspected diverticulitis and carries significant risk of increasing the pericolic contamination.

Often the diagnosis of diverticulitis can be reasonably ensured based on the history and physical examination. If the diagnosis is in doubt, however, three diagnostic tests can be considered: computed tomography (CT) of the abdomen, ultrasound, and the contrast enema. The abdominal CT and ultrasound are noninvasive tests, and both can demonstrate thickened colonic wall and an associated abscess.

The CT is usually the test of choice to confirm the suspected diagnosis of diverticulitis. It reliably detects the location of the inflammation and provides valuable ancillary information such as the presence of an abscess, ureteral obstruction, or a fistula between the colon and urinary bladder (by demonstrating air in the bladder) (Fig. 26-15 A). If an abscess is present, percutaneous drainage under CT guidance is a valuable therapeutic procedure.

The need for contrast enema in the evaluation of acute diverticulitis has diminished considerably since the introduction of CT scanning and ultrasound. Instilling contrast material under pressure into the colon carries the risk of spreading infection by extravasation through the perforated diverticulum. If the infection has not been well contained by the peritoneal defenses, a localized infection may be converted to generalized peritonitis. For this reason, the use of barium should be avoided; barium combined with feces produces a more severe peritonitis than either agent as an isolated contaminant. Water-soluble contrast enemas do not carry the risk of barium-fecal peritonitis but still may result in extravasation of contrast material from the colon and spreading peritonitis. Colonic wall thickening, diverticula, fistula formation, or displacement of the colon by an abscess may be detected by contrast study. However, these findings generally can be demonstrated more safely by abdominal CT scan.

Treatment

Uncomplicated Diverticulitis

The treatment of diverticulitis depends on the severity of the disease. Patients with minimal symptoms or signs of inflammation can be treated on an outpatient basis with a clear liquid diet and broad-spectrum antibiotics. The antibiotics are continued for 7 to 10 days. Pain medications should be avoided. If the inflammation is so severe as to warrant analgesic administration, the patient should be hospitalized and given intravenous antibiotics. The use of morphine especially should be avoided because it increases intracolonic pressure and can aggravate the inflammatory process. Meperidine decreases intraluminal pressure and is a more appropriate analgesic.

Patients with significant signs of inflammation should be hospitalized for bowel rest, intravenous fluids, and intravenous broad-spectrum antibiotics. Nasogastric suction is usually not necessary unless the inflammation is accompanied by ileus or obstruction.

The patient's symptoms usually respond promptly to nonoperative treatment, with noticeable improvement occurring within 48 h. As the clinical situation permits, diet is resumed and investigative studies may be performed after 3 weeks. A colonoscopic examination is usually indicated, after the inflammation resolves, to evaluate the extent of diverticulosis and to rule out the presence of a cancer. A barium enema may demonstrate the extent of diverticulosis, but this is a less accurate test for identifying small polyps or malignant growths in the presence of numerous diverticula.

When a patient recovers from a simple, uncomplicated episode of diverticulitis, a high-fiber diet is recommended. Surgery is seldom indicated in such circumstances because 70 percent of patients who have recovered from one uncomplicated episode of diverticulitis will have no recurrence. However, if the patient suffers recurrent attacks of diverticulitis, surgical treatment should be considered. The chance of a subsequent attack following a second episode of diverticulitis is greater than 50 percent. Such attacks are treated similarly to the initial episode, and elective resection of the diseased colon is scheduled after the inflammation has resolved, usually 4 to 6 weeks later.

Generalized Peritonitis

Spreading generalized peritonitis may occur if the infection originating from the perforated diverticulum is not immediately localized by the normal peritoneal defenses. If the perforation remains unsealed, the entire peritoneal cavity can be contaminated, with resulting generalized fecal peritonitis. This complication is rare, but immediate surgical intervention is mandatory. Patients complain of severe abdominal pain, and there is voluntary and involuntary guarding in all abdominal quadrants. Intraperitoneal free air may be detected by abdominal x-rays, but the absence of free air does not exclude the diagnosis. Leukocytosis with a left shift is generally observed; however, leukopenia may accompany instances of severe sepsis. Urgent celiotomy is required to control the infection. If the peritonitis is not so severe as to prohibit intestinal resection, the diseased segment of bowel containing the perforation should be resected. The distal segment of bowel should be closed or exteriorized as a mucous fistula, and the bowel proximal to the resected segment is used to construct a colostomy. Since the great majority of diverticular perforations occur in the sigmoid colon, there is seldom sufficient length of distal bowel to reach the abdominal wall as a mucous fistula. Therefore, it is usually necessary to suture or staple the distal bowel (proximal rectum) closed. This operation is frequently called *Hartmann's operation* (see Fig. 26-81) after the French surgeon, Henri Hartmann, who described it as a treatment for proximal rectal cancer in 1921. Hartmann's operation is the most common operation for the emergent treatment of diverticulitis.

If the peritonitis is so severe that resection of the perforated segment cannot be accomplished, then diversion of feces with a completely diverting colostomy constructed proximal to

the perforation should be considered. This is a far less satisfactory procedure and should be avoided because the site of contamination remains in the peritoneal cavity.

Diversion of the fecal stream of the source of infection, accompanied by appropriate antibiotics and nutritional support, should permit resolution of the peritonitis. When the patient has recovered completely, usually over a period of no less than 10 weeks, the colostomy can be taken down and intestinal continuity restored with an anastomosis between the descending colon and the rectum.

Diverticulitis with Abscess

An important advance in the treatment of diverticulitis has been the development of percutaneous drainage of abdominal abscesses. A patient with an abscess that has resulted from a perforation of a diverticulum usually has pain localized to the left lower abdomen. A tender abdominal mass may be palpable, and if the abscess resides low in the pelvis, it may be palpable by digital rectal examination. Abdominal CT scan will confirm the diagnosis (see Fig. 26-15 B). Either of these modalities also permits guided placement of a drain through the abdominal wall into the abscess, allowing evacuation of the purulent material. If the abscess is located low in the pelvis and cannot be approached safely through the abdominal wall, it can be drained safely into the rectum through a transanal or transvaginal approach.

These approaches are immeasurably preferable to a celiotomy, during which the abscess contents potentially could contaminate other areas of the peritoneal cavity. In such circumstances, it is usually necessary to perform Hartmann's operation, requiring the patient to have a temporary colostomy. Another option is resection of the diseased segment with primary anastomosis and proximal diverting colostomy. Percutaneous drainage of the abscess provides safe control of the infection, and an elective resection can be anticipated. At the time of elective surgery, it is usually possible to resect the diseased sigmoid colon and construct an anastomosis between the descending colon and rectum, thus avoiding a colostomy. It is mandatory to excise all abnormally thickened colon and to extend the resection to incorporate normal rectal wall. It is felt that a major cause of recurrent diverticulitis following sigmoidectomy is the failure to completely resect the abnormally thickened muscular wall at the rectosigmoid junction. It is seldom necessary to mobilize the rectum distally beyond 2 cm below the sacral promontory to obtain normal bowel for a satisfactory anastomosis. Although diverticula may be present throughout the colon, it is not necessary to incorporate all of them in the resected specimen. Only the colon with a thickened, brittle wall needs to be excised.

Diverticulitis with Fistula

A fistula between the sigmoid colon and other organs, including bladder, vagina, small intestine, and skin, is a relatively frequent complication of diverticulitis. Diverticulitis is a more common cause of sigmoid-vesical fistula than cancer or Crohn's disease. A fistula is usually formed by an abscess that drains into an organ, establishing a tract between the source of the abscess (the perforated sigmoid diverticulum) and the secondarily involved organ. Fistulas between the sigmoid colon and the bladder are more common in men than in women because the uterus lies between the sigmoid colon and bladder in women. Women with fistulas between the sigmoid colon and bladder or vagina usually have had a previous hysterectomy. A fistula from the sigmoid colon to the skin may result from percutaneous drainage of an abscess caused by diverticulitis.

A fistula between the colon and the urinary bladder presents with recurring urinary tract infections, fecaluria, or pneumaturia. In the presence of distal urinary tract obstruction (prostatic hypertrophy in men), a sigmoid-vesical fistula may be complicated by ascending urinary tract infection and sepsis. CT most accurately confirms the fistula by demonstrating air in the bladder. A barium enema will demonstrate the presence of a fistula in fewer than 50 percent of patients, and intravenous pyelography is even less rewarding. Cystoscopy usually reveals cystitis and bullous edema in the area of the fistula.

The presence of a fistula caused by diverticulitis is seldom a cause for emergency surgical treatment. The formation of a fistula often results in improvement in the patient's condition, since it allows natural drainage of an abdominal abscess. Initial treatment should be directed toward control of any associated sepsis. Sepsis in a patient with a sigmoid-vesical fistula and distal urinary tract obstruction should be treated by relief of the obstruction (with a Foley catheter or suprapubic cystostomy) and appropriate intravenous antibiotics.

Before definitive surgical treatment, the cause of the fistula should be confirmed. The second most common cause of fistula between the sigmoid colon and the bladder is sigmoid carcinoma. Colonoscopy should be used to directly visualize the sigmoid mucosa and to exclude carcinoma, for the curative surgical treatment of sigmoid-vesical fistula caused by cancer involves obtaining wider margins of surgical resection than would be reasonable for a patient whose primary disease is diverticulitis.

After diverticulitis has been established as the cause of the fistula and sepsis has been controlled, treatment is directed toward resolution of the localized inflammation in the region of the fistula. This often can be accomplished by antibiotics administered on an outpatient basis for several weeks. Total parenteral nutrition and bowel rest, concomitant with intravenous antibiotics, have been used in some patients to permit more rapid resolution of inflammation in preparation for surgery.

The surgical treatment of a fistula caused by diverticulitis is excision of the diseased segment of the colon containing the site of perforation. We have found it helpful to place stents in the ureters by means of cystoscopy immediately prior to celiotomy. This technique facilitates identification of the ureter in an area in which the normal anatomy may be distorted or difficult to evaluate because of inflammation. Usually the fistulous tract can be interrupted surgically by blunt dissection. If the fistula involves the bladder, the defect in the bladder can be closed primarily, but unless the defect is large, this is seldom necessary. Providing adequate drainage of the bladder with a Foley catheter or suprapubic cystostomy for a week to 10 days will allow the bladder defect to heal. If the local inflammation is not severe, a one-stage operation usually can be accomplished by sigmoidectomy and anastomosis between the descending colon and rectum. If inflammation is too extensive to permit a primary anastomosis, it may be prudent to perform Hartmann's operation, as discussed previously. On some occasions it may be possible to fashion an anastomosis between the colon and rectum, but adjacent inflammation in the pelvis may be unfavorable, causing the risk of anastomotic leak to be unacceptable. In such instances, the anastomosis can be protected by a proximal transverse colostomy or ileostomy. After the patient has recovered and evaluation reveals the anastomosis to have healed satisfactorily, the proximal stoma can be closed.

On rare occasions, the infection will be so severe that sepsis cannot be controlled with the measures outlined above, and the extent of inflammation will not permit a bowel resection to be accomplished safely. In such instances, fecal diversion is mandatory, with a colostomy constructed proximal to the area of perforation and fistula.

Acquired Vascular Abnormalities

Acute lower gastrointestinal hemorrhage most commonly affects older patients. Although most episodes of bleeding stop spontaneously, between 10 and 25 percent of patients eventually will require surgery to control bleeding. The two most common causes of hemorrhage from the colon are *diverticulosis* and *angiodysplasia*. Both these disorders appear to be acquired, are rare in the young but appear with increasing frequency after age 50, and may occur simultaneously in the same individual.

Hemorrhage from Diverticular Disease

Colonic diverticula are formed at the location where the colonic arterioles penetrate the muscular wall of the bowel, and thus there is an intimate association between the arteriole and the diverticulum (Fig. 26-16). In some instances the artery is displaced over the dome of the diverticulum. With the passage of time, structural changes occur in the wall of the affected blood vessel; there are thickening of the intima and simultaneous focal attenuation of the media. These pathologic changes in arterial structure have been demonstrated in the arterioles associated with diverticula that have been observed to have been the source of hemorrhage. It appears that the vessel is prone to disruption at two sites: the dome of the diverticulum or its neck at the antimesenteric margin. Arterial disruption at either site can result in massive lower gastrointestinal hemorrhage; it is rare that diverticulosis is the cause of significant chronic bleeding. It is curious that the vascular disruption almost always occurs on the mucosal side of the artery, and the bleeding is virtually always into the colonic lumen instead of into the peritoneal cavity.

Hemorrhage from Angiodysplasia

Before the 1970s it was assumed that the great majority of cases of colonic hemorrhage were caused by bleeding from diverticula. Unexplained by this assumption was the fact that the bleeding site, when localized, was in the proximal colon most commonly, whereas diverticula are predominately located in the sigmoid colon. Furthermore, a significant number of patients with colonic hemorrhage had no identifiable colonic diverticula. The introduction of mesenteric arteriography provided a method of identifying the site of hemorrhage and demonstrated that a significant number of cases of colonic hemorrhage were associated not with an arteriole adjacent to a diverticulum but with a different type of vascular abnormality. Various names have been applied to this lesion, including *angiodysplasia*, *vascular ectasia*, *arteriovenous malformation*, and *angiectasia*.

It is apparently an acquired lesion, occurring more commonly with increasing age. It is rarely present in individuals under 40 years of age but becomes increasingly common after age 50, the same age group in which colonic diverticula are found. It has been suggested that the cause of colonic angiectasia is chronic, intermittent obstruction of the submucosal veins (Fig. 26-17). This is apparently caused by repeated muscular contractions of the colon that result in obstruction of the venous system. The lesions are located most commonly in the cecum; according to Laplace's law, the tension generated in the spherical cecum is greater than the tension in the remaining cylindrical colon. The lesions are not confined to the proximal colon; as many as 20 percent are found in the descending and sigmoid colon.

With the passage of time, the venules become dilated; eventually, the process extends to the precapillary sphincters, which become incompetent, forming small arterial-venous communications. Angiographic studies of this pathologic process have correlated the earliest stage of the lesion with a slowly emptying vein or a vein that contains contrast material after the dye has cleared from the normal veins during the venous phase of the mesenteric angiogram. When the process becomes more advanced, dilation of a submucosal vein, a *vascular tuft*, may be recognizable by angiography. If the precapillary sphincters have become incompetent and small arterial-venous communications have developed, the angiogram will reveal the presence of an early-filling vein: a colonic vein that fills within 5 s following the injection of arterial contrast material (Fig. 26-18).

The more advanced lesion of dysplasia also can be identified by colonoscopy as a distinct red mucosal patch. These lesions vary in size from 2 mm to over 2 cm. Histologic study reveals tortuous, dilated capillaries in the submucosa.

Management of Massive Lower Gastrointestinal Hemorrhage

Massive lower gastrointestinal hemorrhage can be defined as bleeding distal to the ligament of Treitz that requires transfusion of more than 3 units of blood over 24 h. The causes of such bleeding are numerous, but diverticular disease and angiodysplasia are the most common causes of life-threatening colonic hemorrhage. The exact incidence of each cause is uncertain. Other causes of lower gastrointestinal hemorrhage include inflammatory bowel disease, ischemic colitis, and tumors. Anticoagulant therapy can also cause gastrointestinal bleeding. While these latter conditions are associated with some blood loss, life-threatening hemorrhage is rare. It is important to remember that gastroduodenal hemorrhage is a frequent cause of rectal bleeding, and this source should be excluded immediately by passing a nasogastric tube into the stomach and confirming the absence of blood in the stomach.

The management of the patient with lower gastrointestinal hemorrhage consists of simultaneously restoring the intravascular volume and identifying the site of the hemorrhage so that appropriate treatment can be given (Fig. 26-19). It is more important to identify the location of the bleeding point than to obtain an immediate diagnosis as to the cause of the bleeding. Lower gastrointestinal bleeding occurs most commonly in elderly patients in whom both angiodysplastic lesions and diverticula may be present. Colonic bleeding will stop spontaneously in 85 percent of patients, and the site of bleeding may be impossible to identify once the hemorrhage has stopped. Celiotomy for massive gastrointestinal hemorrhage of uncertain origin should be avoided except in the most extreme circumstances, for it is rare that the site of hemorrhage cannot be otherwise identified, and even the

extreme operation of total colectomy carries the risk of leaving an unrecognized angiodysplastic lesion in the small intestine.

Blood transfusions should be administered as indicated by the clinical situation, and proctoscopy should be performed to exclude an anorectal cause. The sequence of the next diagnostic step varies somewhat among institutions but consists of nuclear imaging, mesenteric angiography, or colonoscopy. These tests are not mutually exclusive, each has advantages and disadvantages, and all depend on the expertise of the examiner.

Bleeding Scans (Scintigraphy)

Nuclear scintigraphy has been reported by some investigators to be very useful in identifying a lower intestinal bleeding site, whereas others have found this approach to be unreliable and do not recommend it as a preferential test. There are two categories of nuclear scans; the first type uses substances that are cleared from the blood after being injected, whereas the second type employs nondiffusible intravascular indicators (autologous red blood cells labeled with technetium-99m). The technetium- sulfur colloid scan has the advantage of immediate availability and excellent sensitivity. Intestinal bleeding at a rate as low as 0.5 mL/min can be identified by accumulation of the isotope in the intestine at the bleeding site. The patient must be bleeding at the time the isotope is administered intravenously, however, for the label is removed from the circulatory system in about 12 min, accumulating in the liver and spleen. These sites of accumulation also may obscure a bleeding site in the region of the hepatic flexure. The autologous labeled red blood cell (RBC) scan is the scintigraphic test most commonly used in the evaluation of gastrointestinal bleeding (Fig. 26-20 A). The patient's labeled RBCs remain in the circulation for as long as 24 h, permitting monitoring when the bleeding occurs intermittently. In as many as half the patients with the bleeding site identified by this technique, the location does not become evident until more than 6 h after the isotope is introduced. The labeled RBC scan can demonstrate bleeding at a rate of approximately 1 mL/min.

Mesenteric Angiography

Selective mesenteric angiography should be performed once the patient's condition is stable and hydration is adequate. A catheter is placed into the mesenteric circulation using a percutaneous approach, usually the femoral artery, and contrast material is introduced. This approach has the advantage of accurately identifying the bleeding site by visualizing the extravasation of contrast material, but bleeding must be at a rate of at least 1 mL/min for this technique to be successful (see Fig. 26-20 B). Mesenteric angiography carries the distinct advantage of therapeutic options: Vasopressin can be infused through a specific mesenteric artery in an attempt to stop the bleeding by the arterial constriction and muscular contraction of the colonic wall induced by the drug. This approach will stop the bleeding in greater than 85 percent of patients in whom the bleeding site is identified. Vasopressin should not be used in patients with a history of myocardial ischemia. A variety of materials also can be introduced through the mesenteric catheter to embolize the area of bleeding. While success has been reported with this approach, intestinal ischemia is obviously a potential complication. Finally, angiographic study of the bowel may demonstrate the presence of angiodysplastic lesions. Unfortunately, these lesions are relatively common, and the mere presence of such vascular lesions, in the absence of extravasation of vascular contrast material, does not reliably indicate the source of the hemorrhage.

As implied earlier, scintigraphy and angiography are not mutually exclusive tests. If the general location of the bleeding site can be determined by scintigraphy, percutaneous arteriography can be immediately directed toward the correct vessel (superior mesenteric artery for lesions of the ascending or transverse colon, inferior mesenteric artery for sigmoid lesions), and vasopressin can be administered into the appropriate mesenteric artery.

Emergent Colonoscopy

A relatively recent approach to the patient with massive lower gastrointestinal hemorrhage has been the use of colonoscopy. The introduction of orally administered electrolyte lavage permits adequate cleansing of the colon within 2 h, and an experienced colonoscopist can reach the cecum in more than 90 percent of the patients. Colonoscopy has the advantage of providing therapeutic options if the bleeding point is identified. Vasoconstricting agents such as epinephrine can be injected at the bleeding site; vasodestructive agents such as alcohol, morrhuate sodium, or sodium tetradecyl sulfate may eradicate the lesion. Finally, thermal modalities such as laser photocoagulation, electrocoagulation, or heater probe coagulation may control the bleeding source via the colonoscope.

Celiotomy

The approaches discussed above are used to avoid an abdominal operation to control intestinal bleeding from a site that cannot be identified. Every reasonable effort should be made to identify and control the bleeding with nonoperative techniques. There will seldom be any signs that will reliably identify the cause of bleeding at the time of celiotomy. Intraoperative colonoscopy rarely may be beneficial in such circumstances, but usually the bleeding rate is too high to permit adequate visualization of the colonic lumen if emergency celiotomy has been required. In such circumstances, the small bowel should be examined carefully to exclude an obvious source of bleeding, such as lymphoma or Crohn's disease. If no specific lesion can be identified, the treatment of choice is abdominal colectomy. The decision to fashion an ileorectal anastomosis or ileostomy should be based on the patient's overall condition. The incidence of recurrent bleeding after abdominal colectomy is less than 10 percent, but the mortality of such bleeding is significant, with reports ranging from 20 to 40 percent.

Ischemic Colitis

Localized ischemia of the intestine occurs most often in the colon. It may be the result of occlusion of a major mesenteric vessel by thrombosis, embolization, or iatrogenic ligation (after aortic aneurysm resection). Ischemic colitis also has been reported in patients with thromboangiitis obliterans or periarteritis nodosa. In the majority of patients, however, no specific vascular disease or underlying cause can be identified and presumably related to low blood flow. The disease occurs most commonly in the elderly, but it occasionally afflicts young adults. It has been reported to occur in women taking birth control pills.

The pathogenesis of the disease that occurs in the absence of mesenteric arterial occlusion is not well understood. The colonic mucosa becomes edematous, and colonoscopy may reveal dark, hemorrhagic mucosa with patchy ulceration. In mild cases the serosal surface of the colon appears normal, but severe cases of transmural ischemia can cause obvious gangrene of the bowel.

The disease usually presents as mild lower abdominal pain accompanied by the passage of bright red blood per rectum. Physical examination often reveals only mild abdominal tenderness, and proctoscopy reveals normal rectal mucosa with blood coming from above. The diagnosis can be made by contrast enema, which reveals submucosal hemorrhage and edema, or "thumbprinting" of the involved colonic segment. Arteriography usually does not demonstrate specific vascular abnormalities, suggesting that this is a disease of the microscopic vasculature of the bowel. The region of the splenic flexure is most commonly involved, perhaps because this is the area at the periphery of both superior and inferior mesenteric arterial blood supply; however, any area of the large intestine can be affected.

Treatment depends on the severity of the ischemia, as manifested by the patient's signs and symptoms. Patients with mild abdominal pain, minimal leukocytosis, and minimal fever should be observed closely, with the anticipation of recovery within 2 or 3 days. Follow-up colonoscopy or barium enema a few weeks later is usually normal. In some patients, however, the ischemia is more severe, with more marked abdominal pain, fever, and leukocytosis. These patients should be observed closely and treated with intravenous antibiotics. Occasionally the ischemic segment will heal with stricture formation, which may be resected electively if it causes obstructive symptoms.

A small number of patients have severe ischemia involving the full thickness of the colonic wall. This condition is characterized by signs of bowel necrosis with severe abdominal pain, peritonitis, and systemic toxicity. Immediate celiotomy is required, with resection of the ischemic intestine and proximal colostomy or ileostomy. The morbidity and mortality are high in this group of patients.

Radiation Proctocolitis

Radiation injury to the intestine is most commonly associated with pelvic irradiation for uterine or cervical cancer, but in recent years high radiation doses have been used for the treatment of prostatic and bladder cancer, as well as adjuvant or even primary treatment for selected cases of rectal malignancy. The new high-energy radiation sources deliver large doses of radiation to the viscera with sparing of the skin; thus the treating physician may not be warned of prior intestinal radiation injury.

Individual responses to radiation vary considerably, but there is a rough correlation of bowel injury with the dosage of radiation. Tissue doses of less than 4000 cGy are seldom associated with intestinal damage. Most patients will exhibit some degree of intestinal injury after exposure to 5000 cGy, and exposure to radiation in excess of 6000 cGy is associated with symptoms of radiation enterocolitis in the vast majority of patients. Factors that increase the likelihood of intestinal complications related to radiation include arteriosclerotic vascular disease, diabetes, hypertension, advanced age, and adhesions from previous abdominal operations that fix the bowel in the pelvis.

Radiation injury to the intestine is manifested in early and late stages. During radiation therapy the proliferating intestinal epithelial cells are affected, and the mucosa becomes edematous and hyperemic. Acute ulceration may occur if reepithelialization is significantly impaired. After therapy is completed, the mucosa usually heals, although the villi are shortened. Patients having radiation to the intestine often have nausea and vomiting, cramping abdominal pain, and diarrhea during the course of treatment. Tenesmus and rectal bleeding may occur during high levels of pelvic irradiation. Usually these symptoms can be managed symptomatically, but sometimes the radiation dosage must be decreased or treatment interrupted until the symptoms subside.

The late stages of intestinal radiation injury occur from weeks to years following completion of radiation therapy and are the manifestations of a progressive vasculitis affecting the arterioles of the submucosa. This radiation arteritis is characterized by thickening of the arteriole wall with subsequent microvascular thrombosis. This chronic ischemic process is associated with thickening and fibrosis of the intestinal wall and ulceration of the mucosa. Progressive ischemia can result in stricture formation, intestinal perforation, or abscess and fistula formation. The rectum is by far the most common site of involvement because of its location in close proximity to the uterus, cervix, and prostate.

Radiation proctitis is associated with tenesmus and bleeding. These symptoms may occur from months to years after completion of therapy and are the result of mucosal ulceration secondary to radiation endarteritis and ischemia. The symptoms often can be managed symptomatically with stool softeners, topical 5-aminosalicylic acid preparations, and occasionally, corticosteroid enemas. Should these measures fail, topical application of 10% formalin to the affected mucosa has been very efficacious. On rare occasions, a proximal colostomy is required for severe tenesmus or bleeding. It is extremely rare for proctectomy to be necessary to control symptoms.

Rectal stricture following pelvic radiation may cause symptoms of colonic obstruction. Recurrent cancer should be excluded by careful endoscopic evaluation and biopsies of suspected radiation-induced rectal strictures. If gentle dilation does not relieve the obstructive symptoms, diverting colostomy may be necessary.

Radiation-induced rectovaginal fistula most commonly occurs in women who have been treated for carcinoma of the uterine cervix. The fistula usually occurs within 2 years after completion of radiation therapy, but it may occur decades later. Patients usually note bleeding and rectal pain in the weeks preceding the passage of stool and gas through the vagina. A rectal stricture is often associated with the rectovaginal fistula, and occasionally there is also a concomitant fistula between the urinary bladder and vagina.

A proximal colostomy is invariably required to relieve the symptoms of radiation-induced rectovaginal fistula. It is mandatory to exclude recurrent carcinoma as the cause of the fistula, a diagnosis that is at times difficult and may require examination under anesthesia. If cancer is excluded and the patient's health is satisfactory, consideration should be

given to treating the fistula by resecting the involved rectum and restoring continuity by low colorectal anastomosis or, more commonly, by coloanal anastomosis. Local excision and closure of the fistula are rarely successful, even when preceded by a diverting colostomy. Bricker has advocated repair of rectovaginal fistula caused by radiation using a technique of reconstructing the anterior rectal wall with nonirradiated sigmoid colon brought down as a fold-over patch graft. All these approaches to repair a complicated radiation-induced fistula require protection by a temporary colostomy or ileostomy. Often the entire pelvis is “frozen” with radiation- injured structures, including multiple loops of small intestine. In this situation, complete fecal diversion with a colostomy should be the definitive treatment, avoiding injury to the irradiated small intestine, which has limited healing capability.

Normal Function of Anorectum

The rectum functions mostly as a storage capacitance vessel. The rectum has very little peristaltic function of its own and relies on external pressure to empty. The outer longitudinal muscle is thick and has some contractility but has lost the organization of the teniae found on the colon. The rectum has a normal manometric resting pressure of approximately 10 mmHg, mostly due to intraperitoneal pressure and resting muscle tone. Conditions such as Crohn's disease or radiation injury cause the rectum to lose its natural compliance. This loss of compliance and capacitance is occasionally incapacitating to the patient.

The normal rectum can hold 650 to 1200 mL of liquid. A rectum that holds more than 1500 mL can be classified as a megarectum. The normal daily volume of stool eliminated by the rectum ranges from 250 to 750 mL of formed feces.

The anal sphincter mechanism is the other component of defecation and continence. Its anatomy and innervation have been described previously (see Fig. 26-5). The external sphincter fibers are responsible for 20 percent of the resting pressure and 100 percent of generated squeeze pressure. The internal sphincter provides 80 percent of anal resting pressure. Both the internal and external sphincter muscles are contracted at rest.

Defecation and continence are coordinated mechanisms. *Continence* can be described as controlled elimination of the rectal contents at a socially acceptable time and place. The coordination of rectal emptying and sphincter contraction and relaxation is very complex, and there are numerous abnormalities that occur.

Defecation can be divided into four components. The first is movement of feces into the rectal vault or capacitance organ. A mass peristaltic wave in the proximal colon and sigmoid colon occurs two or three times per day to pass solid substance into the rectum. The gastrocolic reflex is a well-known phenomenon that results in colonic mass peristaltic movement after distention of the stomach, probably hormonally mediated.

The second component of defecation is the *rectal-anal inhibitory reflex* or *sampling reflex*. Distention of the rectum results in involuntary relaxation of the internal anal sphincter and allows sensation of the rectal contents at the transitional zone. The sampling reflex has been shown by ambulatory manometry to occur frequently throughout the day and night.

The third component of defecation is voluntary relaxation of the external sphincter mechanism. Voluntary relaxation of the pelvic floor, puborectalis muscle, and external sphincter allows the rectal contents to be pushed farther into the anal canal and expelled. The relaxation of the sphincter mechanism is actually a failure to contract rather than an active relaxation, because the mechanism is paradoxically contracted when the rest of the individual is relaxed, yet continent (Fig. 26-21).

The fourth component of defecation is the voluntary increase of intraabdominal pressure, using the diaphragm and abdominal wall muscles. This increase in pressure serves to propel the rectal contents through the anal canal and accomplish defecation.

The passage of flatus also requires coordination of multiple factors. The sensation of gas at the transitional zone and in the anoderm informs the individual that gas is present to be eliminated. If the situation is such that full evacuation of the rectum is not possible but elimination of gas is desired, a voluntary contraction of the pelvic floor including the puborectalis and external sphincter muscles occurs to prevent loss of solid rectal contents. With an increase in abdominal pressure and a coordinated relaxation of some of the external sphincter, selective passage of flatus may be accomplished. During defecation, the gaseous contents of the rectum will be expelled with the solid contents without discrimination. This mechanism of discrimination appears to be learned. Patients eventually can pass flatus selectively even after the rectum has been removed and replaced by a reservoir of ileum for diseases such as ulcerative colitis and familial adenomatous polyposis.

Continence, or the control of rectal contents, requires an adequate rectal capacity and normal compliance. It may be difficult to retain rectal contents in conditions such as Crohn's disease, in which the rectum becomes a rigid tube rather than a soft distensible bag, even if the external and internal sphincter mechanisms work properly. Adequate sensation at the transitional cell zone is required to coordinate pelvic pressure and sphincter tone during the sampling reflex. The external sphincter is most responsible for the fine control of solid, liquid, and gas. The puborectalis muscle has been proposed as the mainstay of the sphincter mechanism and is probably responsible for the control of solid stool. The internal sphincter may be responsible for fine control of gas on the basis of the sampling reflex and constantly provides resting pressure to prevent release of flatus. The pudendal nerves provide both the sensory afferents of the anal canal and the motor efferents to the voluntary muscles of the anal canal.

Dysfunction of Anorectum

Incontinence

Anal incontinence can be defined as the inability to control elimination of rectal contents. Incontinence can be categorized as either constant or intermittent. It also can be characterized by the type of material that the anal canal is unable to control. Incontinence to solid, liquid, and gas is complete incontinence. Incontinence to liquid and gas and incontinence to gas only represent degrees of partial incontinence.

Etiology

Anal incontinence can be caused by mechanical or neurogenic defects in the anal sphincter mechanism. *Mechanical defects* result from trauma to the anal sphincter mechanism. The most common cause of sphincter trauma is obstetric injury during vaginal delivery. The true incidence of anal incontinence after vaginal delivery is unknown, but there is an increased likelihood of incontinence after a midline episiotomy or a perineal tear that extends into the rectum (fourth degree), after multiple vaginal deliveries with midline episiotomies, after infection of an episiotomy repair, and after prolonged labor followed by traumatic vaginal delivery.

The treatment of cryptogenic abscess/fistula disease of the anal canal by fistulotomy also can cause anal incontinence. A fistulotomy performed through a significant amount of external sphincter, in the anterior quadrant, often causes incontinence. This is especially true in women because continence depends on one thin band of anterior external sphincter muscle.

Some systemic diseases also can cause mechanical malfunction of the anal sphincter and result in incontinence. Scleroderma, for example, produces a progressive sclerosis and fibrosis of the muscle and nerve endings and prevents normal function. The internal sphincter is known to become dense, spastic, and occasionally stenotic. The external sphincter may fail to function voluntarily, resulting in total incontinence.

Neurogenic causes for anal incontinence include isolated pudendal nerve injury as well as systemic neurologic disease. Stretch or trauma to the pudendal nerve during prolonged labor at childbirth has been documented to cause partially reversible pudendal nerve injury. Lifelong straining to have a bowel movement has only recently been recognized as a cause of pudendal nerve injury. The pudendal nerve is stretched over the ischial spine as the perineum descends during straining to eliminate the rectal contents. This form of neurogenic incontinence occurs most commonly in elderly women and has been called *idiopathic fecal incontinence*. Systemic diseases may affect selected nerves (e.g., multiple sclerosis) or may cause diffuse damage that affects all nerves and eventually causes incontinence by its effect on the pudendal system (e.g., diabetic neuropathy).

Other causes of anal incontinence that are unrelated to anal sphincter malfunction include severe diarrhea, fecal impaction with overflow incontinence, irritable bowel syndrome, radiation proctitis and fibrosis, inflammatory bowel disease of the rectum, and tumors of the distal colon and rectum.

Physiologic Evaluation of Disordered Anorectal Function

The visual and digital rectal examination remains very important in the evaluation of a patient with anal incontinence. The perineal body is often thin in the setting of incontinence after obstetric trauma. There is usually a scar present when this is the source of the injury. Occasionally fistulous tracts or fistulotomy scars will be evident. An estimate of the resting tone and squeeze pressure can be obtained from the digital rectal examination.

Anal manometry measures the resting pressure, the squeeze pressure, the sphincter length, and the minimal sensory volume of the rectum. Maximal rectal tolerance also can be measured. There are various techniques for performing anal manometry. Normal resting pressure is usually reported as 40 to 80 mmHg. The maximal squeeze pressure generated by the external sphincter is 80 to 160 mmHg, which is twice the normal resting pressure. The average sphincter length of 3 cm is based on measurement of the high-pressure zone of the anal canal. The sphincter is shorter anteriorly where the puborectalis muscle does not complete the circle of fibers at the upper extent of the anal canal. The entire sphincter mechanism is shorter in women than in men. Normal minimal sensory volume is less than 10 mL of air instilled into a latex balloon placed within the rectum just above the anal canal. Resting *rectal* pressure is normally 10 mmHg (Fig. 26-22).

The measurement of the pudendal nerve *terminal motor latency* uses a stimulating electrode on the tip of a gloved finger that is placed over the ischial spine to stimulate the pudendal nerve. The resulting contraction of the external sphincter muscle is detected by means of a recording electrode at the base of the same finger. Normal pudendal nerve terminal motor latency is 2.0 ms, and it has been shown to increase with age, after childbirth, after prolonged excessive straining with perineal descent, and in systemic diseases such as diabetes and multiple sclerosis.

Nerve injury to the sphincter mechanism also can be determined by the electromyographic measurement of single-fiber density of the external sphincter muscle. This uses a thin bipolar needle with a 250-mm recording surface on the side of the needle. Motor unit potentials from a bundle of muscles innervated by a single terminal branch of the pudendal nerve are recorded. The external sphincter and puborectalis muscle maintain a constant resting tone, with spontaneous motor unit potentials documented in the muscle at rest. Injury to a terminal branch of a nerve is demonstrated by a multiphasic reinnervation pattern. As more muscle bundles are reinnervated, the average fiber density in numerous samplings may increase beyond the normal limits, also indicating nerve injury. Surgical correction of a mechanical defect in the muscle will not be successful if the innervation is severely impaired. Electromyographic techniques are also used to “map” the anal sphincter for viable, contracting motor units to document the extent and site of mechanical injury.

Transrectal ultrasound is as accurate as needle electromyography for sphincter mapping and is better tolerated by patients. A 360° 10-mHz probe can image the puborectalis, external anal sphincter, and internal anal sphincter and identify defects in these muscles.

Surgical Correction

If a patient has a minor defect demonstrated by anal manometry or has not responded completely to anal sphincter reconstruction, muscle/sensory retraining may be helpful using biofeedback techniques to improve anal continence. The majority of mechanical problems in the anal sphincter can be corrected by restoring the circular integrity of the sphincter mechanism. The technique most commonly used is the overlapping muscle repair (Fig. 26-23). In a patient with a defect in the circular fibers of the external and internal sphincters, the muscle can be mobilized and the ends of the scar and muscle brought together, in overlapping or end-to-end configuration, to complete the circular ring of the anal sphincter. The normal sphincter length is restored in the anterior quadrant, which may in itself provide a barrier to leakage of liquid and gas.

Extensive injury to the muscles and nerves of the anal sphincter may preclude simple muscle repair. In such cases, some success has been reported from use of the gracilis or gluteal muscle to create an encirclement of the anus; the degree of true continence that this provides is still questioned. Studies are evaluating the use of implantable pacemakers to drive the denervated external sphincter muscle or a transferred muscle, such as the gracilis or gluteus. These techniques are usually reserved for complete denervation of the anal canal or for severe mechanical injury in a patient who wishes to avoid a colostomy. The complicated process of diagnosing and treating anal incontinence is summarized in Fig. 26-24.

Obstructed Defecation

Obstructed defecation or pelvic floor outlet obstruction is a disorder of the anorectum resulting from nonrelaxation of the puborectalis muscle, or internal intussusception of the rectum, these disorders being the major components of a spectrum of disturbances known as *pelvic floor abnormalities*, the end stage of which is complete rectal prolapse.

Anal stenosis is a rare cause of obstructed defecation that results in thin, frequent stools, abdominal bloating, and dilatation of the large intestine. The most common cause of anal stenosis is scarring after anal operation such as an improperly performed circumferential Whitehead hemorrhoidectomy. Other causes of anal stenosis include chronic laxative abuse (particularly mineral oil), radiation injury, recurrent anal ulcer, Crohn's disease, trauma, and anal tumors. Occasionally anal stenosis will respond to dilation. The so-called Whitehead deformity includes a scarred stenosis of the anal canal and an ectropion of mucosa that is pulled from the rectum onto the perineal skin. Both anal ectropion and stenosis can be corrected by advancing a full-thickness pedicled flap of skin into the circumferential scar (Fig. 26-25). The success of this procedure depends on the presence of normal tissue surrounding the obstructing scar.

Nonrelaxation of the puborectalis muscle is a functional disorder in which a normal muscle displays abnormal function or control. Patients complain of straining, incomplete evacuation, need for digital maneuvers to eliminate the rectal content, and anal or pelvic pain. The diagnosis of nonrelaxing puborectalis muscle as a cause of constipation symptoms relies on the documentation of normal colonic transit to the rectum. A persistent image of the puborectalis muscle distorting the posterior rectum is seen on defecography (Fig. 26-26). Defecography is accomplished by instilling methylcellulose-thickened barium into the rectum. Video cinefluoroscopy is used to record a lateral image of the rectum while the patient strains in the sitting position. The inability to expel an air-filled latex balloon from the rectum while sitting on a toilet may be a more accurate means of documenting nonrelaxation (or paradoxical contraction) of the puborectalis. The treatment of this problem is nonoperative. Biofeedback is the most effective method for teaching the patient the cognitive aspects of defecation.

Internal intussusception (or internal prolapse of the rectum) is another cause of outlet obstruction. This condition accompanies laxity of the rectal fixation at the sacrum (Fig. 26-27) and will be discussed in more detail later. The intestine telescoping into itself causes partial obstruction to defecation while simultaneously contributing to a feeling of rectal fullness and a need to defecate.

Other Associated Problems

Proctalgia fugax, or the *levator muscle syndrome*, is spasm of the anal sphincter, which causes pain, often at night. The puborectalis muscle goes into spasm and typically awakens the patient. The pain usually lasts only a few minutes and can be relieved by heat, dilation, or muscle relaxants. It may be stress-related and has been associated with the irritable bowel syndrome or a past history of migraine headaches. Other causes of anal pain such as abscess, foreign body, or cancer should be excluded before this diagnosis can be confirmed.

Fecal impaction and *stercoral ulceration* are occasionally seen in patients with outlet obstruction due to nonrelaxing puborectalis or internal intussusception. The symptoms of bleeding and rectal fullness are treated by regulation of the bowel habits and periodic emptying of the rectum with a bowel regimen consisting of enemas, increased fiber, and stool softeners.

Descending perineum syndrome was in the past thought to be a primary process. It is now known to be the result of chronic straining at defecation. The primary process is, therefore, chronic constipation, disordered defecation, or intussusception. Descending perineum syndrome causes stretching of the pudendal nerve, which leads to neurogenic weakening of the external sphincter and puborectalis muscle, which eventually can lead to anal incontinence.

Rectocele is the result of weakened muscle and soft tissue between the rectum and the vagina after years of straining to defecate. The bulge, or rectocele (see Fig. 26-26 B), occurs in the posterior vaginal wall and in severe cases may even cause protrusion of tissue through the vagina. Symptoms are similar to those of other types of obstructed defecation, and the patient may need to put a finger in the vagina and push the rectum posteriorly in order for it to empty. A rectocele is a signal that a basic problem exists in the rectum or mechanism of defecation. Repair of a rectocele without correction of internal intussusception or outlet obstruction is seldom successful, and recurrence is common in such circumstances.

Psychological Profile of an Outlet Obstruction Patient

Psychological abnormality has never been clearly documented as the cause of disordered defecation, but many patients with constipation have been found by psychological evaluation to have depression, somatization disorder, and an extreme need to control. Patients with nonrelaxing puborectalis have similar findings. These patients also seem to exhibit a cancer fear, which is alleviated when they find that their work-up is normal. It is prudent to obtain a psychological profile on these patients before considering biofeedback or surgery.

Physiologic Evaluation of Disordered Defecation (Pelvic Floor Abnormality)

Disordered defecation or pelvic floor abnormality is a complex problem. An overlap in symptoms between causes may result in a diagnostic dilemma unless a well-planned evaluation has been performed (Fig. 26-28). *Defecography* uses cinefluoroscopy from a lateral position as the patient attempts to empty thickened barium while sitting on a commode. The features of a *normal defecogram* include fixation of the posterior rectum to the curve of the sacrum, rapid rectal emptying with straining, and appropriate relaxation of the puborectalis muscle. Abnormal findings are observed when the rectum moves from the sacrum during straining, when it displays excessive mobility from the sacral curve, and when it forms a funnel configuration (complete intussusception) (Fig. 26-29).

Complete rectal prolapse also can be documented by defecography. Obstructed defecation is suggested if the puborectalis does not relax during attempted defecation (see Fig. 26-26 A).

Colonic *transit time*, as described previously, is important in differentiating colonic from anorectal malfunction in patients with disordered defecation. Outlet obstruction may exist with normal colonic transit to the rectum. Colonic transit studies will show markers within the rectum after the fifth day because the rectum does not empty normally.

Anal manometry rarely has a role in the evaluation of disordered defecation. Its main use in this context is to document the presence of a normal rectal-anal inhibitory reflex to rule out Hirschsprung's disease as the cause of constipation. *Megarectum* is an acquired problem thought to be a result of failure to respond to the normal signals for defecation and can result in an abnormal rectoanal inhibitory reflex. Manometry also shows an increase in the maximal tolerable rectal volume in patients with megarectum.

Balloon expulsion recently has been found to be helpful in establishing the diagnosis of nonrelaxing puborectalis muscle as the cause for disordered defecation. Defecography often reveals a persistent puborectalis shadow and poor relaxation of the sphincter during efforts to evacuate the rectum; however, many of these patients are able to empty their rectum at other times. The overdiagnosis of nonrelaxing puborectalis muscle can be avoided by performing a balloon expulsion test. The inability to pass a 60- mL inflated latex balloon in a private bathroom, sitting on a toilet, seems to be an accurate means of documenting the absence of the normal coordination of muscle function required for defecation.

Therapeutic Options for the Treatment of Disordered Defecation

Nonrelaxation of the puborectalis muscle has been treated with limited transection of the puborectalis muscle itself, an approach that initially was reported to cause improvement in some patients with outlet obstruction; long-term follow-up revealed that almost all patients treated with the operation develop incontinence with very little improvement in symptoms. The most consistent improvement in patients with nonrelaxing puborectalis muscle has come via the use of biofeedback. Several series have now documented a long-term improvement using behavioral modification techniques to retrain the patient to relax the pelvic floor during straining to defecate.

The anatomic abnormalities associated with internal intussusception can be corrected surgically. Retrorectal sacral fixation straightens the rectum and fixes the bowel to the sacrum, eliminating the abnormal mobility (Fig. 26-30 A). Unfortunately, this operation often fails to relieve the symptoms of disordered defecation. Success has been reported with low anterior resection of the rectum (Fig. 26-30 B), but more data concerning long-term results are needed. Most of the patients with internal intussusception as a cause of outlet obstruction respond well to medical management, once the cause of symptoms has been clearly diagnosed. Only those with the most severe problems and impending pudendal nerve damage should be considered for operative correction. It is important to remember that if an outlet obstruction is not corrected before other causes of constipation are treated, the method of treating the constipation will fail. Even the use of abdominal colectomy for severe constipation due to colonic inertia has failed if outlet obstruction is a major component of the problem.

Abnormal Rectal Fixation Internal Intussusception

Internal intussusception has already been described as a cause of outlet obstruction. It is also possible that a nonrelaxing puborectalis muscle induces internal intussusception from prolonged, aberrant straining at defecation. It is further thought that internal intussusception is a precursor to complete rectal prolapse. There seems to be an increased incidence of intussusception of the rectum with aging, connective tissue disorders, and irritable bowel syndrome. Internal intussusception is also now known to be the cause of the *solitary rectal ulcer*. The ulcer develops as the lead point of the intussuscepting tissue becomes chronically ischemic and traumatized. If the ulcer heals, glands may be trapped in the deep tissues, causing the lesion of *colitis cystica profunda*, large glandular lakes within the mucosal and submucosal tissue in the area of the solitary rectal ulcer. Colitis cystica profunda is characterized by the presence of localized or diffuse submucosal cysts and may be mistaken for rectal cancer, resulting in inappropriate resection. It may be found in association with several different clinical conditions, including radiation injury, postoperative changes, adenomatous polyps, malignant growth, or inflammatory bowel disease. Patients may have symptoms of mucous discharge, rectal bleeding, or tenesmus. Feelings of rectal fullness or incomplete evacuation are common patient complaints. Colonoscopy or barium

enema should be performed to rule out malignant growth, and the patient should be evaluated for rectal prolapse or intussusception by defecography. Spontaneous resolution can occur but is rare. Colitis cystica profunda associated with rectal prolapse is treated by correcting the rectal prolapse.

The majority of patients with internal intussusception respond to increased bulk in their diet, stool softeners, and glycerine suppositories or small enemas to stimulate rectal emptying. Indications for surgical correction of internal intussusception include symptoms severe enough to change the patient's life despite maximal medical and psychological therapy, chronic bleeding from a solitary rectal ulcer, or impending anal incontinence because of the descending perineum and subsequent stretch injury to the pudendal nerve.

Rectal Prolapse

Rectal prolapse is the protrusion of the full thickness of rectum and occasionally sigmoid through the anus. Rectal prolapse may be the long- term result of outlet obstruction or internal intussusception of the rectum. There is an increased incidence in elderly women and institutionalized patients treated with constipating psychotropic drugs. There is also an increased incidence of prolapse in women who have had hysterectomy and in spinal cord–injured patients who require routine use of laxatives to maintain bowel function. Surprisingly, there is not an increased incidence of rectal prolapse in women who have had multiple vaginal deliveries.

The symptoms of rectal prolapse include anorectal pain, bleeding, mucous discharge, and incontinence. Some patients will be hesitant to describe the large bolus of protruding tissue and will instead present with symptoms of mucous discharge or progressive incontinence. As the condition progresses, spontaneous reduction may not occur, and manual replacement of the prolapsed rectum becomes necessary. In severe instances, the prolapsed tissue becomes too edematous to replace, and necrosis of the prolapsed segment may occur. Sometimes the patient is unable to demonstrate the prolapse in the physician's office, and defecography is required to reveal increased mobility from the sacrum, marked intussusception, and finally protrusion of the rectum through the anal canal. Occasionally, a solitary rectal ulcer is seen at the lead point of the prolapse. It is important to differentiate rectal prolapse from prolapsing internal and mixed hemorrhoids. In hemorrhoidal disease, there will be a deep groove between the areas of prolapsing tissue as opposed to the concentric rings on the mucosa seen on the rectal prolapse. It is a significant mistake to perform a hemorrhoidectomy when the problem is really complete rectal prolapse.

Complete evaluation of the patient with rectal prolapse should include colonoscopy or air-contrast barium enema to fully evaluate the colon and rule out the presence of a malignant growth that might be serving as the lead point for the prolapse. Anal manometry and electromyography are helpful in evaluating sphincter function in patients with rectal prolapse in order to choose the appropriate corrective procedure.

Therapeutic options in the treatment of rectal prolapse can be divided into three groups (see Fig. 26-30). The first group, patients with full rectal prolapse and normal sphincter function, should be considered for a low anterior resection or retrorectal sacral fixation. If constipation is a significant factor, or if the patient is found to have a very redundant sigmoid colon, low anterior resection with fixation of the lateral rectal ligaments is probably a better choice of procedure. Any operation for rectal prolapse must include full mobilization of the rectum, preceding fixation of the rectum, to prevent forward motion and avoid distal prolapse. Excellent functional results should be anticipated for this group of patients.

The second group of patients has full rectal prolapse and incontinence. Most patients recover sphincter function after repair of the prolapse (60 to 70 percent), but it is difficult to predict which of these patients would tolerate a low resection of the rectum. These patients should be treated by retrorectal sacral rectopexy using a synthetic sling via an abdominal approach or by a perineal proctectomy combined with rectopexy and a posterior sphincter-enhancing procedure (see Fig. 26-30 C). Of patients having surgical correction of rectal prolapse, 30 percent will not recover fecal continence. This is true especially in patients with documented neurogenic incontinence who have prolonged pudendal nerve latency. These patients may be helped by the "posterior repair" of the anal canal using the technique described by Sir Alan Parks. The external sphincter and puborectalis sling are brought together posteriorly to elongate the anal canal, to decrease the anal opening, and to add bulk to the posterior aspect of the canal. In some cases continence is further improved by reconstruction of the anterior aspect of the sphincter as well as posteriorly. This technique has been combined with a perineal proctectomy to correct rectal prolapse and associated anal incontinence.

The third group of patients consists of those at extremely high operative risk or elderly individuals with incontinence and prolapse who have a very short life expectancy. These patients may be considered for an anal encircling procedure (see Fig. 26-30 D). The original procedure used a silver wire (Thiersch procedure) placed in the ischiorectal fat around the external sphincter mechanism. The wire served to cause obstruction of the anal outlet and to compress the sphincter mechanism. Currently, the encircling procedure is performed with a band of soft synthetic mesh, a silicone tube, or a gracilis muscle graft to the anal canal. The majority of these patients require laxatives or enemas to avoid fecal impaction. The complications of this procedure include erosion of the encircling band and infection. This is obviously not the first choice of treatment for rectal prolapse and is used less frequently because the combined perineal proctectomy procedure is available for use in elderly patients under spinal or epidural anesthesia.

Hemorrhoids Pathophysiology

Hemorrhoidal tissue is normally found at the distal end of the rectum within the anal canal (Fig. 26-31 A). These vascular and connective tissue cushions are usually found in the right anterolateral and posterolateral positions and the left lateral position. Internal hemorrhoids are above the dentate line (see Fig. 26-31 B); external hemorrhoids are the vascular complexes under the anoderm of the anal canal (see Fig. 26-31 C).

The pathophysiology of symptomatic hemorrhoids is related to engorgement of the vascular pedicles in the hemorrhoidal complexes, which then dilate, stretch, and cause the cushions to enlarge. The vascular complexes engorge as pressure is applied to the pelvic floor during straining, lifting, or standing. Bleeding occurs from local trauma to the hemorrhoidal complex, usually during defecation. Hard stools, prolonged straining, increased abdominal pressure, and prolonged lack of support to the pelvic floor all combine to increase the likelihood that abnormal hemorrhoidal tissue will develop.

Internal hemorrhoids, which are covered by mucosa, typically bleed or prolapse but do not cause pain. Patients complain of rectal fullness, mucous discharge, and bright red blood dripping into the toilet or on the toilet paper. Occasionally, internal hemorrhoids will incarcerate within the outer ring of the anal canal and develop thrombosis and necrosis. Internal hemorrhoids can be classified as follows: First-degree hemorrhoids bleed; second-degree hemorrhoids bleed and prolapse but reduce spontaneously; third-degree hemorrhoids bleed and prolapse and require manual reduction; and fourth-degree hemorrhoids bleed and incarcerate and cannot be reduced.

External hemorrhoids beneath the anal skin can become enlarged over time as a result of dilatation or repeated thrombosis. The overlying skin may stretch to develop a skin tag, which prevents adequate hygiene. Occasionally, a clot within an external hemorrhoid will cause severe pain and may bleed if the clot erodes through the overlying skin. It is rare for an internal hemorrhoid to cause pain unless it becomes incarcerated and thrombosed. External hemorrhoids usually cause pain even after a small thrombosis because they are beneath richly innervated skin.

The symptoms of hemorrhoids are limited, and caution must be used not to misdiagnose other anorectal problems as hemorrhoids (Table 26-1). Defecography is helpful when symptoms of rectal prolapse or obstructed defecation exist, and sigmoidoscopy, colonoscopy, or contrast enema identifies malignant growths or inflammatory bowel disease.

Treatment

Medical Therapy

The majority of bleeding first- and second-degree hemorrhoids respond to the addition of dietary fiber (e.g., psyllium) or stool softeners and avoidance of straining or prolonged sitting on the commode. It is important to counsel the patient to ingest an appropriate amount of water with the fiber. The goal is to produce soft stool that is easy to pass. Failure of medical therapy is an indication for mechanical intervention, e.g., elastic ligation, excision, etc. of second-degree hemorrhoids. Surgery rarely is indicated for the bleeding of first-degree hemorrhoids.

Elastic Ligation

Elastic ligation of second- and third-degree internal hemorrhoids, 1 to 2 cm above the dentate line, is very effective for control of bleeding and prolapse. The tissue is grasped and pulled into the cylinder of a rubber-band applier, and the rubber bands are placed at the base of the hemorrhoidal tissue (Fig. 26-32). The tissue necroses, sloughs, and leaves a scar in the area of the vascular pedicle, which prevents further prolapse of tissue or bleeding. Care is taken to avoid placement of the bands on the transitional zone or anoderm, both of which are richly innervated with pain fibers. One also must avoid placement of the rubber bands on the deep internal sphincter muscle, since symptoms of spasm may become so severe that urinary retention may result. The majority of patients can tolerate a single- quadrant ligation in the office every 2 weeks. If symptoms are severe or the patient has significant external hemorrhoids and other anorectal problems, three quadrants may be banded in the operating room under local anesthesia.

Severe sepsis has been reported after ligation of hemorrhoids in immunocompromised patients or patients with pelvic floor abnormalities in whom the full thickness of the distal prolapsing rectum has been ligated. This life-threatening complication is heralded by inordinate pain, fever, and urinary retention within 12 h of ligation. The incidence of urinary retention after a single ligation should be less than 1 percent. Urinary retention after multiple ligations may be as high as 10 to 20 percent. The septic patient requires intravenous antibiotic coverage for gram-negative and anaerobic organisms, removal of the rubber bands, and hospitalization.

Normally, the rubber bands fall off with the necrotic tissue after 7 to 10 days. Bleeding may occur, but it is usually minimal. One percent of patients develops significant hemorrhage, which may require cautery or suture ligation of the bleeding site. Nonsteroidal anti-inflammatory drugs or aspirin taken during the 7 to 10 days after banding may increase the incidence of bleeding when the bands and tissue slough.

Photocoagulation

This is a very simple technique by which heat causes necrosis fibrosis of the vascular pedicle in several areas within the anal canal to achieve the same result as elastic ligation on smaller pedicles, but larger hemorrhoids require another form of treatment.

Sclerosis

A mixture of 5% phenol and almond oil developed at St. Mark's Hospital in London may be used to stop bleeding from first- and second-degree hemorrhoids. The mixture is injected through a spinal needle into the submucosa at the base of each hemorrhoid to raise a bleb of mucosa. The depth and amount of sclerosing agent are critical to avoid slough of the entire anal mucosa. Minimizing the amount of IV fluids given during an operation decreases the rate of this complication. Other complications include allergic reaction, local infection, and prostatitis. Many surgeons believe that if tissue can be treated by sclerosis, it probably should not have surgical treatment at all.

Cryosurgery and Direct-Current Coagulation

Hemorrhoidal tissue can be coagulated by freezing with liquid N₂ or a CO₂ probe. Because frozen tissue undergoes liquefaction necrosis and there is poor control of the depth of

freezing, the majority of surgeons avoid using this technique. Patients also complain of profuse, foul-smelling seropurulent discharge.

Direct-current coagulation requires approximately 10 min per site to coagulate the hemorrhoidal cushion. The resulting scar prevents prolapse and further bleeding of first-, second-, and some third-degree internal hemorrhoids. The technique is simply another means of treating internal hemorrhoids in the outpatient or office setting.

Excisional Hemorrhoidectomy

The excision of hemorrhoids should be limited to large third- and fourth- degree hemorrhoids that cannot be treated on an outpatient basis, mixed hemorrhoids with an external anoderm component that are not amenable to ligation, and acutely thrombosed, incarcerated hemorrhoids with severe pain and impending gangrene. Patients on anticoagulant medication are also better treated with closed, excisional hemorrhoidectomy rather than elastic ligation because of the risk of excessive bleeding when the necrotic tissue sloughs. There are numerous techniques for excising hemorrhoids. All of them are best done with the patient in the prone, flexed position. The closed technique uses either a laser, scalpel, scissors, or cautery to elliptically incise the pedicles of the hemorrhoid cushion, including the external and internal components (Fig. 26-33). The incision avoids the underlying internal sphincter muscle as the vascular cushion is dissected from the surface of the muscle. The mucosal defect is then closed after the vascular pedicle is secured with the suture.

Urinary retention is the most common complication after excisional hemorrhoidectomy and has been reported in 10 to 50 percent of patients. Minimizing the amount of IV fluids given during an operation decreases the rate of this complication. Other complications include bleeding, infection, fecal impaction, and sphincter injury. Many of the complications can be prevented by proper use of analgesia to avoid pain and straining to urinate or defecate in the immediate postoperative period. Significant bleeding after hemorrhoidectomy must be managed by complete visualization of the operative site and cautery or suture ligation of the bleeding tissue. This requires adequate anesthesia, usually a spinal or epidural. Anal stenosis is a long-term complication following circumferential excision of the anoderm and dentate line during an improperly performed Whitehead procedure. Excision of the dentate line and anoderm causes circumferential scarring at the new mucocutaneous junction. The result is eversion of rectal mucosa to form ectropion and stenosis of the anal canal. This *Whitehead deformity* often requires an island advancement flap procedure in which normal skin is inserted into the anal canal to relieve the ectropion and the stenosis (see Fig. 26-25).

Acute thrombosis and incarceration of mixed hemorrhoids can be treated conservatively with bed rest and stool softeners or with excisional hemorrhoidectomy after injecting local anesthesia and reducing the edematous rectal tissue. Epinephrine and sometimes hyaluronidase may be added to the local anesthetic agent to decrease swelling and allow the tissue to reduce. A standard three-quadrant excisional hemorrhoidectomy is safe in this setting and should not cause anal stricture and sphincter injury if performed carefully.

Excision of Thrombosed External Hemorrhoid

A patient seen in the office or emergency department with an acutely thrombosed external hemorrhoid may obtain relief by excision of the acutely thrombosed tissue. This is performed with local anesthesia, and the wound is usually left open. Merely incising the hemorrhoid and expressing the clot carries a significant risk of recurrence of the thrombosis and bleeding. The patient who is seen more than 48 h after thrombosis of a hemorrhoid is best treated conservatively with warm soaks, an increase in fiber in the diet, and stool-softening agents.

Prevention

The majority of hemorrhoids occur because of a patient's poor bowel habits. It is important to stress the use of fiber and water in the diet, the avoidance of straining, and the cessation of sitting on the toilet for long periods of time. Symptomatic hemorrhoids can recur despite surgical treatment if these instructions are disregarded.

Anal Fissure/Anal Ulcer General Principles

The very common problem of anal fissure was first described in 1829 by Recamier, who recommended stretching the anal sphincter to treat the condition. An *anal fissure* is a split in the anoderm, usually in either the posterior or anterior midline just distal to the dentate line (Fig. 26-34A). Ninety percent of all fissures occur posteriorly and 10 percent anteriorly. Less than 1 percent of patients will be found to have a fissure in both the anterior and posterior positions. The very characteristic symptoms of the fissure or ulcer include tearing pain on defecation, blood on the toilet paper or stool, and, rarely, blood in the toilet bowl. Patients also may complain of a perineal or perirectal ache or spasm for several hours after defecation.

The cause of the anal fissure has been postulated to be a hard bowel movement or prolonged diarrhea with stretching of the anal canal to cause a split in the anoderm. Increased anal sphincter pressure has been documented on anal manometry in patients with an anal fissure. However, patients in the general population with increased anal pressure have not been found to be at high risk for developing an anal fissure. Fissures may result from ischemia, since anodermal blood flow in the posterior midline less than in other segments of the anal canal, and increased anal pressure correlates with reduced blood flow to this area. The pain and irritation of the fissure result in spasm of the underlying internal sphincter muscle, which then fails to relax during defecation, resulting in further tearing of the anoderm and deepening of the fissure to form an anal ulcer. The presence of this anal ulcer implies a chronic fissure and is accompanied by an external skin tag and a hypertrophied anal papilla in its most mature form (see Fig. 26-34 B). The white fibers of the internal sphincter are found at the base of the ulcer.

Crohn's disease or tuberculosis of the anal canal should be suspected when the fissure or ulcer is in an atypical position. Anal disease is the first manifestation in 10 percent of patients with Crohn's disease. Patients with anal tuberculosis will have a prior history of pulmonary tuberculosis or a concomitant infection. Anal cancer may present as a painless ulcer, but it is usually associated with a mass and persists despite treatment. It is important to biopsy all nonhealing ulcers for this reason. Occasionally, abscess/fistula disease, cytomegalovirus, herpes, or *Chlamydia* infection, or syphilis will mimic some of the features of fissure and ulcer disease.

Diagnosis and Management

The diagnosis is made by visual inspection to identify either the fissure or the ulcer complex at the dentate line. Digital examination can be very painful but, if possible, will reveal not only the ulcer but also the characteristic spasm and hypertrophy of the underlying internal sphincter muscle. Medical treatment of the fissure and ulcer relies primarily on stool softeners and the addition of bulk to the diet. Almost 90 percent of simple anal fissures will heal with simple medical therapy. Stool softeners and bulk, as well as warm sitz baths, will relieve symptoms and produce healing in 60 to 80 percent of patients with a second episode. Topical nitroglycerin 0.2% to 0.5% (a nitric oxide donor) has been used to relax the anal sphincter and has achieved healing of anal fissures in 80 percent of a small series of patients. Botulin toxin injection also has been used to effect transient paralysis of the anal sphincter and promote healing in some patients with anal fissure. Only after the fissure has been shown to be chronic is operation considered.

Surgical Management

The surgical methods of treating fissure and ulcer disease include stretching or dividing the internal sphincter. The four-finger stretch produces uncontrolled fracturing of the internal sphincter even though it may give initial relief of symptoms. Approximately 40 percent of patients treated this way develop recurrence, and a significant proportion are partially incontinent. The lateral internal anal sphincterotomy is considered the procedure of choice by most surgeons. A success rate of 90 to 95 percent is reported after lateral sphincterotomy for a chronic anal fissure or ulcer. Recurrence rates usually are below 10 percent, and minor incontinence, such as leakage of mucus and gas, is reported in less than 10 percent of the patients. The sphincterotomy is usually performed in the posterior lateral anus. The distal third of the internal sphincter is hypertrophied and presents a distinct, palpable band. If this hypertrophied band is not present, another cause of the ulcer should be considered. An open or closed sphincterotomy will give similar results, and each has its proponents (Fig. 26-35). The procedure is performed with the patient in the prone, flexed position, usually with local anesthesia; it is almost always performed on an outpatient basis.

INFECTIONS

Colitis

Pseudomembranous Colitis

Pseudomembranous colitis presents as an acute diarrheal syndrome in patients who are undergoing or have recently completed antibiotic therapy. Antibiotics alter the normal colonic bacterial population, permitting an overgrowth of the pathogen *Clostridium difficile*. This bacterium elaborates two separate exotoxins, a cytopathic toxin and an enteropathic toxin. The cytopathic toxin can be fortuitously neutralized by a cross-reaction with the antitoxin to *C. sordellii*, providing the basis for the tissue culture assay to detect the presence of *C. difficile*. If the cytopathic effects of *C. difficile* can be demonstrated by tissue culture and that effect is eliminated by culture with *C. sordellii* antitoxin, the diagnosis of *C. difficile* infection is presumed. A latex fixation test has been reported to successfully detect the cytopathic toxin. The assay should be performed within 24 h of collection of the stool. The toxin is moderately heat labile and deteriorates rapidly.

C. difficile colitis is usually preceded by the administration of antibiotics, and in as many as one-fourth of the patients the antibiotic has been discontinued before the development of diarrhea. Symptoms may appear as long as 6 weeks after antibiotic therapy. Virtually all antibiotics have been associated with the subsequent development of *C. difficile* colitis, but clindamycin, ampicillin, and the cephalosporins are most commonly implicated. There also is a contagious risk to this disease, with cross- infection occurring in hospitals and nursing homes. For this reason, isolation precautions should be used for patients with *C. difficile* colitis.

The clinical syndrome of this infection varies widely from mild, self-limited diarrhea to severe, invasive colitis associated with megacolon and perforation. When the disease is suspected, the implicated antibiotic should be stopped immediately. Proctoscopy or colonoscopy is a useful measure to detect pseudomembranes, and stool should be sent for toxin assay. Usually mild disease is self-limited. For more severe symptoms, including abdominal pain, fever, and leukocytosis, the patient should be admitted to the hospital for observation and intravenous hydration. Metronidazole is the antibiotic of choice and is usually prescribed in a dosage of 250 to 500 mg orally four times a day for 10 days. Vancomycin (125 mg orally four times a day for 10 days) is also effective, but it is much more expensive. In patients in whom oral therapy is not possible, intravenous metronidazole may be used. Intravenous vancomycin does not achieve therapeutic concentrations in the colon, however, and has no role in this setting. Cholestyramine has been recommended because it binds the toxin and renders it ineffective; however, the resin also may bind the therapeutic antibiotic, rendering it ineffective.

Most patients respond rapidly to the aforementioned treatment. Unfortunately, the recurrence rate is at least 20 percent, but these patients generally respond well to retreatment. Rarely, the disease becomes fulminant and fails to respond to appropriate antibiotic treatment; should the disease progress to a fulminant stage with toxic dilatation of the colon, abdominal colectomy with ileostomy may be necessary.

Amebic Colitis

Amebiasis is an infectious disease caused by the protozoan *Entamoeba histolytica* that infects primarily the colon but may involve other organs (usually the liver) secondarily. In the United States, amebiasis is usually thought to be a tropical disease, to be considered only if the patient has traveled abroad. However, it has been estimated that at least 5 percent of the untraveled population of the United States is infected with *E. histolytica*. The disease is transmitted by the fecal-oral route, which may include sexual contact, contamination of

the water supply, and fecal soilage in day-care centers or institutions for the retarded.

E. histolytica exists in the human colon in two forms: the trophozoite and the cyst. The latter is the infective form that is ingested, causing colonization. The majority of patients infected with the protozoa remain asymptomatic carriers. However, the cysts may transform to invasive *trophozoites* at any time, and the contributing factors leading to the emergence of virulence are not understood. The pathologic lesions of invasive amebiasis are small ulcers that may occur throughout the colon but most commonly affect the cecum. The ulcers may be covered with a yellow exudate. As the disease progresses, the ulcers may grow to more than 2 cm in diameter, with edges of undermined mucosa and necrotic tissue filling the ulcer base. The intervening mucosa may appear completely normal. In rare circumstances the ulcer may perforate, causing fulminant peritonitis. On histologic section, the amebae can be detected at the edge of the ulcer.

The clinical spectrum of amebiasis varies greatly from asymptomatic carriers to acute illness with bloody diarrhea resembling fulminant ulcerative colitis. On rare occasions the disease can cause toxic dilatation of the colon, and emergency colectomy is lifesaving. It is important to distinguish this disease from ulcerative colitis or Crohn's disease; the steroids often prescribed for the latter diseases are contraindicated in patients with amebiasis because immunologic suppression can lead to lethal dissemination of the disease. Microscopic examination of fresh stool specimens will reveal trophozoites in 90 percent of patients with symptomatic invasive amebiasis. Sigmoidoscopy is occasionally helpful if the ulcers are located in the distal bowel; however, these lesions more commonly occur in the cecum. Appropriate stool cultures should be taken to distinguish this disease from infectious bacterial colitis.

Metronidazole is the recommended treatment for invasive amebiasis. This medication, however, might be ineffective against organisms within the bowel lumen and in carriers. Iodoquinol (diiodohydroxyquin) is recommended as an accompanying medication to eliminate luminal organisms.

An uncommon complication of amebic infection that is of interest to surgeons is the *ameboma*, a mass of inflammatory tissue that may cause colonic narrowing and be confused with a carcinoma (Fig. 26-36). Colonoscopy may help in establishing the exact nature of the lesion. This lesion occurs most commonly in the cecum and is difficult to diagnose because the parasite is usually not present in the stools. The indirect hemagglutination test is usually positive in patients with an ameboma, and this test should be done if a lesion is suspicious or if there is a history of previous amebic infection. If the diagnosis is established, these patients should be treated with the antibiotics previously described. If the lesion does not resolve with antibiotic treatment, operation should be performed under cover of antiamebic therapy.

Actinomycosis

Actinomycosis is an uncommon bacterial infection caused by the gram- positive anaerobic or microaerophilic bacterium *Actinomyces israelii*. This organism formerly was misclassified as a fungus; but the species is closer to a bacterium than it is to a fungus. Clinical disease is characterized by chronic inflammatory induration and sinus formation. The organism is a normal resident of the mouth, and the cervicofacial area is the most frequently involved site, followed by thoracic and abdominal involvement.

Abdominal infection most frequently involves the cecal area, usually following appendectomy. An indurated, firm pericecal mass may develop, and, if untreated, the condition progresses to formation of indolent sinuses extending to the abdominal skin.

The diagnosis of actinomycosis should be suspected if an indolent mass or chronic sinus follows an appendectomy. The organisms make *sulfur granules*, which are actually tiny bacterial colonies consisting of a central core of intertwined branching filaments, with club-shaped bodies at the periphery, held together by a polysaccharide protein. These colonies create a sunburst pattern recognized microscopically, from which the organism derives its name. While sulfur granules are highly suggestive of *Actinomyces*, they are neither a consistent finding nor pathognomonic; *Nocardia*, *Streptomyces*, *Actinobacillus*, and *Staphylococcus* may form similar granules. Diagnosis can be confirmed by aerobic and anaerobic cultures. The microbiology laboratory should be alerted to the possible diagnosis of actinomycosis, since it often takes up to a week to isolate the organism in anaerobic culture.

Treatment consists of adequate surgical drainage and antibiotics. Penicillin and tetracycline are both effective; antifungal agents are not effective.

Neutropenic Enterocolitis

Neutropenic enterocolitis occurs most often in patients with acute myelogenous leukemia receiving high-dose cytosine arabinoside chemotherapy, but it may be seen in patients with other types of leukemia or in association with other types of chemotherapy. The syndrome is characterized by sepsis, abdominal pain and tenderness, and neutropenia. High fever, watery or bloody diarrhea, and abdominal distention are common. Tenderness is often localized to the right lower abdominal quadrant, and abdominal x-rays may show partial small bowel obstruction, thickened irregular mucosal folds, a dilated cecum, and air within the bowel wall. CT examination may reveal an edematous cecum with inflammation of the adjacent soft tissues and air in the cecal wall.

The pathogenesis of the syndrome and its proclivity to affect the cecum are not clear. It has been suggested that leukemic infiltrates in the colonic wall become necrotic after chemotherapy, resulting in ulceration of the mucosa and sepsis. Other theories include mucosal hemorrhage accompanying thrombocytopenia causing mucosal disruption and subsequent sepsis or ischemia caused by stasis of the enteric vessels resulting in mucosal necrosis and ulceration.

Treatment consists of complete bowel rest with total parenteral nutrition and intravenous broad-spectrum antibiotics. If the sepsis cannot be controlled, or if colonic perforation occurs, operative intervention will be necessary. In such circumstances, all necrotic bowel must be removed, and the viable limbs of the bowel must be exteriorized as an ileostomy and mucous fistula. It has been suggested that elective right hemicolectomy may be indicated to prevent recurrence in patients who have recovered from neutropenic colitis and require additional chemotherapy.

Cytomegalovirus Colitis

Cytomegalovirus (CMV) infects approximately 90 percent of patients with the acquired immune deficiency syndrome (AIDS) during the course of their illness. It appears to be more commonly associated with homosexuals infected with the human immunodeficiency virus (HIV I). The infection can involve numerous organs, including the lungs, eyes, brain, liver, adrenal glands, intestinal tract, and anus. CMV colitis may occur in up to 10 percent of patients with AIDS and is the most common reason for emergency abdominal operations in this group of patients.

The virus causes mucosal ulceration and submucosal hemorrhage, and symptoms include diarrhea, rectal bleeding, fever, and weight loss. Biopsy will reveal CMV inclusions and inflammation.

Dihydroxypropoxymethyl guanine (DHPG, or ganciclovir) is the drug of choice. If the disease progresses to toxic megacolon or perforation, emergency colectomy is mandatory. Emergency colectomy also might be required for massive colonic hemorrhage. Anastomosis is seldom advisable under such circumstances, and ileostomy is almost always necessary. The prognosis is very poor; fewer than half of patients are alive at 6 months.

Chagas' Disease

Chagas' disease is endemic in Central and South America and is a frequent cause of acquired megacolon in these regions. Infection with the protozoan *Trypanosoma cruzi* results in destruction of Auerbach's myenteric plexus followed by intramural fibrosis. Although various lengths of colon may be affected, the rectum is the site most frequently involved. Loss of normal propulsive ability causes a functional obstruction and results in a proximal megacolon.

Conservative therapy may be possible in patients with mild symptoms. Complications of severe constipation, fecal impaction, or volvulus require resection of the involved colon and rectum. Sigmoidectomy, left colectomy, and subtotal colectomy all have high recurrence rates because the dyskinetic rectum is preserved. Abdominoperineal endoanal pull-through resection with delayed colorectal anastomosis or the Duhamel procedure are the techniques of choice.

Infections of Anorectum

Anorectal Abscess and Fistula Cryptoglandular Abscess

There are several potential spaces around the anorectum that are ordinarily filled with fat and areolar tissue (Fig. 26-37).

Pathophysiology

The *perianal space* surrounds the anus and becomes continuous with the fat of the buttocks. The *intersphincteric space* separates the internal and external sphincters. It is continuous with the perianal space and extends cephalad into the rectal wall. Lateral and posterior to the anus is the pyramidal *ischiorectal space*. It is found below the levator ani and bound medially by the external sphincter, laterally by the ischium, and inferiorly by the transverse septum. This space contains the inferior rectal vessels and lymphatics. The two ischiorectal fossae connect posteriorly above the external sphincter forming the *deep postanal space* between the levator ani and the anococcygeal ligament. An abscess in one ischiorectal fossa can infect the other via the deep postanal space, creating a "horseshoe" abscess. The *supralevator spaces* lie above the levator ani on either side of the rectum. These spaces communicate with each other posteriorly.

The majority of anorectal suppurative disease results from infections of the anal glands (cryptoglandular) (Fig. 26-38) and is usually due to a mixture of fecal and cutaneous flora. The glands are found in the intersphincteric plane, traverse the internal sphincter, and empty into the anal crypts in the anal canal at the level of the dentate line. Not all crypts contain glandular openings. These common infections should not be confused with anorectal manifestations of Crohn's disease, pilonidal disease, hidradenitis suppurativa, or the rarer tuberculosis or actinomycosis. The *acute* phase of cryptoglandular disease results in formation of an abscess. If the abscess drains and an infected tract persists, the process enters the *chronic* phase of fistula-in-ano.

Diagnosis

Severe anal pain is the most frequent presenting complaint. Walking, coughing, or straining can aggravate the pain. A palpable mass is often detected by inspection of the perianal area or by digital examination of the rectum. Occasionally patients present with fever, urinary retention, or life- threatening sepsis.

Infection of an anal gland initially results in the formation of an abscess in the intersphincteric plane (Fig. 26-39). As the abscess enlarges, it spreads in one of several directions. A *perianal abscess* is the most common manifestation. It forms when pus spreads downward between the two sphincters and presents as a tender, erythematous bulge at the anal verge. If the pus penetrates the external sphincter below the level of the puborectalis and expands into the ischiorectal fossa, an *ischiorectal abscess* is formed. A *supralevator abscess* is formed when an intersphincteric or ischiorectal abscess expands upward. Pus is found above the levators, lateral to the rectum, and below the peritoneal reflection.

These abscesses can become quite large, are seldom visible, and are palpated by digital rectal examination as a bulge above the puborectalis. These abscesses are uncommon, accounting for less than 5 percent of patients with anorectal abscesses.

Treatment

Perirectal abscesses should be treated by drainage as soon as a diagnosis is established. It is not necessary to delay operation to allow erythema or purulence to develop. Antibiotics are not indicated unless there is extensive cellulitis or the patient is immunocompromised, has diabetes, or has valvular heart disease. In this group of patients, expeditious treatment is critical, because the infection can spread rapidly and become life-threatening. Usually the abscess can be localized by digital examination. If an indurated area is all that is found, examination under anesthesia and needle localization are indicated. Once adequate anesthesia has been obtained, proctosigmoidoscopy should be performed to rule out other anorectal disease.

An *intersphincteric abscess* is treated by an internal sphincterotomy. The sphincterotomy drains the abscess and destroys the infected crypt. *Perianal* and *ischiorectal abscesses* are treated by drainage through an incision in the perianal skin overlying the abscess. The abscess must be drained completely, either by excising an adequate amount of overlying skin or by inserting a mushroom catheter into the abscess cavity. After the acute phase has resolved, the mushroom catheter may be injected with dilute methylene blue or povidone-iodine. If there is no evidence of a fistula (the presence of an internal opening), the catheter is removed. If a fistula is present, it should be treated as described below. *Supralelevator abscesses* should be drained transanally by incising the rectal wall overlying the abscess. A supralelevator abscess should not be drained externally or through the ischiorectal fossa; this may cause a supralelevator fistula that will seldom heal without a colostomy for diversion of feces (Fig. 26-40).

Drainage alone results in cure for 50 percent of patients. The other 50 percent develop an anal fistula with persistent symptoms of local infection. If only skin pathogens are cultured from an anorectal abscess, there is a decreased incidence of fistula formation. This is probably because such an abscess originates from the skin instead of from an anal crypt. The majority of the recurrences are in patients with ischiorectal abscesses. Although an argument can be made to do a primary fistulotomy at the time of incision and drainage, it is not always possible to find the internal opening in the setting of acute inflammation. Unless the internal opening can be clearly identified, it is probably preferable to treat ischiorectal abscesses by incision and drainage, because a complicated fistulotomy may result in injury to the anal sphincter.

The immunocompromised patient may present with anal pain but no evidence of a fluctuant mass because of the paucity of white blood cells. These patients should be examined under anesthesia. In such circumstances, a painful indurated area should be drained even if no purulence is found. The tissue should be biopsied (to exclude a leukemic infiltrate) and cultured (to aid in selection of antibiotics).

Necrotizing Anorectal Infections

Pathophysiology

Necrotizing anorectal infections (nonclostridial and streptococcal cellulitis, clostridial gangrene, and Fournier's gangrene) are perianal and perineal infections caused by either a monobacterial or mixed synergistic flora. The source of infection may be anorectal or urogenital.

Diagnosis

Examination of the perineum may reveal only limited necrosis of skin overlying crepitant tissue, but extensive tissue necrosis may be rapidly progressing in the deeper tissue. Many patients present with systemic toxicity. Most patients have an underlying chronic systemic disease. Cultures may reveal the causative organism, but surgical therapy should not be delayed while waiting for results.

Treatment

Early recognition and aggressive surgical debridement are important. Excision of all nonviable perineal and perianal tissue is required. Antibiotics are an adjunct to adequate surgical therapy. Orchiectomy is seldom necessary because the blood supply to the testicles is different from that to the scrotum and penile skin. Colostomy may be required to decrease fecal contamination or if the anal sphincters are destroyed by the infective process. Hyperbaric oxygen has been used for clostridial infections but should not interfere with or delay adequate debridement of devitalized tissue. Early diagnosis and treatment are necessary for a successful outcome; the mortality is approximately 50 percent.

Fistula-in-Ano

Pathophysiology

Acute infection of the anal crypt leads to an anorectal abscess; an anal fistula represents the chronic form of this infection (Fig. 26-41). The fistula consists of the infected crypt and a tract extending to the site of drainage. It is therefore important to drain an abscess as close to the anus as possible to avoid the formation of a long fistulous tract. Fistulas also may be secondary to infections due to trauma, fissures, tuberculosis, Crohn's disease, carcinoma, radiation, actinomycosis, and *Chlamydia* infection.

Diagnosis

Patients present with persistent purulent drainage from the external and internal openings of the fistula. An indurated tract is often palpable. The internal opening is usually detectable by anoscope examination. Goodsall's rule can be used as a guide in determining the location of the internal opening (Fig. 26-42). It states that if the anus is bisected by a transverse line, external openings anterior to the line will connect to an internal opening by a short, direct fistulous tract. External openings posterior to this line join tracts that curve toward an internal opening in the posterior midline. An exception to this rule occurs if an anterior external opening is greater than 3 cm from the anal margin. These fistulas are more likely to have an internal opening in the posterior midline.

Treatment

The mainstay of treatment for anal fistula is eradication of sepsis with preservation of anorectal function. A simple fistula is treated by *fistulotomy* (opening the fistulous tract), curettage or cautery of the tract, and healing by secondary intention. *Fistulectomy* (complete excision of the fistulous tract) is not indicated because the magnitude of tissue loss associated with the procedure increases the risk of compromise to sphincter function.

An anal fistula that results from a drained ischiorectal abscess is usually a *transsphincteric fistula*, crossing both the internal and external sphincters. Incising the tract results in an internal sphincterotomy as well as division of a portion of the external sphincter. Continence usually will be maintained if the puborectalis muscle is not divided.

Horseshoe fistulas usually have an internal opening in the posterior midline of the anus and may extend anteriorly and laterally to one or both ischiorectal spaces by way of the deep postanal space (see Fig. 26-37). Incision and drainage of horseshoe abscesses are associated with a high recurrence rate. Adequate treatment of a horseshoe abscess requires incision and drainage of the postanal space and counterdrainage of the lateral ischiorectal spaces through separate incisions (Fig. 26-43). Counterdrainage is maintained until the incision over the deep postanal space is healed. This method is preferable to a fisulotomy, which would require a long skin incision to unroof the entire tract. High, complicated fistulas should not be treated by fistulotomy because of the risk of significant sphincter damage. A *seton* (see Fig. 26-57) used in this situation acts as a drain for the primary tract, but use of the seton to transect the muscle carries a significant risk of incontinence. Primary drainage followed by an endorectal sliding flap has been used to treat complicated anterior fistulas without impairment of the sphincter. This method, a modification of the technique used to treat rectovaginal fistulas (Fig. 26-44), removes the involved crypt and closes the internal opening with a sliding endorectal flap. The fistulous tract is counterdrained, and the drain is removed after the endorectal flap has healed.

Untreated fistulas may cause chronic intermittent infections, episodes of sepsis, rare cases of carcinoma, and, occasionally, death. The use of a seton is associated with an incontinence rate of 17 percent, which is increased to 39 to 63 percent after the tract is laid open at a second stage. Therefore, our technique has been to establish drainage of the primary abscess to control the sepsis. Anorectal function is preserved, and an endorectal sliding flap is used to close the internal opening. The technique is the same as will be described for rectovaginal fistula, except for the perineal counterdrain.

Rectovaginal Fistula Pathophysiology

A *rectovaginal fistula* is a communication between the rectum, proximal to the dentate line, and the vagina. These fistulas are classified by their location. The rectal opening in a *low rectovaginal fistula* is close to the dentate line with the vaginal opening just inside the fourchette. The vaginal opening in a *high rectovaginal fistula* is near the cervix, and the opening of a *midrectovaginal fistula* is between the high and low locations.

The cause varies with the location of the fistula. Low rectovaginal fistulas are commonly caused by obstetric injuries (especially midline episiotomies that become infected), foreign body penetration of the rectum, or Crohn's disease. Midrectovaginal fistulas may result from extension of an undrained ischiorectal abscess. Crohn's disease, surgical excision or fulguration of an anterior rectal tumor, radiation injury, or extensive childbirth trauma. High rectovaginal fistulas are caused by Crohn's disease, radiation injury, operative injury, and diverticulitis. These septic processes are more prone to occur through the vaginal cuff after hysterectomy. Diverticulitis may cause a colovaginal fistula. Carcinoma of the rectum, cervix, or vagina also may result in a rectovaginal fistula at any level.

Diagnosis

Patients describe the very characteristic symptoms of passing gas and stool through the vagina and report varying degrees of continence. Diagnosis is usually evident on anoscopic or vaginal examination. A barium enema or vaginogram may reveal the fistula. Sometimes the history is characteristic, but the site cannot be identified. In such circumstances, administration of a dilute methylene blue enema may stain a tampon placed in the vagina, confirming the presence of a fistula.

Treatment

The treatment of a rectovaginal fistula depends on its cause, location, and the condition of the adjacent tissue. If the fistula is caused by a cryptoglandular abscess, appropriate drainage of the infection may result in spontaneous closure of the fistula. Fistulas resulting from obstetric injury may heal spontaneously, and it is wise to wait at least 3 months before attempting surgical repair. This also allows the acute inflammatory reaction to resolve.

Low and midrectovaginal fistulas are best treated by an endorectal advancement flap (see Fig. 26-44). If a sphincter injury is also present, it may be repaired simultaneously. A protecting colostomy is not necessary, but full bowel preparation and systemic antibiotics are important, because infection is the major source of failure of the repair. Treatment of a high rectovaginal fistula requires a transabdominal approach, allowing resection of the diseased tissue that caused the fistula.

Fistulas caused by Crohn's disease, radiation injury, or cancer almost never heal spontaneously. Intensive medical treatment and conservative local drainage procedures establish adequate drainage of abscesses and preserve the anal sphincter in patients with perianal Crohn's disease. Local sliding endorectal flaps for Crohn's fistulas in which the sepsis has been controlled and the disease is in remission have been used with good results, as will be described in the section on Crohn's disease. Proctectomy may be required for patients

with severe perineal Crohn's disease unresponsive to local anorectal surgical procedures or for patients with extensive destruction of the rectum.

Rectovaginal fistulas due to pelvic radiation are seldom amenable to transanal repair because the tissue is permanently damaged. Tissue repairs constructed with irradiated bowel are doomed to fail. Biopsy of the fistulous tract is mandatory in patients with a history of cancer or radiation therapy. It is sometimes very difficult to clinically differentiate radiation injury from recurrent cancer.

Pilonidal Disease Pathophysiology

Pilonidal disease consists of a hair-containing sinus or abscess that involves the skin and subcutaneous tissues in the postsacral intergluteal region. The cause of the disease is not clear; it may be a congenital or an acquired condition. Most investigators believe that the condition is due to ingrown hair. Pilonidal disease can occur in any age group but is most prevalent from adolescence to the third decade, with a male predominance. The sinus is superficial to the periosteum of the sacrum and may be single or have multiple extensions. An external orifice is not always visible but is usually in the midline approximately 5 cm from the anus. The soft tissues surrounding the sinus exhibit acute and chronic inflammation. Patients present with complaints of pain, swelling, and purulent drainage when these sinuses become acutely infected. Once an acute episode has resolved, the patient rarely remains asymptomatic; recurrent infections are common.

Diagnosis

Pilonidal disease is confirmed as the cause of an abscess when the typical midline pits are identified in association with the inflammatory process. Adequate anal examination should rule out an anal fistula as the source of the abscess.

Treatment

There are numerous treatments reported in the literature, all with imperfect results. Incision, drainage, and curettage with secondary healing constitute the simplest method of treatment and are used for acute abscesses and chronic sinus tracts. This method often can be performed with local anesthesia on an outpatient basis. All hair must be removed, and the wound must be kept clean and shaved until complete healing occurs. Complete excision and secondary healing leave a large open wound and are associated with a prolonged healing time. Excision and primary closure are often successful for small sinuses but are complicated in extensive disease. Excision of the sinus followed by a Z-plasty to obliterate the intergluteal cleft has been used in some centers with low recurrence rates. The Z-plasty also serves to eliminate complicated recurrent pilonidal disease.

Hidradenitis Suppurativa

Hidradenitis suppurativa is an infection of the cutaneous apocrine sweat glands. Infected glands rupture and form subcutaneous sinus tracts. The infection spreads by direct extension and can spread to the perineum, labia, and scrotum. The perianal skin can become a network of superficial abscesses and sinuses. This infection may mimic complicated anal fistula disease but stops at the anal verge because there are no apocrine glands in the anal canal.

The diagnosis is made by clinical appearance. The glands are destroyed by the inflammation resulting from their rupture; thus there are no specific histologic identifying features. Abscesses are initially treated by wide incision, including resection of the severely indurated, infected skin that overlies the infected tracts. A colostomy is almost never indicated. Recurrence is frequent, and any recurrent infections are best incised early in the process.

Pruritus Ani Diagnosis and Pathophysiology

Pruritus ani (severe perianal itching) is a symptom complex that may be caused by a number of factors but often must be classified as idiopathic because a specific etiologic agent cannot be discovered.

Patients complain of perianal itching that is often worse at night. Many have tried numerous over-the-counter medications to relieve their discomfort. A special problem occurs when products containing local anesthetic agents cause contact dermatitis and magnification of the original symptoms. A careful history and physical examination should identify any obvious causes.

The appearance of the perianal skin varies. The skin may appear normal, or it may be erythematous and weeping serum from numerous excoriations. Chronic pruritus ani is characterized by thickened, leathery, white skin.

Pinworms (*Enterobius vermicularis*) are the most common cause of perianal itching in children and may be transmitted to other family members. The worms come out of the anus at night and deposit eggs on the perianal skin, causing intense nocturnal pruritus. The diagnosis is confirmed by swabbing the perianal skin with cellophane tape and examining it under the microscope to identify the *Enterobius* eggs.

Common anal lesions (prolapsing hemorrhoids, villous tumors of the rectum, and rectal prolapse) often cause excessive moisture in the perianal region. This moisture causes irritation and itching of the perianal skin.

Dietary habits, cleansing habits, or a combination of the two are common causes of pruritus. Overzealous, as well as careless, cleaning of the perianal skin can lead to itching. Frequent application of topical ointments keeps the area moist, causing further irritation.

Treatment

The treatment of pruritus ani is directed at the cause. Hemorrhoids, villous adenomas, and rectal prolapse should be specifically treated as discussed in this chapter. Piperazine citrate is the curative treatment for pinworms. Patients must be instructed that frequent washing and topical ointments should be avoided. The perianal area should be kept as dry as possible, and scratching should be avoided. Short term course of topical steroids is likely to provide symptomatic relief. Coffee, tea, cola, beer, chocolate, and tomatoes all can induce perianal itching, and abstaining from these foods, as well as use of proper anal hygiene, for about 2 weeks should alleviate most of the symptoms. It is, however, very common for the symptoms to recur and require reinstitution of the treatment regimen.

Proctitis

Proctitis and *anusitis* are nonspecific entities characterized by inflammation of various degrees within the anus or rectum. The differential diagnosis can be complex and includes infection, inflammatory bowel disease, and various forms of trauma.

Inflammatory Proctitis

Ulcerative colitis always involves the rectum as well as variable amounts of colon. There is a specific, mild form that involves only the rectum and is referred to as *ulcerative proctitis*. Ninety percent of these patients respond to medical management and never go on to have involvement of the colon. Patients with Crohn's disease also may have isolated rectal involvement, making the differentiation between the two difficult at times.

Diagnosis

The diagnosis may be suggested by a good history. Rectal bleeding or discharge, diarrhea, and tenesmus are common presenting symptoms. Endoscopy reveals an inflamed and friable rectal mucosa and biopsy will verify the diagnosis. The remainder of the colonic mucosa may have a normal appearance.

Treatment

The initial treatment consists of steroid retention enemas given over a course of 2 weeks. A short course of oral steroids may be considered in patients who do not respond to the enemas, with the dose being tapered over 4 to 6 weeks. Sulfasalazine may be given orally or its active ingredient, 5-aminosalicylic acid (5-ASA), may be administered as retention enemas or suppositories. Dietary measures to decrease the frequency of bowel movements, such as avoidance of milk, fruit, and fiber, and the addition of bulk-forming agents such as psyllium products may be beneficial, as well as antimotility agents such as loperamide.

Radiation Proctitis

Another cause of proctitis is radiation injury. An accurate history usually reveals the diagnosis. Early symptoms are diarrhea, rectal bleeding or discharge, tenesmus, rectal pain, and incontinence. Late symptoms appear several months to years after irradiation. Bleeding is the most common late symptom, but rectal strictures and fistulas to the bladder or vagina can develop. Examination reveals telangiectasia and friable mucosa.

Treatment

Bulk-forming agents, antispasmodics, and antidiarrheal medications are used initially. Steroid or 5-ASA retention enemas or suppositories may be helpful. Laser coagulation of bleeding telangiectasias may be required to control bleeding. Topical application of 10% formalin may be helpful in patients with persistent bleeding. Rarely, the rectum will be so badly injured that intestinal diversion is necessary.

Sexually Transmitted Disease

The differential diagnosis of proctitis and especially anusitis must include the spectrum of sexually transmitted diseases are common in individuals who practice anal intercourse.

Gonococcal Proctitis

Anorectal gonorrhea is caused by *Neisseria gonorrhoeae*, an intracellular gram-negative diplococcus. It is associated with anal intercourse, and vaginal discharge can transfer the disease to the rectum in women. Initially, patients are asymptomatic or present with rectal bleeding or discharge. Perianal excoriation, painful defecation, and fistulas also may occur. Proctoscopy reveals edematous and friable mucosa and a purulent exudate. The diagnosis should be established by cultures of anus, vagina, urethra, and pharynx, since these areas are also frequently infected.

Intramuscular procaine penicillin G with oral probenecid is the treatment of choice. Penicillinase-producing *N. gonorrhoeae* requires treatment with spectinomycin or ceftriaxone and tetracycline. Follow-up examination and cultures are necessary to avoid inadequate treatment. Sexual partners also should be evaluated.

Anorectal Syphilis

There is a rising incidence of syphilis in the Western Hemisphere. The causative organism is the spirochete *Treponema pallidum*. After an incubation period of up to 8 weeks, the primary lesion (chancre) is found at the site of inoculation. Syphilitic chancres are usually asymptomatic and can easily go unrecognized. Ulcers, proctitis, pseudotumors, and

condylomata lata also may be found. Condylomata lata are the lesions of secondary syphilis and are highly contagious. They are hypertrophic papules found in areas of warmth, moisture, and friction.

Dark-field examination of exudate may reveal the spirochetes. The rapid plasma reagin (RPR) test and Venereal Disease Research Laboratory (VDRL) slide test are nonspecific screening tests that become positive 1 to 3 weeks after the appearance of the primary chancre. These tests are positive in 75 percent of patients with primary syphilis. Both the RPR and VDRL tests remain positive for long periods even after treatment. The fluorescent treponemal antibody-absorption test (FTA-ABS) is more specific and sensitive but is also more expensive. Syphilis is treated with benzathine penicillin G or erythromycin.

Chlamydia trachomatis

Chlamydia trachomatis is an obligate intracellular parasite that causes chlamydial proctitis and lymphogranuloma venereum (LGV). Chlamydial proctitis is associated with anal intercourse or may spread via the lymphatics in the rectovaginal septum in women. The infection may be asymptomatic, or the patient may complain of anal pain, pruritus, rectal bleeding, or rectal discharge.

The primary lesion of LGV is a small shallow ulcer that often is not noticeable. Inguinal adenopathy develops and can form large matted nodes. If left untreated, the patient could develop hemorrhagic proctitis and late rectal strictures. The LGV complement fixation test or the microimmunofluorescent antibody titer confirm the diagnosis. Tissue cultures can be used to isolate the organism.

Tetracycline (or doxycycline) is the treatment of choice, but erythromycin may be used as an alternative. Antibiotics should be continued for 14 to 21 days. Rectal strictures may be dilated early in the disease. Complete obstruction is rare and requires colostomy.

Chancroid

Chancroid, caused by *Haemophilus ducreyi*, is a disease manifested by soft, multiple lesions that are very painful and bleed easily. Multiple lesions may develop as a result of autoinoculation and are often seen in the perineum and around the anus. Inguinal adenopathy is a common finding. Lymph nodes become fluctuant and rupture through the skin, causing draining sinuses. Treatment is with erythromycin or trimethoprim- sulfamethoxazole.

Herpes Proctitis

Herpes proctitis is caused primarily by type II herpes simplex virus, although the type I virus can cause the same symptoms. The virus is transmitted by direct mucosal contact. Patients present with tenesmus, anal pain, rectal bleeding, discharge, and perianal ulcerations. The pain may be so intense that the patient is reluctant to have a bowel movement. Fever and malaise are frequently present.

The anorectal lesions appear 3 to 7 days after sexual contact. The primary lesions are clusters of vesicles that rupture, coalesce, and form painful ulcers. The ulcers become infected secondarily. These lesions tend to recur, and the recurrent form is usually less painful than the original. The disease is contagious while the lesions are present. Diagnosis is made by viral cultures of vesicular fluid or biopsies.

There is no cure; treatment is symptomatic. Oral acyclovir reduces the frequency and duration of recurrent episodes and reduces virus shedding, but acyclovir does not prevent recurrences and does not alter the long-term natural history of the disease.

Condylomata Acuminata

Condylomata acuminata are warts that occur in the anorectal and urogenital regions, caused by infection with the human papillomavirus. In developed countries, genital warts are now the most common viral sexually transmitted disease, estimated to be three times more common than genital herpes. Anal intercourse is the most common means of transmission of anorectal warts, and there is a high association with other sexually transmitted diseases. Although condylomata are found more commonly in homosexual men, heterosexual men and women can be infected. This virus is autoinoculable and has an incubation period that can be as long as 1 year; both factors account for frequent treatment failures.

Diagnosis

Most patients present after noticing the visible perianal growths. Pruritus, anal discharge, bleeding, and pain are occasionally present. The diagnosis is usually obvious because of the characteristic appearance of the warts. Anoscopy and proctosigmoidoscopy should be performed on all patients because condylomata may be found within the anal canal. Tests for other venereal diseases, e.g., HIV and hepatitis B are indicated. Condylomata lata associated with secondary syphilis are usually smoother, flatter, and lighter in color and can occur simultaneously with condylomata acuminata.

Treatment

There are numerous treatments for condylomata acuminata. Evaluation of treatment efficacy is complicated by occasional spontaneous regression of lesions, reinfection, and lack of prospective, randomized, controlled trials. Bichloroacetic acid and 25% podophyllin application are the two most common treatments performed in the office setting. These treatments are satisfactory only if the warts are small and few in number. Weekly evaluation and treatment are necessary until all the warts are gone.

Excision and electrocoagulation under local anesthesia are effective treatment for the patient with a large number of warts. The warts are destroyed using the cautery; care is taken to preserve the adjacent skin and anoderm. The smoke generated from the procedure contains viable organisms and should be thoroughly evacuated. Carbon dioxide lasers have been used to eradicate condylomata but have been shown to be associated with a slightly higher recurrence rate. Small lesions can be treated using local anesthesia; larger lesions require regional or general anesthesia. Immunotherapy with an autogenous vaccine and intralesional injection of interferon have had variable results. All treatments for condylomata have been associated with recurrence rates up to 70 percent.

Malignancy

There appears to be an association between the human papillomaviruses 16, 18, and 31 and carcinomas of the anal and genital tracts. Malignant degeneration to squamous carcinoma has been reported with increasing frequency in anal condylomata. Immunosuppression, most commonly encountered today with HIV disease, appears to confer additional risk. Resected condylomata specimens should be examined microscopically, and recurrent or persistent condylomata should be submitted for histologic examination. The incidence and rate of malignant transformation are unknown. Radical surgical removal of the rectum has been the usual treatment for these cancers but is associated with a high recurrence rate. Combined chemoradiation has been reported to shrink large lesions, making them surgically curable. The choices of treatment for anal cancer will be discussed later.

Giant condylomata acuminata (Buschke-Lowenstein tumors) involving the perianal region may appear to be histologically benign but exhibit malignant behavior. They are thought to represent a low-grade, well-differentiated variant of conventional squamous cell carcinoma. These tumors do not respond to conservative therapy, and radical surgical excision offers the only hope for cure or adequate palliation.

Acquired Immune Deficiency Syndrome (AIDS)

Acquired immune deficiency syndrome (AIDS) has been associated with numerous gastrointestinal complications. Tumors such as Kaposi's sarcoma and non-Hodgkin's lymphoma, as well as infections with fungi, viruses, atypical bacteria, and protozoa, can involve the entire gastrointestinal tract in immunosuppressed patients. Heterosexuals with AIDS get fewer anorectal infections and malignancies than homosexual men with AIDS.

Kaposi's sarcoma is the most common malignant tumor in AIDS patients. The lesions may appear as small telangiectasias or larger purple nodules; their color is due to extravasation of blood. The skin is the most frequently affected site, but the gastrointestinal tract may be involved as well. These lesions are often multiple and may occur anywhere along the gastrointestinal tract. Usually patients are asymptomatic, but the lesions can cause hemorrhage or obstruction. Diagnosis is made by the endoscopic appearance of the lesions, and biopsies for histologic confirmation should be taken. These lesions are submucosal, and superficial biopsies are often nondiagnostic; deeper biopsies are necessary, and although the lesions have a vascular appearance, they rarely bleed after biopsy. Therapy is palliative and may require chemotherapy with an agent such as vincristine or vinblastine. Surgery is seldom indicated. If surgery is required for complications such as persistent hemorrhage, obstruction, or intussusception, only the involved section of bowel should be resected.

Lymphoma is found with increased frequency in patients with immune deficiencies, particularly in patients with AIDS. Most are non-Hodgkin's B- cell lymphomas, which are highly aggressive, and multiple organ involvement is common. Rectal involvement is rare but has been reported in homosexual men and should be considered in the differential diagnosis of anorectal masses in this population. Patients present with fever and perirectal masses. The lesions should be biopsied for diagnosis. Chemoradiation is the treatment of choice, but the response rate in patients with AIDS is poor. Surgical resection or bypass may be necessary for obstructing or perforating gastrointestinal lymphomas.

CMV is ubiquitous in the homosexual AIDS population. CMV ileocolitis is the most common intestinal infection in patients with AIDS. Patients frequently present with diarrhea, melena or hematochezia, fever, and weight loss. Examination reveals diffuse ulceration that can occur anywhere along the gastrointestinal tract. Biopsies of the ulcers show CMV inclusion bodies. The diagnosis is confirmed by viral culture of biopsy specimens or antigen assay of washings taken from ulcers. Resection may be required for the complications of bleeding or perforation. The mortality is high.

Infections with *Campylobacter*, *Shigella*, *Mycobacterium avium- intracellulare*, *Cryptosporidium*, *Isospora belli*, *Entamoeba histolytica*, and *Giardia lamblia* may cause enterocolitis or intractable diarrhea in patients with AIDS. Many patients present with persistent diarrhea, lower abdominal cramping, and, occasionally, bloody stools. Anoscopic and sigmoidoscopic findings may range from normal-appearing mucosa to friable ulcerations. Diagnosis is based on special stool cultures, examination of fresh stool specimen for ova and parasites, and biopsies of ulcers sent for histologic examination. Treatment is directed at a causative agent, and surgery is rarely necessary.

Sexually transmitted anorectal diseases are common in AIDS patients, including gonorrhea, syphilis, chlamydia, herpes simplex, and papillomavirus infection. Anorectal disease in AIDS patients has been approached with nihilism in the past because of reports of high morbidity and mortality postoperatively and because of short overall survival from AIDS. However, now that new drug treatments are increasing the life expectancy in AIDS patients, it becomes more important to diagnose and aggressively treat anorectal disease to avoid disabling complications.

INFLAMMATORY BOWEL DISEASE

General Considerations

The term *inflammatory bowel disease* encompasses two major entities: *ulcerative colitis* and *Crohn's disease*. There are several synonyms for each of these diseases. Ulcerative colitis is also referred to as *mucosal ulcerative colitis* and *idiopathic ulcerative colitis*. Crohn's disease is also called *segmental colitis*, *regional enteritis*, *regional ileitis*, *terminal ileitis*, *granulomatous colitis*, or *transmural colitis*.

Ulcerative colitis is an inflammatory process involving the colonic mucosa, characterized by alterations in bowel function and symptoms of intestinal inflammation. The most frequent sign of ulcerative colitis is hematochezia, the passage of red blood from the rectum. The presence of abdominal pain, fever, and weight loss depends on the severity of the inflammation (Table [26-2](#)). Ulcerative colitis has been recognized since the nineteenth century, when it was reported in several London hospitals.

Crohn's disease was first described as "chronic interstitial enteritis" by the Scottish surgeon Dalziel in 1913, but the description of the clinical entity "regional ileitis" in 1932 by Crohn, Ginsberg, and Oppenheimer from Mt. Sinai Hospital in New York led to widespread recognition of the disease involving the ileum. Subsequently, Lockhart-Mummery and Morson reported the colonic involvement by transmural inflammation, distinguishing the entity from ulcerative colitis. It is now recognized that Crohn's disease can occur anywhere in the gastrointestinal tract, from mouth to anus, and that the extraintestinal manifestations of the disease may precede those of the intestine.

The annual incidence of both diseases in western Europe and North America increased up to 1970 and has since stabilized or decreased. While men more frequently have ulcerative colitis and women more frequently have Crohn's disease, genetic and environmental factors seem to be important etiologic considerations. Other factors under investigation include smoking, oral contraceptives, alcohol, refined sugar, infectious diseases, and Jewish heritage. A defect in immune regulation in the gastrointestinal tract, in either an initiating or a perpetuating role, is also being investigated. Some observations implicate gut-associated lymphoid tissue in the pathobiology of both ulcerative colitis and Crohn's disease, for example, the granuloma in Crohn's disease (indicative of T-cell-mediated immunity), the lymphoid infiltrate resembling an Arthus reaction in ulcerative colitis, the presence of extraintestinal manifestations of *inflammatory bowel disease (IBD)* in other immunologically mediated diseases, and IBD patients displaying hypergammaglobulinemia and autoantibody production as seen in an immunologically activated state.

Recent studies focus on a fundamental defect of mucosal immune regulation leading to immunologic hyperactivation rather than on a direct autoimmune assault on a specific target cell, the focus of earlier studies. This fundamental defect, leading to uncontrolled immune reaction to environmental agents and antigens, can cause an unregulated immune response, including lymphocyte proliferation, cytokine release, secondary recruitment of auxiliary effector cells, excessive neutrophil products (especially leukotriene B₄ and reactive oxygen metabolites), and ultimately, tissue injury (Fig. [26-45](#)). The actual tissue damage probably reflects a nonspecific "innocent bystander" injury.

Crohn's disease and ulcerative colitis appear closely related; when the disease exists in the colon, 15 percent of cases cannot be clearly differentiated (Table [26-3](#)). Distinguishing Crohn's colitis from ulcerative colitis carries therapeutic significance: There is a significant recurrence of disease involving the small intestine in patients with Crohn's disease treated by colectomy, whereas such recurrence is rare in patients with ulcerative colitis. The two entities are similar because they are both characterized by exacerbations and remissions of intestinal inflammation. Each also has a chronic phase. Ulcerative colitis is a disease of the mucosa that is limited to the colon and rectum. Crohn's disease, on the other hand, can involve any part of the gastrointestinal tract; this chapter will deal mainly with its involvement of the large intestine.

The symptoms also overlap and generally consist of those which would be caused by inflammation of the intestine, the specifics varying with the depth of the inflammation and the location along the intestinal tract (see Table [26-2](#)). Many diagnostic modalities may be used to differentiate and define the extent of the disease as well as the degree of inflammation. These include contrast studies of the intestine, endoscopic evaluation, CT, magnetic resonance imaging (MRI), ultrasonography, and special nuclear scans (Table [26-4](#)).

The tissue changes of ulcerative colitis are depicted in Fig. [26-46](#), which shows the progression from superficial, homogeneous ulceration and purulent discharge, to the intense inflammatory reaction resulting in edema and pseudopolyp formation, to the chronic changes of a flattened, potentially dysplastic lining of the large intestine. The disease generally involves only the mucosal layer of the large intestine. Crohn's disease, as depicted in Fig. [26-47](#), shows the early phase of the disease to involve only isolated areas of the mucosa in the form of aphthoid ulcers. As the activity of the disease progresses, deep ulceration and inflammation extend through the full thickness of the intestine, with thickening of the intestinal wall. It is this full-thickness destruction of tissue that results in the eventual perforation of the intestine, leading to many of the complications of Crohn's disease. In the chronic phase of Crohn's disease, the intestinal wall is thickened, fibrotic, and nondistensible, causing symptoms of intestinal obstruction. In all phases of the disease there may be areas of normal mucosa that are spared the inflammatory process; in contrast, ulcerative colitis is a more homogeneous inflammatory process.

Both ulcerative colitis and Crohn's disease have systemic manifestations contributing to the hypothesis that there is some type of autoimmune phenomenon involved in the causation of IBD (Table [26-5](#)).

The gross as well as the radiologic appearance of the intestine in the two diseases is depicted in Fig. [26-48](#). Ulcerative colitis is characterized by an intestinal wall of normal caliber with severe inflammatory changes of the mucosa and a normal-appearing mesentery. In Crohn's disease, the intestine is markedly thickened, as is the adjacent mesentery, with enlargement of the mesenteric lymph nodes. The serosal surface of the intestine is inflamed, and the fat takes on a strange configuration of enveloping the intestine ("fat wrapping"). Ulcerative colitis appears in contrast radiography as homogeneous ulceration with mucosal edema, on the other hand, chronic Crohn's disease is characterized by segmental narrowing with thickening of the loops of intestine involved.

A better understanding of the differentiation of the two diseases can be accomplished by looking at the effect of the two diseases as they involve the colon (Table [26-6](#)). Ulcerative colitis always involves the rectum and may extend, in continuity, to the more proximal colon; Crohn's disease has a segmental distribution with areas of normal bowel located between diseased segments. Rectal bleeding occurs in both diseases but is usually more of a problem in ulcerative colitis. The distinction between rectal and anal involvement is an important differentiating feature. In patients with ulcerative colitis, the rectum is always involved; only 50 percent of patients with Crohn's colitis will have rectal involvement. Anal disease, on the other hand, is highly suggestive of Crohn's disease. Anal fistulas are rare in patients with ulcerative colitis but are found frequently in patients with Crohn's disease.

Mucosal ulceration exists in both diseases, but the ulcers are more irregular and in a continuous distribution with ulcerative colitis, as opposed to being linear with transverse fissures (cobblestone appearance) in Crohn's disease. Strictures occur rarely in ulcerative colitis; when they do, they are often indicative of cancer. On the other hand, benign stricture formation is a characteristic of Crohn's disease. Ulcerative colitis is considered a premalignant condition. While there is an increased risk of large bowel cancer in patients with Crohn's disease, the risk is less than that of ulcerative colitis. Toxic dilatation, or more appropriately, fulminant colitis, occurs in both diseases. In patients with fulminant ulcerative colitis, the inflamed colon can dilate, presenting the clinical situation termed *toxic megacolon*. Severe inflammation can cause colonic perforation in patients with Crohn's disease, but because of fibrosis and thickening, dilatation is less frequently a component of the severe colonic involvement.

Each of the diseases can present with variable intensity, and the clinical management depends on the activity of the disease. In some cases, there can be rapid progression to the fulminant phase. Management of the patient with IBD requires knowledge of the *acuity* of the disease as well as the geographic *distribution* within the intestinal tract. *Mildly active* Crohn's disease consists of mucosal aphthoid ulcers; mildly active ulcerative colitis consists of mucosal edema and friability. These subtle changes are sometimes difficult to recognize by endoscopy. As the activity progresses to the *moderately severe* form, Crohn's disease is characterized by thickening of the intestinal wall, deep linear ulceration of the mucosa, edema of the bowel wall, wrapping of fat about the serosa, mesenteric thickening, and enlarged mesenteric lymph nodes. In moderately severe ulcerative colitis, there is progression of the mucosal disease with more purulent discharge, diffuse ulceration, and bleeding. As the diseases enter their *severe acute* inflammatory phase, Crohn's disease is characterized by deep ulceration of the bowel wall, increased serosal fat wrapping, and extension of the inflammatory process beyond the wall of the intestine, resulting in a phlegmon or abscess. Ulcerative colitis, on the other hand, is associated with the formation of pseudopolyps with severe fluid, electrolyte, and sometimes blood loss from the intense inflammatory process involving the mucosal surface of the colon and rectum.

The *chronic phase* of both diseases is depicted in Fig. [26-49](#). In ulcerative colitis, the mucosa takes on a dry, granular appearance, and the colon is markedly contracted, as seen in the radiograph. The colon loses its distensibility, and in this phase, where most of the lining of the intestine looks abnormal by endoscopy, the identification of malignant change becomes important. Crohn's disease, in its chronic phase, is characterized by intense fibrosis of the intestinal wall without acute mucosal inflammatory changes. If the rectum is involved, the loss of the reservoir function results in intolerably frequent, and often difficult to control, bowel movements, to the point where patients can be incapacitated with the chronic form of the disease and yet not have any acute inflammatory changes. Steroids and immunosuppressive medications will have little effect on this chronic, fibrotic state.

The two diseases are also similar in the use of antibiotics, anti-inflammatory agents, and immunosuppressive drugs in their management. Medical therapy is summarized in Table [26-7](#).

Ulcerative Colitis

Indications for Surgery

The indications for surgery in the treatment of ulcerative colitis include *active disease unresponsive to medical therapy*, *risk of cancer*, and *severe bleeding*. Because only the large intestine is involved, a proctocolectomy should cure the patient of intestinal disease. Patients requiring operation because of active disease usually have extensive edema and ulceration of the intestine; fluid, electrolyte, and blood loss; and sometimes dilatation of the colon with possible perforation (Table [26-8](#)). These patients with *toxic colitis* or *toxic megacolon* must be treated aggressively with bowel rest, antibiotics, and corticosteroids. Should their clinical situation deteriorate, they should have emergency abdominal colectomy or, if walled-off perforation is encountered, a diverting loop ileostomy combined with a decompressing colostomy, as described for fulminant Crohn's colitis. Barium enema, antidiarrheal agents, and morphine should be avoided because they might intensify the colonic dilatation. Careful observation of the patient is more important than serial radiographs of the abdomen looking for an arbitrary limit to cecal dilatation as an indication for surgery.

The patient at risk for cancer usually has had total colonic involvement for a period of at least 10 years. Studies have shown a 10 percent incidence of cancer in patients with such chronic disease. The risk of cancer in ulcerative colitis is related to the extent of colonic mucosal involvement and to the duration of the disease but apparently not to the intensity of the inflammation. Unfortunately, some patients who had childhood ulcerative colitis that became quiescent develop colon cancer in early adult life. The cancer risk was not recognized because the disease was in long-term remission. Such a colon at risk for developing malignant changes is illustrated in Fig. [26-49](#). Before the advent of colonoscopy, it was observed that almost all patients who developed cancer of the colon manifested changes of mucosal dysplasia that could be detected by sigmoidoscopy and biopsy of the rectum. It was this finding that led to a cancer surveillance plan for patients with chronic ulcerative colitis. Now colonoscopic evaluation of the entire large intestine is possible, and a plan for organized surveillance of patients with chronic ulcerative colitis has been developed, as illustrated in Fig. [26-50](#).

Colonoscopic surveillance has been recommended by the American Society of Gastrointestinal Endoscopy after 7 years of disease in patients with pancolitis and after 10 years of disease in patients with left-sided colitis. Colonoscopy and biopsy for dysplasia should be performed every 1 or 2 years in these patients. If severe dysplasia is identified, surgical

removal of the colon and rectum is indicated. The surveillance regimen includes random segmental biopsies and biopsies of any suspicious lesions. Random biopsies are necessary because malignant growths that occur in ulcerative colitis are frequently of the endophytic, rather than exophytic, type of cancer and are difficult to detect with the colonoscope (Fig. 26-51). Colon cancer in patients with ulcerative colitis is a major exception to the concept of polyp–cancer sequence, i.e., the cancer may arise in a flat mucosal surface. The pathologic determination of *dysplasia* is often difficult because the early changes of dysplasia are subtle, and dysplasia cannot be diagnosed in the presence of active inflammation.

Surgical Management

The characteristic inflammation of ulcerative colitis is confined to the mucosa of the large bowel. Complete excision of the large intestine is curative in this disease (unlike in Crohn's colitis, which carries a significant risk of recurrent disease following excision of involved intestine). These characteristics provide the rationale for sphincter-saving reconstructive procedures as treatment for ulcerative colitis. Since the disease is confined to the mucosa, the rectal sphincters can be preserved; since the disease does not involve the small intestine, a reservoir can be fashioned from the ileum that can serve as a substitute for the excised rectum. If the anal sphincter cannot be preserved, a “continent ileal reservoir” can be constructed (Table 26-9). The ileorectal anastomosis or any other procedure that leaves the rectum in situ carries the risk of recurrence of inflammation as well as the risk of subsequent cancer. The surgical options are illustrated in Fig. 26-52.

The current preference is to give a patient the choice for a sphincter- preserving operation as long as the individual can tolerate the increased magnitude of the surgery and has a demonstrated adequate anal sphincter function. It is most important that the patient not have Crohn's disease. In a few patients, this distinction is difficult to make, and a subset of patients has been recognized as having “indeterminate colitis.” The patients treated by a sphincter-preserving operation have an increased risk of complications at the ileoanal anastomosis and have an increased risk of requiring resection of the pouch because of subsequent complications. Total colectomy with ileal pouch–anal anastomosis also requires, in most cases, a temporary ileostomy, making this a two-stage procedure. Patients who have an impaired anal sphincter or those who do not wish to accept the increased risk of the reconstructive procedure should be treated by total proctocolectomy and ileostomy. If the surgeon treating a patient with fulminant colitis is unfamiliar with the sphincter-saving procedures, an abdominal colectomy with closure of the rectum allows the patient to recover from the life-threatening colitis. Subsequently, this patient can be referred to an institution where the more complex reconstructive procedure can be accomplished.

The use of the continent ileostomy or Kock pouch (see Table 26-9) has decreased since the advent of the new ileal pouch–anal anastomotic procedures and probably should be offered only to patients with ulcerative colitis who have permanent ileostomies that cannot be maintained properly because of severe skin allergies or to patients who have been treated by total proctocolectomy but desire a form of continence. The only advantage of the continent ileostomy is avoidance of the need to wear a permanent appliance. It requires intubation six to eight times a day to empty the pouch. There is also the risk of delayed complications related to the loss of the nipple valve created by intussusception of the ileum.

The preferred sphincter-saving operation for a patient with ulcerative colitis is total proctocolectomy with preservation of the anal sphincters, construction of an ileal reservoir, and anastomosis of the reservoir to the anus (Fig. 26-53 A,B). This removes the distal disease (some current forms of the operation leave a small amount of distal rectum at the anastomosis), maintains transanal defecation with reasonable continence, and provides a satisfactory quality of life without a permanent ileostomy. The main disadvantages of the procedure are frequent bowel movements (approximately five to seven per day), nocturnal fecal soilage (occasional in 55 percent of patients), and “pouchitis” (usually a transient inflammation of the pouch that responds to metronidazole).

The details of this operation include excision of the entire colon and proximal rectum, but only the diseased mucosa of the distal rectum and proximal anal canal is removed. Three centimeters of rectal muscularis, the pelvic floor, and the entire anal sphincter mechanism are preserved. Then 30 cm of terminal ileum is fashioned into a neorectum using either a J or S configuration (J is simpler; S has larger capacity and reaches the anus with less difficulty). Intestinal continuity is restored by connecting the pouch to the anal canal at the dentate line. A temporary ileostomy allows healing of the pouch and anastomosis and protects the patient from sepsis should an anastomotic leak occur. After 2 or 3 months, integrity of the pouch, anastomosis, and sphincter mechanism is evaluated; if all is well, the ileostomy is closed, restoring intestinal continuity. The first 3 months after ileostomy closure can be difficult because of the frequent need to pass stool. The pouch capacity increases with time, and the need to empty the pouch can be reduced to an average of four or five times per day.

None of the surgical procedures for ulcerative colitis should prevent completion of normal pregnancy in women who have been cured of the disease. Even normal vaginal delivery is a possibility after any of the resective or restorative procedures.

Crohn's Disease

General Considerations

The diagnosis of Crohn's disease usually takes longer to establish than that for ulcerative colitis, possibly because of the nonspecific nature of the symptoms: diarrhea, abdominal pain, weight loss, and anal disease. The mean time from onset of symptoms to diagnosis is approximately 35 months.

Surgery for Crohn's disease is indicated only for the management of complications, not for cure of the disease, since recurrence is common. Surgery for Crohn's disease is preferably done as an elective procedure rather than waiting until severe complications make urgent surgical intervention necessary, increasing the risk of postoperative complications. The choice of surgical management for Crohn's disease must take into account the geographic distribution of the disease within the intestinal tract and the nature and severity of the complication requiring surgery.

The intensity of activity of the Crohn's disease may determine the type of complication that occurs and may affect the prognosis, including risk and rate of recurrence. Crohn's disease may be considered to have two phases of activity. The *acute* inflammatory process consists of intense inflammation of the involved intestine, deep ulceration into the bowel wall with possible perforation, phlegmon, or abscess, and adherent loops of bowel not involved with the primary disease. The *chronic* fibrotic phase of the disease exists for a longer period and results in narrowing and stricture formation. Any adjacent organs may be involved by the inflammatory process. Hydronephrosis of the right kidney can result from ileocolonic Crohn's disease causing partial occlusion of the ureter. The condition is relieved when the Crohn's disease is resected.

The surgeon should prepare the patient adequately so that a single operation will suffice to correct the complication requiring treatment. This often involves the use of anti-inflammatory medications, antibiotics, complete bowel rest (including parenteral nutritional support), and when indicated, percutaneous drainage of abscesses. The nutritional fitness of the patient undergoing surgical treatment probably contributes more to a favorable outcome than most other variables. By careful preoperative management, the surgeon can avoid operating on a phlegmon with resulting injury to normal, uninvolved intestine with the risk of postoperative disruption of the anastomosis and subsequent fistula formation leading to the eventual creation of a short bowel syndrome.

Surgical Management

Surgical management of Crohn's disease involves resection of the diseased segment of intestine responsible for the complications requiring operation. The extent of resection usually encompasses intestine that is grossly involved with the disease. Recurrence of disease is not reduced by obtaining wide margins of uninvolved intestine or by verifying normal margins by frozen-section histology. The uninvolved margin of resection usually can be detected by demonstrating a normal margin at the mesenteric-intestinal border. It is not necessary to include a large segment of mesentery in the resected tissue. If normal-appearing bowel remains after resection and there is not significant intraabdominal infection, intestinal continuity can be restored by primary anastomosis. Should there be significant intraabdominal infection or intestine that is so severely inflamed that it cannot be removed safely, an intestinal stoma fashioned from bowel proximal to the inflamed intestine will allow the patient to resume eating and taking medication until the inflammatory process resolves and definitive surgical resection can be accomplished.

The symptoms requiring surgical resection depend on the geographic distribution of the disease.

Ileocolonic Crohn's Disease

As shown in Table 26-10, ileocolonic Crohn's disease presents with internal fistula and abscess formation as often as it does with intestinal obstruction. Here the surgical principle of draining an abscess and allowing a phlegmon to resolve becomes critically important. The timing of operation should be such that sepsis is controlled by total bowel rest and percutaneous drainage of abscesses before definitive resection is carried out. Usually intraabdominal abscesses associated with ileocolonic Crohn's disease can be drained percutaneously with the aid of guidance by CT. This allows definitive surgery after the inflammation has resolved, avoiding injury to noninvolved small intestine and other intraabdominal structures that may be involved secondarily by the inflammatory process. Drainage of such abscesses can lead to fistula formation because of the factors summarized in Table 26-11. For this reason, abscesses should be drained from locations on the abdominal wall that would be amenable to application of an appliance, as used for an intestinal stoma. This allows containment of the output even if it should be of high volume and potentially damaging to the abdominal skin.

Colonic Crohn's Disease

Patients with colonic Crohn's disease often require operation after a shorter duration of symptoms than is usual for patients with small intestinal or ileocolonic disease (see Table 26-10 B). Colonic perforation can occur in patients with Crohn's disease and toxic colitis without the dilatation of the colon that is seen in patients with toxic ulcerative colitis. This is probably due to the thickening of the colonic wall, which prevents dilatation; thus the term *toxic megacolon* is not usually appropriately applied in Crohn's colitis. The prognosis and risk of recurrence for Crohn's colitis depend significantly on the presence of simultaneous involvement of the small intestine. The surgical management depends somewhat on the presence and severity of rectal involvement. The risk of recurrence in Crohn's disease prohibits the use of the more sophisticated sphincter- preserving procedures described for ulcerative colitis and sometimes for familial polyposis.

The choices of surgical procedure for Crohn's colitis include abdominal- perineal proctocolectomy with a permanent end ileostomy, total abdominal colectomy with an end ileostomy (with retention of the rectal segment as a Hartmann's pouch), and total abdominal colectomy with ileorectal anastomosis (sometimes protected with a proximal loop ileostomy). The specific choice of surgical procedure is governed by the condition of the rectum and anal sphincter, as well as by the patient's metabolic and emotional status. The latter becomes an important factor because this disease, which frequently involves young adults, should not result in construction of a permanent stoma unless the individual can participate in the decision. A permanent ileostomy presents a change in body image; and although it can be life-preserving, the choice of a permanent stoma should be made by the patient, not by the parents or the physician. Since an ileostomy is not truly permanent until the rectum is removed, significant effort should be made to retain the rectum and preserve the hope of restoration until the patient realizes that the ileostomy is tolerable; the diseased rectum can interfere with long-term health and normal life activity. Recent reports indicate that the retained rectum, even for Crohn's disease, is at some increased risk of developing cancer. Therefore, surveillance of the rectal stump by regular endoscopy is necessary.

The patient presenting with *acute, catastrophic* Crohn's disease is so ill that urgent operation is often necessary. These patients should be managed as previously described for toxic ulcerative colitis. Intense medical treatment, including intravenous hyperalimentation and antibiotics, should be instituted. Some patients revert to a more stable condition, but others require urgent surgical treatment. If peritoneal signs and the general status of the patient do not improve rapidly, operation is required. A site should be marked

preoperatively for a possible ileostomy and, if the colon is massively dilated, for a possible decompressing transverse colostomy as well. Sigmoidoscopy should be done in the operating room, after induction of anesthesia, to determine the status of rectal disease.

At the time of operation, if a severely diseased intact colon is found, it should be removed along with any involved terminal ileum. If the rectum is uninvolved and the patient is stable, consideration can be given to constructing an ileorectal anastomosis. If the status of the rectum is questionable, an ileorectal anastomosis can be performed but should be protected with a proximal loop ileostomy. If the rectum is significantly diseased, it should be left in place, the colon removed, and an end ileostomy established. Once the patient achieves an adequate level of postoperative health and nutrition, the rectum can either be excised or incorporated in an ileorectal anastomosis. If the rectum is severely diseased, it is safer to make a formal mucous fistula than to close the end of the rectum and leave it in place. This mucous fistula decreases the risk of rectal perforation. If a severely diseased colon with free perforation is found at the time of exploration, colectomy is necessary, but no anastomosis should be made. The abdominal cavity should be irrigated, and an end ileostomy should be constructed with the rectum left in place.

In rare situations, exploration will reveal a sealed-off perforation of the colon, usually in the region of the splenic flexure, with a tense, thin-walled colon. Resection in this situation is unsafe; the surgeon should treat this with a loop ileostomy and a "blowhole" decompressing colostomy, as described by the late Rupert Turnbull and his colleagues. This operation requires both diversion and decompression; neither component should be used by itself.

In a patient with *long-standing colonic Crohn's disease*, especially with severe anal or rectal involvement, a single-stage proctocolectomy and ileostomy may be done. The proctectomy should be done differently from that performed for cancer in order to preserve normal sexual function (Figs. [26-54](#) and [26-55](#)).

Rarely, a colonic stricture can form in an isolated segment of Crohn's disease, causing obstructive symptoms. Resection of such a segment should be considered if the obstructive symptoms are significant. Colonoscopic dilation of strictures has been described in situations where surveillance of the proximal colon was not possible. Since the colon has no critical physiologic function, stricturoplasty to retain a colonic segment is seldom indicated; involved segments should be resected and intestinal continuity restored. Complete colectomy is the better procedure for extensive colonic disease.

Rectal Crohn's Disease

Crohn's disease can involve the rectum or the anus, either separately or simultaneously. It is critical to differentiate these anatomic distributions of Crohn's disease. If a patient with rectal Crohn's disease has severe bleeding and tenesmus that cannot be relieved by medical management (including bowel rest with total parenteral nutrition), consideration should be given to removal of the rectum or at least temporary diversion with a colostomy or an ileostomy. Once the rectum has become so fibrotic that it has lost its reservoir function (see Fig. [26-49](#)), proctectomy should be considered because of intolerably frequent, and sometimes poorly controlled, bowel movements. In the rare situation where only the rectum is involved with Crohn's disease, a completely diverting colostomy may allow the rectum to regain its functional capacity. This is especially important in the patient who has had previous small-bowel resections, when it becomes more important to preserve the water-absorbing capacity of the right colon. An end colostomy also may be considered for patients with colonic sparing but severe destructive anal or perineal Crohn's disease. In such cases, diversion may be needed to control sepsis or anal incontinence. If a colostomy is done in a patient with Crohn's disease, it should be constructed as a protruding spigot similar to that of an ileostomy.

The proctectomy in a patient with Crohn's disease should be performed by confining the rectal dissection to the immediate proximity of the rectum (see Figs. [26-54](#) and [26-55](#)), but any fistulous tracts extending to the perineum should be radically debrided.

Certain precautions must be taken during proctectomy for IBD. The nerves to the genitalia and the urinary bladder lie in close approximation to the rectum (see Fig. [26-8](#)). There is a significant incidence of injury to these nerves when a proctectomy is done for cancer, because some of the surrounding tissue must be included in the resection. For IBD, however, the dissection of the rectum need not involve adjacent tissue, and the dissection should be precisely on the rectal wall.

Another complication of removal of the rectum for IBD is nonhealing of the perineal wound, usually more of a problem in Crohn's disease than in ulcerative colitis. This can be avoided by preserving the external sphincter and levator muscles. The operation involves removal of the rectum from the perineal field and secure closure of the musculature, accompanied by drainage of the pelvis from above, as depicted in Fig. [26-55](#).

Another complication of proctectomy occurs in women when the ovaries, which are normally intraperitoneal structures, fall into the space formed by removal of the rectum. The peritoneum may reconstitute in a fashion that excludes the ovaries from the peritoneal cavity, in effect making them extraperitoneal organs (Fig. [26-56 A](#)). As ovulation occurs, large cystic structures can form in the pelvis, and the woman obviously has reduced fertility. This situation can be avoided by simply suturing the ovaries out of the pelvis after the rectum is removed (see Fig. [26-56 B](#)).

Anal and Perineal Crohn's Disease

Involvement of the anus occurs in 35 percent of all patients with Crohn's disease at some time during the course of their illness. Four percent of patients present with only anal disease as the first manifestation of Crohn's, but all go on to have other intestinal involvement. It is important to recognize the possibility of anal disease being secondary to the presence of Crohn's, because the surgical treatment should be much less radical than for non-Crohn's-related anal disease. Procedures such as excisional hemorrhoidectomy and extensive fistulotomy should be avoided because of the necessity of preserving anal sphincter function in a patient with inflammatory bowel disease who is prone to have diarrhea and loss of rectal capacity. Of particular importance is preservation of the anterior aspect of the sphincter mechanism in a woman, in whom continuity of the sphincter muscle is maintained by only one band of external striated muscle. Through the use of conservative and often prolonged and tedious management of anal and perineal disease, very few patients with Crohn's disease require proctectomy because of the anal-perineal disease. This is especially important as more effective medical management becomes available. Conservative surgical drainage of septic processes and the use of intensive medical treatment allow preservation of the rectum and anal sphincter function and in many instances allow the Crohn's disease to go into remission so that the source of the sepsis can be resolved to one small internal opening, which can be closed with the endoanal sliding flap advancement technique (see Fig. [26-44](#)). The medications most advantageous for treating anorectal Crohn's disease include (see Table [26-7](#)) the antibiotic metronidazole, the anti-inflammatory agents 5-aminosalicylic acid and cortisone (by enema or suppository), and such immunosuppressive agents as azathioprine (Imuran), 6-mercaptopurine (6-MP), and cyclosporine. The long-term effectiveness of the agents is difficult to evaluate because of the intermittent nature of the disease. This conservative management is not possible when severe destruction of the rectum exists. These patients often require eventual proctectomy because the loss of the reservoir capacity of the rectum is intolerable.

Anal Crohn's disease should be suspected when a patient presents with inflamed edematous tags (mistaken for external hemorrhoids), anal or perineal fistulas, or abscesses (Fig. [26-57](#)). Often there are multiple abnormalities and poor healing after minor anal surgical procedures. Management of such a patient should begin with examination under anesthesia, including sigmoidoscopy, to determine the presence of disease within the rectum. The edematous skin tags can be removed if they are causing symptoms. This tissue has been shown to serve as a rich source for *granulomas*, which can help confirm the diagnosis of Crohn's disease. The objective of treatment then becomes the management and control of septic processes, in anticipation of medical treatment to put the Crohn's disease into remission. This can be accomplished with mushroom-tipped catheters or unroofing of abscesses and with the use of setons to provide drainage of fistulous tracts. A superficial fistulotomy can be performed if the external sphincter mechanism is not involved. Special care must be observed for preservation of the anterior component of the sphincter mechanism in women. These conservative and tedious drainage processes may go on for years in an attempt to preserve the rectum. As part of this preservation, evaluation of the sphincter mechanism should be carried out, including manometry and electromyography. Intestinal continuity should never be restored to a poorly functioning anorectum or to one that is extensively destroyed. Such patients would do better with resection and a good intestinal stoma.

Outcome of Surgical Treatment of Crohn's Disease

The incidence of the need for operation and the development of recurrence in the postoperative period for a large series of patients with Crohn's disease, followed over a long period of time at the Cleveland Clinic Foundation, can be seen in Table [26-12](#). After the immediate perioperative period, the major complication of Crohn's disease is the high rate of recurrence. Its true incidence can only be discovered by long-term observation of a large, carefully defined population. This series, originally reported by Farmer, attributed the risk of recurrence to the *pattern of initial presentation* of Crohn's disease within the intestinal tract, suggesting that ileocolonic Crohn's disease had a higher recurrence rate. After 15 years, the rates of recurrence, based on patterns of presentation, converged. Thus the reliance on disease distribution to predict recurrence was no longer useful. Their series showed, in addition, that the rate of recurrence seemed to be biphasic, with the slope of the recurrence curve improving 8 years after surgery, suggesting that the longer the patient lives without recurrence, the less likelihood there is of the disease presenting itself again. Both Farmer and Sachar, from another institution, presented statistics to indicate that the *severity* of the presenting symptoms was perhaps a better predictor of risk for severe complications and early, frequent recurrence of disease. Perforation and intense inflammatory changes as presenting symptoms, including severe anal or perineal disease, carried a much higher likelihood of postoperative recurrence than more indolent symptoms such as obstruction. We also have seen a more rapid as well as increased incidence of recurrence and a more severe intensity of disease in the African American population. Although Crohn's disease was initially said to exist primarily in the Jewish population and to spare blacks, we found that not only did African-Americans acquire Crohn's disease, but also a higher percentage required operation. There was also an increased requirement for second operations, and all of them had extraintestinal manifestations of Crohn's disease. Recent studies have suggested a relationship between the rate of recurrence and the immunologic status of the patient, showing a lower and slower recurrence rate in patients with ileocolonic resections who had the immunosuppressive effect of preoperative blood transfusions. The interpretation of this clinical finding is still not clear.

An important principle to which one must adhere is that the need for surgical management of a complication of Crohn's disease should not be ignored because of the risk of recurrence. In addition, the fact that one operation for resection of Crohn's disease is necessary does not increase the subsequent need for further surgery, nor does it mandate the eventual creation of a short bowel syndrome.

Ileostomy Construction and Management

An ileostomy can be constructed with different configurations, depending on the goal to be accomplished by its construction. A permanent ileostomy is usually constructed with the end of the ileum, and enough of the mesentery is preserved to allow adequate vascularization (Fig. [26-58](#)). The stoma is everted on itself to prevent the serious problem of "ileostomy dysfunction" that historically led to severe fluid and electrolyte imbalance due to partial obstruction of the small intestine. This technique of eversion was described by Brooke and by Crile and Turnbull in 1954. A *loop* of ileum also may be used for construction of a stoma (Fig. [26-59](#)). This leaves the mesentery intact and is usually constructed on a temporary basis to protect distal, precarious anastomoses or the new restorative procedures for maintaining anal continence, although these latter procedures are not indicated in patients with Crohn's disease because of the risk of recurrence. A loop ileostomy can be constructed with the bowel in continuity or separated (Fig. [26-60](#)).

The techniques of ileostomy construction are based on the fact that the output from an ileostomy is relatively high in volume, liquid in consistency, and contains active proteolytic enzymes. Thus a properly constructed ileostomy must be situated precisely on the abdominal wall at a site chosen carefully preoperatively with the patient in various positions (Fig. [26-61](#)). The abdominal incision performed on patients for IBD should always preserve the right lower quadrant skin, even if the possibility of an ileostomy seems remote at the time of the operation. The ileostomy stoma must be constructed with a protruding spigot configuration to allow an appliance to seal perfectly to the skin (Fig. [26-62](#)). An ileostomy stoma must have adequate blood supply, a goal not easy to accomplish, especially in the obese patient.

An individual with a well-constructed and properly located ileostomy should be able to empty the appliance four to six times a day and should not need to have complete change of the appliance more often than every 4 or 5 days. Neither belts nor supporting straps should be necessary, and the ostomate should be able to perform all routine activities (Table 26-13). The only true disability caused by a properly constructed ileostomy is a change in body image.

Long-term management of an ileostomy includes provision of adequate fluid replacement in hot weather, because the individual is at risk for rapid dehydration during periods of intense physical activity. Oral electrolyte- containing solutions are beneficial for this purpose. The ileostomy patient also should be advised to have available antidiarrheal agents for control of episodes of high output from the stoma. Large quantities of fibrous foods should be avoided, because they can cause blockage of the intestinal stoma (Fig. 26-63). If this occurs, the patient must be seen in the emergency room, where the ileostomy can be irrigated through a #24 Foley catheter inserted just below the fascia with 3 mL of saline in the balloon. Then 50 mL of warm saline is used to irrigate the stoma. The diagnosis of food blockage can be confirmed by the passage of particles of food after each irrigation. These patients often require intravenous fluids and nasogastric decompression to alleviate the severe cramps and dehydration that result from these episodes of small-intestinal obstruction. If the obstruction is not diagnosed as due to food blockage, other causes should be sought, and a water-soluble contrast study should be performed through the intestinal stoma. This can serve as a therapeutic as well as a diagnostic modality because of the cathartic action of the contrast solution. Digital inspection of an ileostomy is inappropriate, because the stoma is smaller than that of a colostomy, and insertion of a finger may be injurious.

A patient with a properly constructed ileostomy should have a normal life expectancy, assuming no malignant growth existed at the time of initial operation, and should anticipate no restrictions of normal life activity, with the exception of the extra time needed to provide meticulous care of the intestinal stoma.

NEOPLASTIC DISEASE

Etiology

Although the exact cause of colon and rectal cancer is not known precisely, recent laboratory work has provided a veritable explosion in the understanding of the molecular basis of colorectal carcinogenesis. These advances have significantly altered many concepts about the development of cancer of the large bowel, and studies at the molecular level promise to radically alter colorectal cancer screening, diagnosis, treatment, and prognostication in the coming years. It is now recognized that colorectal cancer is a genetic disease, meaning that the cause of the uncontrolled proliferation of cells that we recognize as carcinoma resides in abnormalities or alterations in the genetic code. These genetic mutations may be in the germ line, in which case the disease is inherited, or somatic mutations in the genes may occur, which would result in the type of cancer that we usually have referred to as *sporadic*.

Genetic Predisposition

It has been estimated that approximately 10 to 15 percent of colorectal cancer is familial and that 1 person in 200 may carry high-risk alleles of the genes causing inherited colorectal cancer. It also is estimated that the same genes may be involved in as many as 15 percent of sporadic cancers.

Familial Adenomatous Polyposis

Pathophysiology and Diagnosis

Familial adenomatous polyposis (FAP) is the prototypical hereditary polyposis syndrome. It is an autosomal dominant disorder diagnosed when a patient has more than 100 adenomatous polyps in the large bowel or when a member of an FAP family has any number of colonic adenomas detected. FAP includes a spectrum of syndromes formerly defined by various extraintestinal manifestations. FAP also includes Gardner's syndrome (epidermal inclusion cysts, osteomas of bone, and colonic polyps) and Turcot's syndrome (colonic polyps and brain tumors). The genetic defect responsible for adenomatous polyposis has been detected on chromosome 5, near the q21 locus. The gene has been called the APC (adenomatous polyposis coli) gene. FAP occurs in anywhere from 1 in 8000 to 1 in 29,000 individuals. The common expression of the syndrome is the presence of multiple polyps in the colon and the associated extraintestinal manifestations, including epidermoid cysts, desmoid tumors in the abdomen, osteomas of the bones, and brain tumors (usually gliomas or medulloblastomas). The gene expression occurs in 100 percent of patients with the defect. Autosomal dominance results in expression in 50 percent of offspring. Between 10 and 20 percent of all patients with FAP appear to harbor the genetic defect as a result of spontaneous mutation. All patients with the defective gene will develop cancer of the colon if left untreated. The average age of discovery of a new patient with FAP is 29 years. The average age of a patient who is newly discovered to have colorectal cancer related to FAP is 39 years.

Two rare variants of FAP include *attenuated adenomatous polyposis coli* (AAPC) and *hereditary flat adenoma syndrome* (HFAS). AAPC patients have relatively few polyps but retain a high risk of colorectal cancer. Patients with HFAS develop small adenomas (usually less than 100) that are frequently dysplastic and prone to malignant transformation. One interesting marker for FAP is congenital hypertrophy of the retinal pigmented epithelium (CHRPE), which is seen on indirect ophthalmoscopy in approximately 70 to 80 percent of FAP patients.

Our understanding of the genetic basis for colorectal cancer has resulted in clinical applications for screening of patients in FAP kindreds by DNA analysis for the APC gene or its products in peripheral blood leukocytes. This test, which requires only venipuncture, has obvious advantages over the standard screening, which consists of an annual endoscopy. This is particularly useful for young patients for whom screening is indicated. A single negative blood test, showing the noninheritance of a mutated APC gene, suggests that the individual will not develop polyposis and may not require intensive screening. This depends on the genetic defect's being clearly identified by testing the blood of known involved members of the family. It is believed that occasional surveillance colonoscopy is still indicated.

Upper endoscopy should be used to screen for gastric and duodenal adenomas beginning at age 30, even though such polyps have been found in patients under 20 years of age. Prior to the development of genetic screening for the APC gene, young children were screened by ophthalmoscopic examination. If the patients were found to have an area of retinal pigmentation, the chances of developing colonic polyps approached 100 percent. DNA analysis probably will replace the need for such screening procedures.

Surgical Management

Appropriate treatment for patients with FAP is removal of all colonic and rectal mucosa. Total proctocolectomy with permanent ileostomy accomplishes this goal, but the development of sphincter-saving procedures has virtually eliminated this operation as an acceptable form of therapy, unless the patient should have already developed a cancer of the distal rectum before operative treatment.

Total proctocolectomy with ileal reservoir and ileoanal anastomosis is the most commonly recommended operation for patients with FAP. The procedure must include a mucosectomy to the level of the dentate line to remove all mucosa at risk. Patients who undergo this procedure for FAP have a better functional result than patients similarly treated for ulcerative colitis, in that the incidence of inflammation in the ileal pouch (pouchitis) is much less in patients with familial polyposis. An alternative approach, total abdominal colectomy with ileorectal anastomosis, has been used successfully by several institutions. This latter approach was used extensively before the development of the technique of ileal pouch–anal anastomosis. The disadvantage of subtotal colectomy is that the precancerous rectum remains in situ, and repeated endoscopic procedures are needed every 6 months to detect and eradicate (by fulguration or snare excision) any adenomatous polyps that appear in the rectum. The early fears of unacceptable incontinence, intractable diarrhea, and impotence in men that might accompany proctocolectomy with ileal pouch–anal anastomosis, thus making subtotal colectomy with ileorectal anastomosis the preferred procedure, have not been confirmed.

FAP patients treated by proctocolectomy of any type remain at risk of death from upper gastrointestinal cancer. In one series this risk was greater than the risk of death from cancer in the residual rectum after total abdominal colectomy and ileorectal anastomosis. These patients require continued periodic surveillance by gastroduodenoscopy.

Extracolonic Manifestations

Some patients with FAP develop epidermoid cysts, osteomas, desmoids, and upper gastrointestinal neoplasms (Gardner's syndrome). The epidermoid cysts require no treatment unless they become symptomatic or cause significant cosmetic deformities. It is extremely rare for osteosarcomas to arise from osteomas in patients with FAP. The intraabdominal *desmoid* is probably the only incapacitating, lethal extraintestinal manifestation. After surgical procedures, dense fibrous tissue forms in the mesentery of the small intestine and encroaches on the vascular supply, tethers the small intestine, and causes obstruction or ischemia. Treatment has been attempted with chemotherapy, radiation, tamoxifen, and sulindac. Operative therapy is completely inadequate and should be reserved only for decompression of obstruction or to relieve incapacitating symptoms. Sulindac has been reported to cause tumor regression in some patients with desmoid tumors.

Upper gastrointestinal neoplasms in patients with FAP (Gardner's syndrome) include duodenal or ampullary adenomas, which are premalignant, and hyperplastic fundic polyps in the stomach, which do not become malignant. The current recommendation is for upper gastrointestinal surveillance after the age of 30 and endoscopic polypectomy, if possible, to remove all large adenomas within the duodenum. If an ampullary cancer is discovered at an early stage, pancreatoduodenectomy (Whipple procedure) is indicated.

Turcot's syndrome (colonic polyps and intracranial gliomas or medulloblastomas) is very rare, with only 50 patients reported in the literature. Patients with Turcot's syndrome usually die at an early age, before developing colonic polyposis and before having children.

Hereditary Nonpolyposis Colorectal Cancer

Aldred S. Warthin first recognized an increased incidence of cancer in certain families in 1913. One family characterized by Warthin became the prototype of cancer families. This family was afflicted by an excess of uterine and gastric cancers. The family was evaluated four times in the past 80 years, and a predominance of colorectal cancers has been detected in family members. A half century after Warthin's observations, Henry Lynch collected data that accurately characterized these cancer-prone families. Colorectal cancer is the most predominant cancer in this syndrome, but it is not associated with multiple polyps, which characterize familial polyposis. For this reason, the syndrome was named *hereditary nonpolyposis colorectal cancer (HNPCC)*. In 1984 the terms *Lynch syndrome I and II* were proposed to indicate two apparently distinct clinical presentations. In Lynch syndrome I, the cancer usually occurs in the proximal colon, is site- specific within an affected family, and is the only type of tumor these patients develop. Lynch syndrome II is characterized by development of colorectal, endometrial, gastric, upper urinary tract, ovarian, and other types of cancer.

There are no phenotypic markers for HNPCC, so a standard clinical definition of the syndrome is necessary. The most widely accepted definition uses the *Amsterdam criteria*, which require (1) at least three relatives with histologically verified colorectal cancer, one of which is a first- degree relative of another, (2) involvement of a least two generations, and (3) at least one colorectal cancer diagnosed before the age of 50 years. HNPCC is characterized by an autosomal dominant inheritance pattern, an early age of colorectal cancer development, a predominance of right-sided cancers, and a tendency toward the development of synchronous and metachronous cancers.

Molecular scientists recognized widespread genetic instability in most cancers studied in patients with HNPCC, and for this reason, a DNA “mutator” gene was suspected as playing

a carcinogenic role. The particular genetic instability in tumors from patients with HNPCC is characterized by variations in length of “microsatellite” DNA, which are short repetitive DNA sequences located throughout the genome. While microsatellites vary in length among individuals (a fact used in forensic DNA “fingerprinting”), they should be the same in all cells from a single individual. HNPCC tumors have microsatellites that differ in length from the normal cells of the same individual, indicating a loss of fidelity in DNA repair. This defect in the DNA repair process has been referred to as a *replication error* (RER). To date, four mismatch repair genes have been identified, and these are responsible for the majority of, but not all, HNPCC cancers. These genes are *hMSH2*, located on chromosome 2p; *hMLH1*, located on chromosome 3p; *hPMS1*, on chromosome 2q; and *hPMS2* on chromosome 7p.

The cancers that arise in patients with HNPCC are more often poorly differentiated and mucin-producing than sporadic colorectal cancer. “Signet ring” cancers are also more common in HNPCC tumors. Despite these findings, which are usually associated with a poor prognosis, the prognosis for patients with HNPCC is more favorable than that of patients with sporadic colorectal cancer of the same pathologic stage. Approximately 20 percent of HNPCC patients with colon cancer will have a synchronous cancer, and about 25 percent will develop a metachronous cancer of the remaining colon. It is recommended that patients with this syndrome be treated with subtotal colectomy.

Colorectal cancers arising in HNPCC patients develop through a typical adenoma-carcinoma progression. The hypermutable cancer may accelerate the time frame in which malignancy occurs. The optimal intervals for surveillance in this disease have not yet been ascertained. Because most of the tumors will occur in the proximal colon, flexible sigmoidoscopy is not an effective screening approach. A 3-year interval of screening significantly reduces the incidence of colorectal cancer, but surveillance intervals of 1 or 2 years may be more appropriate for this syndrome. There has been no significant study on the impact of any diagnostic screening to prevent the development of upper gastrointestinal tract cancers or gynecologic cancers.

Assays undoubtedly will become available that identify mismatch repair gene mutations in suspected HNPCC kindreds. This advance will be of practical and theoretical importance: practical, to identify gene carriers in a given HNPCC kindred, and theoretical, because HNPCC kindreds are currently defined by family history alone, which may be incomplete or misleading. Individuals may carry the disease trait but not fulfill the current clinical criteria. Future definitions of HNPCC will be based on genetic analysis rather than on family history alone.

Acquired Somatic Defects

The chromosome defect in FAP is a germ cell line defect and is present from birth. The majority of patients with colorectal cancer have no inherited component, and the initiating genetic mutation that causes neoplastic growth occurs in a single cell or in a group of mucosal cells. These tumors are not inherited, but there may be a familial predisposition to their development. Lynch has shown that the risk for developing colorectal cancer is increased threefold to ninefold if one or more first-degree relatives have had colorectal cancer.

Available evidence suggests that adenomas of the colon develop after such a gene mutation. The adenomatous tissue contains crypt cells with atypical genetic expression. Tumor formation may be the result of more than one mutation. A *somatic* mutation, associated with an *inherited* mutation, may be expressed as the adenoma or tumor progresses. The inherited mutation is felt to be tumorigenic but may not be expressed or detectable in the tumor. Fearon and Vogelstein have proposed that, once the adenoma has developed, a second mutation may be required to change the adenoma to a cancer. A progression of mutations actually can occur in the course of the development of a malignancy. Oncogenes function by initiating events that result in abnormal cellular growth and the development of tumors. For example, the *ras* gene family has been found in 50 percent of colorectal cancers and in adenomas larger than 1 cm, but it has been found in less than 10 percent of adenomas smaller than 1 cm.

As a tumor develops, subsequent mutation can result in the deletion of suppressor genes. This loss usually occurs on chromosomes 17, 18, and 5. For example, the p53 gene is absent from chromosome 17 in 85 percent of colorectal cancers. The chromosome 5 MCC gene (mutated colorectal cancer) is found in some tumors in patients with nonpolyposis cancers. The chromosome 18 DCC gene (deleted in colorectal cancer) is found in more than 70 percent of colorectal cancers. These mutations are postulated to affect cell-to-cell relationships in all tumors and thereby also affect the development of metastases.

Fearon and Vogelstein have proposed a model for genetic changes leading to colorectal cancer (Fig. 26-64). First, a mutational *activation* is initiated by oncogenes. This mutational activation is accompanied by a mutational *inactivation* of multiple suppressor genes. The mutations allow uncontrolled growth in varying degrees. More mutations are required for the development of cancer than for the development of benign polyps. The progression of cellular changes from normal mucosa to adenoma to cancer could be related to a progression of mutations over time.

Environmental Factors Diet

Diet has been proposed as a significant etiologic factor in the development of colorectal cancer. Fats have been suggested to be toxic to the colonic mucosa. Diets high in oleic acid (found in olive oil, coconut oil, and fish oil) cause no increase in cancer in animals. The monounsaturated fatty acids (omega-3 and omega-6 fatty acids) appear to be less carcinogenic than polyunsaturated or saturated fats. Epidemiologic studies of populations that consume less than 5 percent of their diet as fat have shown a lower incidence of colorectal cancer, whereas those having diets with 20 percent of their fat as either corn oil or safflower oil have an increase in cancer. Therefore, it appears that unsaturated animal fats and highly saturated vegetable oils may cause some event to occur in the colonic mucosa that stimulates the development of colorectal cancer.

Some studies suggest that certain elements of the diet protect against the development of neoplasia. Selenium, dithiothiones, thioethers, terpenes, and carotenoids might act as “anticarcinogens” by reducing free oxygen radicals in the colon at the mucosal surface. An increase in dietary fiber may lower the incidence of cancer in patients who have a high-fat diet. An increase in fat, however, can negate the beneficial effect of the fiber. A high-fiber, low-fat diet has been described as better protection against neoplasia than a fat-free diet. An increase in bulk also tends to dilute carcinogens, shorten colonic transit time, and as a result, decrease the growth of harmful bacteria. So far, controlled trials have shown little relationship between dietary fiber and colon cancer, but the studies used a relatively low dose of fiber.

The National Research Council established interim guidelines for diet recommendations in 1982 until definitive studies could be performed. These recommendations included

- (1) Decrease fat in the diet from 40 to 30 percent of total calories.
- (2) Increase consumption of fiber-containing foods.
- (3) Limit salt-cured, pickled, and smoked foods.
- (4) Limit food additives shown to be potential carcinogens.
- (5) Establish research protocols for detection of mutagens in foods.
- (6) Limit alcohol consumption.

Epidemiologic studies indicate that diet has a role in the development of colon cancer. People in less-industrialized nations are noted to have a low incidence of colorectal cancers. These people have fewer processed foods and more natural fiber in their diets. People from these countries who immigrate to the industrialized United States or United Kingdom acquire the high risk of colon cancer found in these countries.

Exposure to Carcinogens

No clear relationship has been established between specific carcinogens and colorectal cancer, but potential agents currently under investigation include bile acids (such as chenodeoxycholic acid), food additives, alcohol, and ionizing radiation. Each of these may function as a promoter of mutational changes in the colonic mucosa, but none has been clearly shown to have a carcinogenic effect. Oxygen free radicals, which are unstable by-products of oxygen metabolism, may serve as promoters or chemical stimulants to the development of altered gene expressions. This is not a true carcinogenic effect but the creation of a “milieu” that predisposes to the development of mutations.

Premalignant Conditions Ulcerative Colitis

The overall incidence of neoplasia in patients with pancolitis is 1 percent per year after 10 years; so the cumulative risk of a cancer is 10 percent by 20 years' duration of disease. Dysplasia identified in the colonic mucosa by repeated colonoscopic biopsies has been shown to be a precursor of cancer. The cancers tend to be more advanced because they are difficult to detect at an early stage. Approximately 35 percent are Dukes' C or D lesions. There is a controversy over the implication of adenomatous polyps found in patients with chronic ulcerative colitis. The question arises as to whether this represents dysplasia and is an indication for total colectomy or whether this is a benign polyp arising from inflamed mucosa.

The most effective method for preventing colon cancer in patients with ulcerative colitis is to remove the colon once dysplasia has been identified. The risk of cancer in the presence of dysplasia is approximately 30 percent.

Crohn's Disease

The overall incidence of cancer occurring in patients with Crohn's disease is approximately 7 percent over 20 years. Most of the intestinal cancers were found in the excluded segments of patients treated by segmental bypass. This procedure is no longer recommended because of a 20 percent incidence of adenocarcinoma and persistent disease in these bypassed segments. Patients with colonic strictures have a slightly higher incidence of adenocarcinoma at the site of the fibrotic narrowing. Reports of adenocarcinoma arising at the site of stricturoplasty in the small intestine have made it apparent that there is a risk of adenocarcinoma in these lesions. For this reason, a biopsy of the intestinal wall should be performed at the time of stricturoplasty. There also have been reports of squamous cell cancer and adenocarcinoma arising in chronic fistulas (especially of the anus and perineum) in patients with Crohn's disease. Overall, the risk for neoplasia in patients with Crohn's disease is only slightly higher than that of the general population and is certainly lower than in patients with ulcerative colitis or familial adenomatous polyposis.

Detection

Tests for Occult Blood

A colorless phenol in guaiac gum is converted to a blue quinone by oxidation. The reaction relies on the presence of a peroxidase catalyst. Oxidation is accomplished by the addition of hydrogen peroxide in the presence of a catalyst, such as hemoglobin. Unfortunately, there are other catalysts in the diet. Careful elimination of some of these products may be necessary to improve the accuracy of the test. Red meat, some vegetables, and other intrinsic peroxidases may cause the reaction in the absence of hemoglobin. The

guaiac occult blood test will detect 20 mg of hemoglobin per gram of stool or 20 mL/day.

An immunofluorescent test for occult blood relies on the conversion of hemoglobin to fluorescent porphyrins and detects between 5 and 10 mg of hemoglobin per gram of stool. The false-positive rate is prohibitively high.

There are inherent problems in using the detection of occult blood as a screening test for colorectal cancer. Other sources of bleeding may produce positive tests, or the test may be too sensitive for blood. Cancers might bleed only intermittently or not at all, and the occult blood test may be falsely negative. Test processing, dietary manipulation, ingestion of gastrointestinal irritants (aspirin), the number of tests, and the interval of testing all affect the accuracy of the occult blood screening process.

There have been numerous uncontrolled studies of occult blood testing for screening populations for colorectal cancer. These reports usually show 2 to 6 percent positive tests. The positive predictive value for cancer in these studies has been 2 to 5 percent, and the positive predictive values for a polyp in these studies has been close to 20 percent. Cancer has been detected in 0.03 to 0.2 percent of all tested subjects, and polyps have been found in 0.6 percent.

The results of screening 46,000 individuals for occult blood using the guaiac test revealed a sensitivity of 89.3 percent, a specificity of 92.7 percent, and a positive predictive value for cancer of 2.5 percent. Long- term follow-up results from a University of Minnesota study showed an increase in more advanced lesions B and C in the control group, which translates to a decrease in mortality for the screened group. The cost- effectiveness of the test for occult blood as a screening tool is questionable. False-positive tests result in expensive evaluations to rule out real disease. False-negative tests result in patients presenting with a more advanced lesion that ultimately may cause more expensive, radical treatment.

Diagnosis Rectal Examination

The digital rectal examination can reach approximately 8 cm above the dentate line. It was long believed that 50 percent of colorectal cancers were within the reach of the finger. It has been shown that almost 50 percent of all colorectal cancers and polyps will occur proximal to the splenic flexure. The digital rectal examination remains an important tool in the detection of colorectal cancer and should not be abandoned. Twenty percent of colorectal cancers can be palpated and their degree of fixation to surrounding tissue evaluated.

Proctosigmoidoscopy

Rigid proctosigmoidoscopy uses a lighted tube 2 cm in diameter and 20 to 25 cm long. The examination usually reaches 20 to 25 cm from the dentate line, but the acute angulation of the rectosigmoid junction can prevent complete insertion of the instrument. This examination should detect 20 to 25 percent of colorectal tumors. Rigid proctosigmoidoscopy is safe and effective for screening low-risk adults under 40 years of age when combined with evaluation of the stool for occult blood. The endoscopist must remember to follow the contour of the rectum, advancing the scope only when the lumen can be seen. An enema is sometimes used to prep the patient before the examination.

Flexible Sigmoidoscopy

The flexible sigmoidoscope is a fiberoptic or video scope that measures 60 cm in length and should reach the proximal left colon or even the splenic flexure in most patients (and will identify 50 percent of colorectal cancers). The patient does not need a complete bowel prep because the left colon can be cleansed by one or two small enemas. The flexible sigmoidoscope should not be used for therapeutic polypectomy, cautery, and the like, except in special circumstances, such as after an ileorectal anastomosis. Polypectomy or biopsy using electrocautery in unprepped colon can be disastrous if an explosion of colonic methane gas occurs or perforation allows leakage of feces into the peritoneal cavity.

Flexible sigmoidoscopy every 5 years beginning at age 50 is the current endoscopic screening method recommended for asymptomatic persons at average risk for colorectal carcinoma. An adenomatous polyp found on flexible sigmoidoscopy is an indication for colonoscopy, because even small (<10 mm) adenomas in the distal colon are associated with aggressive neoplasms in the proximal colon in 6 to 10 percent of patients.

Contrast Studies

The most commonly used contrast study to detect colorectal cancer is the air-contrast barium enema. An air-contrast (double-contrast) barium enema has been found to be 90 percent sensitive in detecting polyps larger than 1 cm. The air-contrast barium enema, in conjunction with the flexible sigmoidoscope, has been suggested as a cost-effective alternative to colonoscopy for patients who do not tolerate colonoscopy or for use on an alternating basis with colonoscopy for long-term follow-up in patients who have had a previous polyp or cancer excised. The risk of perforation of the colon by barium enema is low (less than 0.02 percent). If perforation is a possibility, water-soluble contrast should be used rather than barium. Barium peritonitis is a disastrous complication that causes severe infection and subsequent intraperitoneal fibrosis. The patient usually requires an emergency operation to clear the barium from the abdomen and construction of a colostomy to protect the repair of the perforation. Water-soluble contrast material unfortunately does not provide the necessary detail for detecting small colonic mucosal lesions.

Colonoscopy

Colonoscopy allows visualization of the mucosa of the entire colon and rectum and usually the terminal ileum. Colonoscopy originally depended on flexible fiberoptics as the source of the imaging. New technology has allowed a video camera to be placed at the tip of the flexible colonoscope, allowing greater flexibility and better resolution of image. The standard colonoscope is 160 cm in length. Colonoscopy is the most accurate method of detecting polyps less than 1 cm in diameter. It also allows biopsy, polypectomy, control of hemorrhage, and dilation of strictures. Adequate colonoscopy includes good patient compliance, good bowel preparation, patient sedation in most cases, and monitoring of vital signs and blood oxygenation. Colonoscopy is a safe procedure, and major complications (including bleeding, anesthetic complications, and perforation) should occur in less than 0.2 percent of patients. Colonoscopy is useful in the diagnosis and management of inflammatory bowel disease, nonacute diverticulitis, sigmoid volvulus, gastrointestinal bleeding, nontoxic megacolon, colonic foreign body, colonic strictures, neoplastic disease, and unexplained diarrhea.

Complications after colonoscopy are more frequent after therapeutic colonoscopy (polypectomy, stricture dilation, control of hemorrhage) than after diagnostic colonoscopy. The most common complications of diagnostic or therapeutic colonoscopy are related to the use of conscious sedation. The most common major complication after therapeutic colonoscopy is hemorrhage after polypectomy (up to 3 percent). Postpolypectomy hemorrhage may occur immediately, in which case local measures are appropriate for control (epinephrine injection at bleeding site, electrocautery, resnaring polyp stalk). Management of delayed hemorrhage requires colonoscopy or angiography if the patient is bleeding massively. The source is usually arterial when the eschar separates from the polypectomy site. The majority of patients stop bleeding spontaneously, but colonoscopic measures or vasopressin via arteriogram catheter may be required if bleeding persists. Perforation is the most common major complication after diagnostic colonoscopy and is associated with free intraperitoneal air, abdominal pain, and variable degrees of peritoneal signs depending on the contamination of the abdominal cavity. A patient who develops free air without peritonitis and minimal contamination of the abdomen may be observed closely and treated with bowel rest, antibiotics, and intravenous fluids. The presence of (or progression to) diffuse peritonitis indicates the need for exploration. Depending on the size of the perforation, degree of contamination, and status of the patient, the patient may be treated with colectomy with primary anastomosis (plus or minus proximal diverting stoma), colectomy with colostomy and mucous fistula or Hartmann stump, or debridement and closure of the perforation.

Imaging Techniques

As technology progresses, more sophisticated imaging techniques are becoming available. CT, MRI, positron-emission tomographic (PET) scan, and transrectal ultrasonography are specific imaging techniques and not screening tests. They are important in the evaluation, staging, and follow- up of patients with colorectal cancer. They are used to evaluate primary cancers and their potential metastases to adjacent or distant organs.

Computed Tomography

Computed tomography (CT) allows preoperative evaluation of the abdominal cavity in patients with colorectal cancer. The CT scan can identify metastases to the liver, adrenal gland, ovary, lymph nodes, and other organs in the pelvis, and it can verify the integrity of the urinary tract because intravenous contrast material is excreted by the kidney. The CT scan is useful in searching for recurrent or residual cancer in patients with an elevated *carcinoembryonic antigen (CEA)* level after surgery for colorectal cancer. A combined angiogram and CT scan with dynamic imaging (scan during arterial injection of contrast material) is useful to evaluate the vascular anatomy of the liver and to accurately identify all possible metastases in the liver. The sensitivity of the angio-CT is approximately 95 percent.

The CT scan plays a special role in patients with rectal cancer because of the difficulty in staging the lesion preoperatively. Treatment options other than total rectal excision are available for selected patients with favorable lesions, as will be discussed in the section on rectal cancer. Sphincter- saving operations and local therapy may be appropriate if the tumor has not invaded deeply or metastasized to regional lymph nodes. Pelvic CT scanning documents rectal wall invasion by the tumor with a 60 to 90 percent accuracy and detects lymph nodes larger than 1 cm in 75 percent of patients.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a relatively new technique for evaluating colorectal cancer. Axial and sagittal views can be used to detail anatomy. T₁- and T₂-weighted images facilitate the differentiation of cancer from normal tissue and fibrotic scar. A coil placed in the rectum allows enlarged, detailed images of the rectum that may facilitate staging of rectal cancer. MRI is a lengthy study that requires the patient to lie quietly in a small tunnel, causing patients with claustrophobia to dislike the technique. The test does not require the use of contrast material. Early experience suggests that MRI offers little or no advantages over CT in the preoperative evaluation of patients with colorectal cancer. MRI is more specific for tumor in the liver than CT scanning and is frequently used to clarify a questionable lesion seen on CT scan. Because it is more sensitive than the CT scan, it is also used to identify other liver metastases when a liver resection is contemplated.

Positron-Emission Tomography

Positron-emission tomography (PET) is still investigational but eventually may be most helpful in evaluating recurrent tumor in the pelvis where scar formation is very dense. The PET scan detects differences in tissue metabolism using an injected substrate, fluorodeoxyglucose. Scar and recurrent tumor appear essentially the same on CT. The PET scan, however, will show an increase in metabolism in an area of cancer. Positron emission is greatest in hypermetabolic tissue such as tumors.

The major advantage of the PET scan may be to identify extrahepatic or intraperitoneal disease before performing a liver resection for metastasis or wide radical resection of local recurrence. This reduces unnecessary surgery and improves the selection of patients who will benefit from resection of recurrent or metastatic disease (Fig. [26-65](#)).

Transrectal Ultrasound

Transrectal ultrasonography or endosonography of the rectum is useful for staging rectal cancer. There is no radiation risk, and the procedure takes little time and can be directed to the upper rectum using a modified proctoscope. The ultrasound image allows clear delineation of the layers of the rectal wall and, as a result, a very accurate definition of the depth of invasion of the tumor into the bowel wall: Stage T1=limited to mucosa and submucosa; Stage T2=limited to muscularis propria; Stage T3=through the entire bowel wall (Fig.

26-66). The most important aspect of staging is the identification of those lesions which invade through the bowel wall into surrounding fat. Ultrasonography easily detects small lymph nodes, but the positive predictive value for cancer is only 50 percent for those under 1 cm. Larger lymph nodes (>1 cm) seen on ultrasonography have a 70 percent positive predictive value for cancer. It is difficult to perform endosonography in a patient with an obstructing lesion because the rectal probe cannot pass beyond the lesion. Endosonography of a rectal cancer also may reveal invasion of the tumor into adjacent prostate, sacrum, or vagina (Stage T4).

Endoluminal Ultrasound

Colonoscopic ultrasound evaluation of a colonic lesion is available. The small side-scanning transducer is placed on the colonoscope tip. The planar view allows determination of depth of invasion, attachment to adjacent structures, and spread to close mesenteric lymph nodes by tumors in the colon. The usefulness of such information is unclear at present because treatment is determined by the portion of the colon involved rather than depth of invasion.

Ultrasound-guided biopsy of the lymph nodes or extraluminal tumor in the pelvis is possible with newer ultrasound equipment. This has the benefit of documenting local recurrence of tumor in the pelvis in patients after previous sphincter-sparing resections of a rectal cancer. Patients who have tumor present in lymph nodes at the time of diagnosis of a rectal cancer are not candidates for local treatment.

Monoclonal Antibodies

Hybridoma techniques have allowed the development of a library of monoclonal antibodies to colorectal cancer. There are two types of monoclonal antibodies: one that recognizes CEA and the other that recognizes cell membrane features of colon cancer. Monoclonal antibodies tagged with a radioactive isotope allow tumor imaging. Therapeutic destruction of rectal or colonic cancer using this technology may be possible in the future.

Surveillance Screening

Screening must be differentiated from evaluating patients who are at high risk for developing colorectal cancer. Screening programs are aimed at the general population. Cost-effectiveness, reliability, ability to follow up on findings, and a reduced mortality in the screened population are factors to consider when evaluating a screening program. *Sensitivity* of the screening method is based on the *true positives* in the population and is often difficult to prove. *Specificity* is determined by the *true negatives*, which are defined by a negative test. This also is difficult to evaluate because lengthy follow-up is necessary.

Even though screening is an excellent idea, there are limitations. The ideal situation might indeed be to perform colonoscopy on the entire population and remove all polyps to prevent the occurrence of colorectal cancer. This obviously is impractical because of the inordinately high expense, tremendous amount of time involved, and risk of complications of colonoscopy. The medical community has defined *high-risk* groups and is focusing its attention on these groups in order to reduce the incidence of cancer and to improve the cost-benefit and risk-benefit ratios for its patients. Traits of high-risk individuals are listed in Table 26-14.

Recommendations

The summary of recommendations of the US Preventive Services Task Force (USPSTF) is as follows: Screening for colorectal cancer is recommended for all persons aged 50 and older with annual fecal occult blood testing (FOBT), or sigmoidoscopy (periodically unspecified), or both.

There is sufficient evidence to determine which of these screening methods is preferable or whether the combination of FOBT and sigmoidoscopy produces greater benefits than does either test alone.

There is also insufficient evidence to recommend for or against routine screening with digital rectal examination, barium enema, or colonoscopy, although recommendations against such screening in average-risk persons may be made on other grounds. Persons with a family history of hereditary syndromes associated with a high risk of colon cancer should be referred for diagnosis and management.

Colonoscopy is recommended for blood in the stool, a past history of adenomatous polyps, a positive family history of colorectal cancer, a history of ulcerative colitis or FAP, or an adenomatous polyp found on flexible or rigid sigmoidoscopy. The last recommendation is based on the fact that about 30 percent of patients with a rectosigmoid adenoma will have synchronous neoplasms elsewhere in the colon.

High-Risk Individuals

Previous Neoplasia

Metachronous neoplasms, polyps or cancer, have been reported in 40 to 50 percent of patients with a primary colorectal cancer. Cancer has developed in 2 to 5 percent of these patients. The risk of subsequent cancer in patients who have had an adenomatous polyp removed varies from 2.7 to 7.7 times that of the general population. The risk of cancer is increased by polyp number, polyp size, and the presence of villous architecture in the polyp. The National Polyp Study reported that 32 percent of patients who had adenomatous polyps removed via colonoscopic polypectomy were found to have metachronous neoplasms at colonoscopy 3 years later. In addition, colonoscopic polypectomy resulted in a lower-than-expected incidence of colorectal cancer in these patients, suggesting that routine surveillance may reduce the incidence of metachronous cancer. The influence of surveillance on survival is unknown. Patients who have had resection of a cancer or a polyp should have routine surveillance.

The following are guidelines for surveillance after resection of a colorectal cancer. These recommendations may vary depending on the number of polyps, the ease of the examination, and the age of the patient (Fig. 26-67). The plan can be summarized as follows:

- (1) Colonoscopy in the perioperative period to remove all synchronous polyps and detect synchronous cancers.
- (2) Colonoscopy 1 year after surgery to check for missed synchronous lesions and to detect any new lesions. Colonoscopy can be repeated yearly until no further lesions are identified.
- (3) Colonoscopy 3 years after a negative 1-year examination.
- (4) Colonoscopy every 5 years after a negative 3-year examination.

A similar surveillance strategy can be used after colonoscopic removal of adenomatous polyps, with the exception that the first follow-up colonoscopy can be performed 3 years after polypectomy.

Family History

There is a significantly increased risk of cancer in individuals with a family history of colorectal cancer. This risk depends on the number of first-degree relatives who have had cancer. This information should be used to begin earlier screening with colonoscopy of the group at risk. Patients who are members of families with FAP and Lynch syndromes also should have routine surveillance. The National Polyp Study also has demonstrated an increased risk of colorectal cancer in individuals whose first-degree relatives have adenomatous polyps.

It is very common for young patients (under 40 years) with colorectal cancer to have first-degree relatives with colon cancer (89 percent) and to be part of an HNPCC family (18 percent). Therefore, it is advisable for family members of young patients with colon cancer to be screened with colonoscopy at an early age (25 years old).

Hereditary Nonpolyposis Colon Cancer

HNPCC is an autosomal dominant disease that carries a 50 percent risk of cancer in family members. Metachronous colon cancer is also possible in affected individuals. The appropriate screening or surveillance protocol for these individuals is not known. Genetic testing for microsatellite instability as an indicator of errors in mismatch repair is not adequate because a large proportion of tumors in patients without HNPCC have microsatellite instability. The use of white blood cell mutational analysis in family members of patients with HNPCC is under investigation. Recommendations for individuals with HNPCC include total abdominal colectomy and ileorectal anastomosis with yearly screening of the residual rectum. Family members in HNPCC families should have colonoscopy every 2 years beginning at age 18 or at least 5 years before the age at diagnosis of the youngest family member with cancer. Mutational analysis eventually may be used to detect one of the at least four defects (MSH2, MLH1, PMS1, PMS2) in the mismatch repair genes on chromosomes 2 and 3 in individuals at risk for cancer. The need for prophylactic colectomy rather than continuous surveillance colonoscopy has not been established.

Ulcerative Colitis

Patients with ulcerative colitis for longer than 10 years should be examined by colonoscopy and biopsies for dysplasia, as discussed previously, with treatment based on the fact that the cancer risk in the setting of severe dysplasia is 30 percent (see Fig. 26-50).

Polyps

Hamartoma Peutz-Jeghers Syndrome

The combination of hamartomatous polyps of smooth muscles and mucocutaneous pigmentation was first described in 1896 by Hutchinson. The existence of *familial* hamartomatous polyps originally was described by Peutz and then again in 1949 in a follow-up article by Jeghers. The syndrome consists of pigmentation, produced by melanin in the mucocutaneous areas, and of hamartomas of intestinal smooth muscle throughout the gastrointestinal tract. This is an autosomal dominant syndrome and carries with it a well-documented risk of cancer in the breast, cervix, ovary, fallopian tubes, thyroid, lung, skin (basal cell), gallbladder, pancreas, testicle, and bile ducts. An overall risk of 2 to 13 percent for developing intestinal cancer arising in the stomach, small intestine, and colon has been documented in these patients.

Symptoms of the hamartomas include bleeding or intestinal obstruction secondary to intussusception. Treatment includes removing any polyp that is larger than 1.5 cm, even if asymptomatic. If a surgeon is forced to operate for symptoms, every effort should be made to completely clear the polyps with intraoperative endoscopy and polypectomy. Cancer, if present, is usually found in the mucosa overlying or adjacent to the hamartoma. After an operation to remove the polyps, colonoscopy is recommended every other year to survey for recurrence of colonic polyps. Patients also should undergo periodic screening for breast, cervix, testicle, ovary, upper gastrointestinal, and pancreatic cancers.

Juvenile Polyposis

Juvenile polyposis was first described by Verse in 1908. The familial inheritance of these polyps was reported by Veale at St. Mark's Hospital in 1966. Juvenile polyps occur as

cystic dilatations of glandular structures within the fibroblastic stroma of the lamina propria. Juvenile polyps usually cause symptoms of gastrointestinal bleeding or obstruction due to intussusception, with the polyp as the lead point. These complications require surgical or endoscopic removal of the polyp. There are varying patterns of occurrence of juvenile polyps. Approximately 1 percent of all infants have a single juvenile polyp discovered in infancy. A second peak incidence is seen at 25 years of age. The most common cause of gastrointestinal bleeding in a child is a juvenile polyp. There has been no malignant potential documented for these isolated juvenile polyps.

Isolated polyps are best treated with polypectomy, primarily for establishing the correct diagnosis. No other treatment is indicated. Autoamputation of the polyp often causes the initial bleeding. Once the bleeding has stopped, no further treatment is necessary.

Multiple polyposis coli (diffuse juvenile polyps) is an autosomal dominant syndrome with high penetrance that carries an increased risk of cancer in the gastrointestinal tract and in other nongastrointestinal tissues. The syndrome usually causes bleeding, obstruction due to intussusception, or malnutrition due to excessive protein loss. Diffuse juvenile polyposis occurs in less than 1 percent of patients with juvenile polyps and may carry an even higher risk of cancer when mixed or heterogeneous types of polyps are present (polyps of both adenomatous and juvenile types). Adenomas have been documented on juvenile polyps as well as in adjacent mucosa. The malignant potential has been estimated at approximately 10 percent in patients with the multiple polyposis coli type of juvenile polyposis and is even higher in patients with the diffuse variety. As in the other polyposis syndromes, the risk of cancer appears to increase with the duration of the disease and the age of the patient.

The multiple juvenile polyposis syndrome requires colonoscopic polypectomy in patients with a few colonic polyps because of the possibility of bleeding, intussusception, and a low but real incidence of adenomatous change. In patients with this condition, it is sometimes best to perform an abdominal colectomy with ileorectal anastomosis and to continue to screen the rectum. The entire colonic and rectal mucosa should be removed in patients with the diffuse form of juvenile polyposis and an ileal reservoir constructed with ileal pouch–anal anastomosis.

Hyperplastic Polyps

Hyperplastic polyps are common in the colon. They have been estimated to be 10 times more common than adenomas. Their typical histologic findings consist of epithelial dysmaturation and hyperplasia. They are usually less than 5 mm in diameter and show a thickened mucosa without atypia. There is usually no evidence of adenomatous change. They are associated with an increase in myofibrils and T cells in the lamina propria. The clinical significance of hyperplastic polyps continues to be the subject of some debate, but most studies indicate that they are not premalignant lesions and are not associated with adenomatous change elsewhere in the colon. *Serrated or mixed polyps* have a combination of hyperplasia and adenoma. These polyps have been associated with the development of cancer.

Hyperplastic polyps tend to increase in incidence with age. There is also a rare inherited syndrome of hyperplastic polyposis in which larger hyperplastic polyps occur in young adults. Hyperplastic polyps can be biopsied, destroyed, or removed during colonoscopy.

Other Lesions

There are several other less well-known and rare hamartomatous syndromes. *Cowden's syndrome* is an autosomal dominant defect that includes hamartomas of all three embryonal cell layers. Tricholemmomas, breast cancer, thyroid disease, and gastrointestinal polyps with occasional gastrointestinal cancers are typical of the syndrome. Treatment is based on symptoms. Polyps should be removed, and the patient should be screened for other cancers.

Cronkhite-Canada syndrome is typified by nonfamilial juvenile polyps with other epidermal changes but no evidence of malignancies. The polyps should be removed only if they cause symptoms.

The *Ruvalcaba-Myhre-Smith syndrome* includes developmental abnormalities, microcephaly, and juvenile polyps. It is a very rare syndrome that occurs in males, and to date, no cancers have been documented. The polyps should be removed when symptomatic, and other members of the family should be screened.

Adenoma Pathology

The *adenoma*, or *adenomatous polyp*, of the colon or rectum is a benign neoplasm. It represents proliferation and unrestricted cell division in a well- circumscribed area of glandular epithelium within the colonic mucosa. The surface is usually rounded and lumpy, like a raspberry, or flat and velvety with a shaggy appearance. Polyps can exist in a pedunculated (with a stalk), sessile (flat), or semisessile (raised) form. Each of these macroscopic forms may have similar histologic adenomatous appearances of the epithelium and the same cytology. The histologic architecture of the polyps may vary from branched tubular glands, to fingerlike, elongated villi, to villous configuration.

Polyps are classified as *tubular adenoma*, *tubulovillous adenoma*, or *villous adenoma* depending on the predominant pattern. In general, 65 to 80 percent of all adenomatous polyps removed are tubular adenomas, 10 to 25 percent are tubulovillous, and only 5 to 10 percent are pure villous adenomas. Tubular adenomas usually are pedunculated, and villous adenomas usually are sessile. However, all types of histologic characteristics can be found in each type of polyp. Incompletely differentiated epithelial cells are found along the fingerlike villi. The degree of differentiation decreases as the polyp becomes more like a cancer. The term *severe atypia* is now accepted for the finding of a malignancy in the epithelium of a polyp that has not invaded the muscularis mucosae of the bowel. Such lesions previously were referred to as *carcinoma in situ*.

Polyp-Cancer Sequence

It is generally accepted that a benign adenoma has been the precursor to a colorectal cancer in the majority of patients. However, Shimoda has shown that cancers also may arise de novo in nonadenomatous mucosa. These flat cancers usually are advanced and seem to have arisen as intramucosal cancers with no evidence of preceding polyp formation. This finding is consistent with a rapid progression of the genetic changes postulated by Fearon and Vogelstein as a polyp changes to a cancer (Fig. [26-68](#)).

The evidence for the existence of an adenoma-carcinoma sequence includes the following:

(1) Polypectomy has been shown to decrease the incidence of cancer. Patients with small adenomas have a 2.3 times increased risk of cancer after polypectomies as compared with an 8-fold increase in the risk of cancer in patients who do not undergo polypectomy.

(2) Colonic adenomas are observed more frequently in patients with cancer. Nearly 30 percent of all patients with a colorectal cancer also will have a synchronous polyp.

(3) Larger adenomas are found to contain cancer and severe dysplasia more often than are smaller polyps. Sixty percent of adenomas larger than 2 cm are villous adenomas. Villous adenomas have a 40 percent risk of containing a cancer, as compared to a 22 percent risk of cancer in tubulovillous adenomas and a 5 percent risk of cancer in tubular adenomas; however, the size of the polyp seems to be more important than the histologic features. The risk of a cancer in a villous adenoma, tubulovillous adenoma, or tubular adenoma larger than 2 cm is 53, 46, and 35 percent, respectively.

(4) Severe dysplasia has been shown to progress to cancer in polyps. The differentiation between carcinoma in situ and invasive cancer relies on the depth of invasion of the dysplastic area. Invasion beneath the muscularis mucosae defines invasive adenocarcinoma in a polyp because the cancer cells at this depth have access to the vascular and lymphatic systems.

(5) Residual adenomatous tissue is found in the majority of invasive cancers. This is presumptive evidence that the cancer has arisen within the polyp and has progressed to replace the majority of the polyp.

(6) Patients with familial adenomatous polyposis (>100 polyps) develop cancer 100 percent of the time if the colon is not removed. The adenomas of FAP are histologically the same as sporadic adenomas.

(7) There is a high prevalence of adenomas in populations with a high rate of cancer. This seems to indicate that the polyp is a precursor or a necessary factor in the development of cancer.

(8) It has been shown that the presence of an adenoma places a patient at a lifetime risk for the development of cancer.

(9) The peak incidence for the discovery of colonic polyps is at 50 years of age. The peak incidence for the development of cancer is at 60 years of age. This would indicate that there is a 10-year time span for the conversion of the adenomatous polyps to cancer. Investigators at the Mayo Clinic reported that a polyp larger than 1 cm has a cancer risk of 2.5 percent in 5 years, 8 percent in 10 years, and 24 percent in 20 years.

The adenoma-carcinoma sequence suggests that colonic mucosa progresses through stages to the final development of an invasive cancer. Normal epithelium is stimulated by some event to allow undifferentiated dividing cells to leave the crypts and move to the surface of the villi. Continued proliferation of these cells on the surface of the mucosa results in a polyp or adenoma. As proliferation continues, the glands begin to resemble normal tissue less and less. Eventually this disorganized growth crosses the muscularis mucosae barrier and becomes invasive cancer.

Treatment Pedunculated Polyps

Pedunculated polyps have a stalk that is usually less than 1.5 cm in diameter. Colonoscopic polypectomy is the best treatment for most pedunculated polyps. The cautery snare is placed at a point on the stalk to adequately remove the polyp with a clear margin from the head but distant enough to avoid thermal injury to the colonic wall at the base of the stalk. It is also important to avoid transmission of electric current via the head of the polyp to the opposite wall of the colon. Complete removal and retrieval of the polyp are essential for histologic diagnosis (Fig. [26-69](#)).

Semisessile Polyps

Semisessile polyps have a broad pedicle, larger than 1.5 cm. These polyps usually can be removed in piecemeal fashion if they do not have a malignant appearance. Bites of 0.5 to 1.5 cm are taken with the cautery snare until the stalk is seen. It is important to mark the area with 0.1 mL of India ink as a tattoo in the mucosa to facilitate repeat colonoscopic examination or surgical removal of the same site.

Sessile Polyps

Sessile polyps larger than 2 cm usually can be removed in piecemeal fashion also (Fig. 26-70). Small, 0.5- to 1.5-cm bites are taken to avoid burning the full thickness of bowel. A lower current also helps to avoid full-thickness burning. A “second look” colonoscopy and polypectomy is necessary if residual tissue is left behind as in the semisessile lesion.

Colectomy

Cancer in the head of a polyp may be treated by colonoscopic polypectomy (Fig. 26-71). Colectomy must be performed when the cancer within the head of the polyp has unfavorable features. Because the incidence of nodal metastases in these patients varies widely (1 to 25 percent in different series), various characteristics of excised polyps have been examined to improve the estimation of this risk, including level of invasion of the malignant process into the polyp; local invasion of lymphatic, vascular, and neural tissues; degree of differentiation of the malignant cells; and DNA ploidy. If the neoplasm is limited to the mucosa or invades only the head of a pedunculated polyp, the risk of lymph node metastasis is small (0 to 3 percent), and patients may be considered for close follow-up rather than resection. If the carcinoma reaches the base of the polyp, or if the lesion is sessile, then the risk of lymphatic spread may be as high as 10 to 27 percent. Intermediate lesions that invade only the neck or the stalk of the polyp are rarely associated with lymph node metastases unless the lesion is poorly differentiated or there is lymphovascular invasion. According to Nivatvongs, the finding of lymphatic or vascular invasion, poorly differentiated adenocarcinoma, or cancer close to the resection line of the stalk indicates the need for colonic resection. The risk of local recurrence is at least 10 percent, and the risk of spread of cancer to lymph nodes is approximately 10 percent in these patients. It should be noted that if a patient does not undergo resection, colonoscopic follow-up may not detect persistent or recurrent tumor because the lesion may reside outside the lumen of the bowel. Each case must be individualized, taking into consideration the life expectancy, operative risk, and personal philosophy of the patient.

A sessile polyp, by definition, has no stalk, and the submucosa is immediately adjacent to the muscularis propria. The invasion of the cancer through the muscularis mucosae into the submucosa puts the patient at high risk for local recurrence and spread of cancer to regional lymph nodes. A fold of adjacent mucosa is occasionally confused as the pedicle of a sessile polyp. Therefore, even if completely excised, the sessile polyp with an invasive carcinoma requires colectomy for complete cure. Laparoscopic colectomy may be a less invasive approach to managing polyps that are benign in appearance and have no obvious cancer present on biopsy analysis but are too large to be removed colonoscopically.

Villous Adenoma of the Rectum

Villous adenomas of the rectum are typically flat, velvety, and soft sessile polyps within the rectum. They occasionally present with watery diarrhea and hypokalemia. The finding of a large (>4 cm) villous adenoma in the rectum, with areas of induration, indicates a very high risk for cancer (90 percent). Transrectal ultrasonography of the villous lesion is helpful to determine the thickness of the lesion. Ultrasonography, however, is unable to distinguish between carcinoma in situ and invasive cancer. Because of the risk of sampling error in these large, exuberant lesions, the only adequate biopsy is complete excision, preferably transanally if possible. If cancer is present, the patient should be treated with appropriate resection. The transanal excision should include adequate peripheral margins for complete cure if cancer is present. It is helpful to inject the submucosal layer with epinephrine both to raise the lesion from the underlying muscle and to reduce blood loss. The defect should be closed with suture to provide primary healing and to mark the site. There is a high risk of recurrence if inadequate excision is performed.

Occasionally, a villous tumor will cover the entire mucosal surface of the distal rectum. If the polyp appears to be benign, a coloanal anastomosis after distal mucosal proctectomy and full-thickness proximal proctectomy are the preferred treatment. The coloanal anastomosis allows preservation of the anal sphincter and removes all the mucosa at risk in the rectum. It is important to document by frozen section the absence of cancer in this situation unless the tumor is well above the anal canal and there is very little risk of invasion into the levator or external sphincter muscles.

The use of a colonic J-pouch in selected individuals may improve their function after a coloanal anastomosis. The terminal 10 to 15 cm of the left colon are folded into a J, and the intervening septum is divided and the two limbs stapled together with a linear-cutting gastrointestinal anastomosing stapler introduced through a hole in the apex of the J. The coloanal anastomosis is then performed using an end-to-end intraluminal circular stapler or a hand-sewn anastomosis to the anal canal mucosa. A protecting loop ileostomy is usually placed if there is concern for the anastomotic integrity for any reason.

Thermal ablation of the villous adenoma has been used in some patients in whom local excision is not possible. The disadvantages of thermal ablation include the risk of undiagnosed cancer or cancer developing in the base of a repeatedly cauterized, scarred segment. If the villous adenoma recurs after several laser treatments, the patient should be referred for surgical treatment.

In the past, transsphincteric (York-Mason) and transsacral (Kraske) approaches to the rectum were used to remove villous adenomas and allow sphincter preservation. These methods are currently used infrequently because the coloanal anastomosis and low stapled anastomotic techniques allow preservation of anorectal function and appear to be more effective in treatment of the disease.

Colon Cancer

Incidence

Cancer of the colon and rectum is the most common cancer of the gastrointestinal tract. In women, colorectal cancer is second only to breast cancer as a cause of cancer-related death. In men, it is the third most common lethal cancer, preceded by carcinoma of the lung and prostate. The incidence of this malignant tumor increased in the United States during the first part of this century but has stabilized over the past four decades.

It is estimated that approximately 131,000 new cases of colorectal cancer (94,000 colon and 37,000 rectal) will develop in the United States in 1997, and approximately 58,300 patients will die of the disease. Rectal cancer is slightly more common in men, whereas there is a slight predominance of colon cancer in women. An American has approximately a 5 percent probability of developing colorectal cancer during a 70-year life span. Most cases of colorectal cancer are diagnosed in patients over the age of 50, and the incidence of the disease rises steadily after that age. Despite the clear relationship with aging, colorectal cancer is not strictly a disease of the elderly; between 6 and 8 percent of cases occur in individuals under the age of 40. The onset of familial and hereditary forms of the disease occurs at a much earlier age, typically around the third decade.

Signs and Symptoms

The signs and symptoms of colorectal cancer are varied and nonspecific. The symptoms that most often prompt patients to seek medical attention include rectal bleeding, a change in bowel habits, and abdominal pain. Whether a patient has any noticeable symptoms or what form the symptoms take depends somewhat on the location of the tumor and the extent of disease.

Subacute Presentation

Tumors in the *right colon* typically do not cause changes in bowel habits (although large, mucus-secreting tumors may cause diarrhea). Patients might notice dark or tarry stools, but more often these tumors cause truly occult bleeding, which is not detected by the patient. Such chronic blood loss may cause iron-deficiency anemia, with resulting fatigue, dizziness, or palpitations. When a postmenopausal woman or an adult man develops iron-deficiency anemia, colorectal cancer should be suspected and appropriate diagnostic studies performed. Because the bleeding associated with colon tumors tends to be intermittent, negative tests for occult blood in the feces do not rule out the presence of a large-bowel cancer.

Lower abdominal pain is more often associated with tumors located in the narrower left colon. The pain is of a cramping nature and may be relieved by bowel movements. These patients are more likely to notice a change in bowel habits and the passage of bright red blood.

Less common nonacute symptoms of large-bowel cancer include weight loss and fever. About 50 percent of patients who have other symptoms also report a loss of body weight, but weight loss is almost never the sole manifestation of a colorectal tumor. Fever is an unusual presenting symptom. Septicemia is rare but can occur with any stage of large-bowel tumor; *Streptococcus bovis* bacteremia is highly suggestive of colorectal carcinoma.

Rarely, colon cancer will present as the lead point of colonic intussusception in an adult. In all adults who are found to have obstruction or partial obstruction because of intussusception, a colonoscopy or air-contrast barium enema is warranted to rule out a colon cancer.

Acute Presentation

A significant proportion of patients are first seen with acute symptoms reflecting obstruction or perforation of the large bowel. Colonic obstruction is highly suggestive of cancer, particularly in older patients. Complete obstruction occurs in less than 10 percent of patients with colorectal cancer, but it is an emergency that requires immediate diagnosis and surgical treatment. Ideally, this surgical emergency could be avoided by early recognition of changes in bowel habits. Patients with complete obstruction complain of inability to pass flatus or feces, cramping abdominal pain, and abdominal distention. Examination reveals a distended, tympanitic abdomen; occasionally the obstructing tumor can be palpated as an abdominal mass.

If the obstruction is not relieved and the colon continues to distend, the pressure in the intestinal wall can exceed the capillary pressure, and oxygenated blood will not reach the bowel wall, resulting in ischemia and necrosis. In such situations the patient will complain of severe abdominal pain, and abdominal examination will reveal rebound tenderness and decreased or absent bowel sounds. If not treated immediately, the necrosis will progress to perforation with fecal peritonitis and sepsis.

The large bowel also can perforate at the tumor site, probably because a transmural tumor loses its blood supply and becomes necrotic. Such cases are easily mistaken for acute diverticulitis, and the inflammatory process may be confined to the site of the perforation; however, in some cases a perforation may not be contained, resulting in generalized peritonitis. In rare cases, perforation into an adjacent organ (usually the bladder or vagina) occurs and is manifested by pneumaturia, fecaluria, or feculent vaginal drainage.

Unfortunately, the first signs of colon cancer may be caused by metastatic disease. Massive liver metastases may cause pruritus and jaundice. Ascites, enlarged ovaries, and scattered deposits in the lungs detected by a chest radiograph can be caused by an otherwise asymptomatic colon cancer. Treatment is rarely beneficial in such advanced disease.

Diagnosis and Evaluation

In patients with symptoms suggestive of colon cancer, the definitive diagnosis is usually established by endoscopy (flexible sigmoidoscopy or colonoscopy) or barium enema. Other tests are necessary to evaluate the extent of disease and search for metastases.

Barium studies of the colon may demonstrate the primary lesion (Fig. 26-72) and reveal a synchronous cancer elsewhere in the colon (which occurs in 3 to 5 percent of patients). Dietary restrictions and cathartics usually are necessary to prepare the patient for barium enema. However, if acute colonic obstruction is suspected and plain films of the abdomen

reveal dilated colon, water-soluble contrast study may be obtained with no preparation, to reveal the location and nature of the obstructing lesion.

Colonoscopy is currently the most accurate and complete examination of the large bowel. Colonoscopy and barium enema studies should be regarded as complementary to each other. The purpose of a *complete* colon and rectal evaluation for patients with large-bowel cancer is to rule out synchronous carcinomas and polyps.

A chest x-ray should be taken routinely to rule out pulmonary metastasis. CT of the abdomen is being used increasingly to assess the extent of invasion of the primary tumor and to search for intraabdominal metastatic disease.

Elevated liver function test results, including alkaline phosphatase, lactate dehydrogenase, bilirubin, and transaminase enzyme determinations, can suggest liver metastases. These tests are useful in the postoperative follow-up and, if elevated, can suggest the need for CT scan to search for hepatic metastases.

The serum level of CEA is important in the evaluation of patients with colorectal cancer. This antigen, a glycoprotein, is present in embryonic and fetal tissue and in colorectal cancers but is absent in normal adult colonic mucosa. In patients with cancer that has not penetrated the bowel wall, the serum CEA level is usually not elevated. The test is somewhat nonspecific, since it can be elevated in patients with tumors of the lung, breast, stomach, or pancreas. Levels also may be elevated in smokers and in patients with cirrhosis, pancreatitis, renal failure, and ulcerative colitis. Thus the test is not useful for screening.

An elevated CEA level does have a high correlation with tumor recurrence and with the presence of metastases from colorectal cancer. If the CEA level is elevated before operation, the recurrence rate is higher, regardless of the disease stage. Postoperative CEA levels may indicate the completeness of surgical resection or the presence of occult metastases. A rising CEA titer after treatment may indicate a recurrence, before clinical disease becomes evident. When used alone, the CEA assay is 70 percent accurate in predicting the development of liver metastases within 1 year; in combination with CT scans, accuracy increases to up to 90 percent.

Surgical Treatment Preoperative Bowel Preparation

Wound infection and intraabdominal abscess formerly were common problems following colorectal surgery. The primary source of infection was the endogenous bacteria in the bowel lumen, with *E. coli* being the most common aerobic species and *Bacteroides fragilis* the most common anaerobic organism involved in septic complications. The large bowel cannot be sterilized prior to surgery, and colorectal operations are classified as “clean-contaminated” procedures by the American National Research Council; operations during which the large bowel is opened virtually always expose the operative field to colonic bacteria. Accordingly, steps must be taken before the operation to reduce the bacterial population as much as possible. The preoperative preparation generally includes two components: *mechanical cleansing* and *antibiotic administration*.

There is no absolute consensus regarding appropriate mechanical bowel preparation. A decade ago the most common method was a 3-day regimen consisting of a low-residue or liquid diet combined with laxatives (magnesium citrate or castor oil) and enemas. This approach was time- consuming, was often hampered by incomplete patient compliance, and severely restricted the patient's caloric intake for 3 days before major surgery. Other approaches are used more commonly today, including mono- and dibasic sodium phosphate purgatives and whole-gut lavage with polyethylene glycol (Golytely or Colyte).

A commonly used bowel preparation today is an isotonic lavage solution containing polyethylene glycol (PEG) in a balanced salt solution. PEG also acts as an osmotic purgative, but it does not serve as a culture medium for bacteria and is not associated with the production of explosive colonic gas. With this method of cleansing, the patient may have a light meal at lunch and clear liquids the evening before surgery. To achieve adequate cleansing of the colon, an average of 4 L of PEG solution must be ingested within 4 h. The solution can be drunk or instilled through a small nasogastric tube. This method has been associated with less fluid retention than saline lavage but is still somewhat uncomfortable for the patient. Metoclopramide may be given one-half hour before starting the lavage to reduce nausea.

Mono- and dibasic sodium phosphate acts as a purgative when given in adequate volume (45 mL) and accompanied by large volumes (24 oz) of clear liquids. Two doses (noon and 6 p.m.) the day prior to surgery are adequate to achieve a complete mechanical bowel preparation. Patients with renal failure or severe hypertension on sodium restriction should be given this preparation cautiously because hypocalcemia, hyperphosphatemia, hypernatremia, and acidosis may occur. Patients with severe congenital megacolon may develop hypernatremic dehydration. Fleet's phospho soda and clear liquids has become a preferred alternative to whole-gut irrigation and is better tolerated than high-volume preparations, with equal results in bowel cleansing.

Mechanical bowel preparation will reduce the absolute number of colonic bacteria, but the bacterial concentration in the lumen seems unaffected, and the postoperative wound infection rate without antimicrobial use remains unacceptably high (between 30 and 60 percent). Reduction of colonic bacterial concentration can be achieved by oral or intravenous antibiotic preparations, and there is no consensus concerning the ideal preparation. Most surgeons use a combination of oral antibiotics (neomycin plus erythromycin, 1 g each at 1 p.m., 2 p.m., and 11 p.m. the day before operation or 0.5 g of neomycin instead of erythromycin) and an intravenous broad-spectrum antibiotic administered immediately before the operation. The timing of the administration of antibiotics is important; the postoperative infection rate can be reduced to well below 10 percent with proper bowel cleansing and preoperative antibiotic administration, but postoperative antibiotics do not appear to be effective.

Operative Technique: Colon Carcinoma

The objective in the treatment of carcinoma of the colon is to remove the cancerous segment of bowel, the mesentery containing its lymphatic drainage, and any organ that has been directly invaded by the tumor. Since lymphatic channels of the colon accompany the main arterial supply, the length of the bowel resected depends on which vessels are supplying the segment containing the cancer (Fig. 26-73). Surgical techniques that minimize the risk of intraoperative spillage of tumor cells are essential. The arterial supply to the uninvolved intestine must be preserved so that intestinal continuity can be restored by anastomosis.

During the operation, the peritoneal cavity is carefully explored, and the liver is examined to identify any unsuspected metastases. The intestine is encircled with ties, proximal and distal to the tumor, to prevent intraluminal spread of cancer cells during manipulation of the bowel. The main segmental artery supplying the cancerous segment of bowel is ligated and divided, allowing resection of the mesentery in continuity with the bowel. If the tumor is adherent to or invading adjacent organs such as the small bowel, ovaries, uterus, or kidney, an en bloc resection should be performed if technically feasible. Adhesions between the tumor and contiguous organs are often inflammatory, but this cannot be determined before resection. If the adhesions prove to be cancerous and they are merely divided instead of removed by an en bloc resection, the chance for cure will be lost. If all tumor cannot be removed, a palliative colon resection is generally indicated to relieve symptoms and prevent future obstruction and bleeding from the primary tumor. In the rare situation of a nonresectable tumor, a bypass operation should be considered to relieve or prevent intestinal obstruction. If multiple colon carcinomas are present, or if a colon carcinoma is associated with multiple neoplastic polyps, a subtotal colectomy (total abdominal colectomy) with ileorectal anastomosis should be considered.

Carcinoma of the right colon is treated by right hemicolectomy, which includes excision of approximately 10 cm of terminal ileum. The ileocolic artery, right colic artery, and right branch of the middle colic artery are ligated and divided, and an anastomosis is fashioned between the ileum and the left transverse colon. The anastomosis can be either hand-sewn or achieved by stapling instruments.

Carcinoma of the hepatic flexure or the right transverse colon is treated by extending the hemicolectomy to include the middle colic artery along with its left branch (see Fig. 26-73).

Carcinoma of the midtransverse colon is treated by extending the hemicolectomy even farther and performing the anastomosis between the ileum and the proximal descending colon. An alternative, in this situation, is to ligate only the middle colic artery and excise the transverse colon, making an anastomosis between the ascending and descending colon.

In a similar fashion, the surgical treatment of left colon cancer incorporates resection of the arteries and accompanying mesentery, as shown in Fig. 26-74.

The resected specimen should be inspected immediately by a pathologist to assess the surgical margins. A 5-cm margin of normal bowel on either side of the tumor is ideal; however, margins as small as 2 cm from the tumor edge may be satisfactory if adequate mesentery is resected with the specimen. Wider margins are required for a poorly differentiated or anaplastic carcinoma. While the anastomosis can be fashioned by either hand-sewn or stapling techniques, it is essential that the bowel edges be meticulously approximated to form a lumen of adequate size, that there be no tension on the anastomosis, and that the blood supply to both sides of the anastomosis is adequate.

The functional results after colectomy vary according to the segment of colon resected. The right colon absorbs approximately 1000 mL of fluid delivered from the terminal ileum. After a right colectomy, the stool generally is softer, and 750 mL/day is passed rather than 250 mL. A left colectomy or sigmoid colectomy generally results in several small stools per day but the same volume as before surgery. Patients who require a total abdominal colectomy with ileorectal anastomosis will experience 4 to 10 bowel movements per day immediately after their operation from 1000 mL of ileal contents entering the rectum during a 24-h period. Over time this will decrease as the ileum compensates for loss of the right colon and begins absorbing bile salts and more water. Younger patients tolerate removal of the colon better than the elderly, but this option should not be withheld from older patients.

Laparoscopic colectomy has been described for all types of colorectal disease. The first cases were reported in 1991, and since that time, numerous reports have entered the literature. Several large series have documented the safety and potential efficacy of this technique in the management of colorectal disease, including cancer. Randomized control trials have not confirmed these findings, and the use of laparoscopic colectomy in the treatment of cancer is controversial. Laparoscopic techniques such as laparoscopic segmental colectomy, abdominoperineal resection, rectopexy, ostomy construction and takedown, and small-bowel resection can provide earlier recovery (shorter postoperative ileus and hospital stay) and less pain for most patients regardless of the reason for operation. Laparoscopic colectomy requires advanced skill, thorough knowledge of intraabdominal anatomy, and an ability to recognize structures from unusual angles and perspectives. The same principles of surgical technique that guarantee a good outcome for open operations apply to laparoscopic procedures.

Emergency Operations

Obstruction

When a colon or rectal cancer causes complete obstruction, immediate surgical treatment is necessary. Obstructing cancer of the right or transverse colon usually can be treated by resection and primary anastomosis. The entire colon proximal to the obstruction should be resected, and the terminal ileum then should be anastomosed to the remaining normal colon. In such cases, diversion by proximal ileostomy is rarely needed.

Obstruction by a left-sided cancer presents a more difficult problem. If the obstruction is only partial, with feces and gas present in the rectum, it may be possible to proceed with preoperative cleansing of the colon as described earlier. However, when radiography reveals a distended colon with no air in the rectum, the obstruction is complete, and emergency surgical treatment is mandatory. Water-soluble contrast should be performed to confirm the presence of an obstruction and to determine its exact location.

The appropriate surgical procedure depends on the location of the obstruction, the initial surgical findings, and the surgeon's experience and judgment. The current trend is toward

immediate resection of the lesion. However, if the colon is extensively dilated and the patient's condition unstable, it may be appropriate to relieve the obstruction by a colostomy placed proximal to the lesion (described in the section on rectal cancer). After the patient's condition has stabilized, the bowel can be properly prepared and the obstructing lesion resected at a later time. If the surgeon feels that it is possible to resect the lesion, several options are available, including the following:

(1) Primary resection without anastomosis, bringing the proximal end of the colon to the abdominal wall as a colostomy and the distal end to a separate location on the abdominal wall as a mucous fistula. If the tumor is located so far distally that the distal end of colon cannot be exteriorized, the Hartmann procedure is used.

(2) Primary resection with anastomosis. Anastomosis between two segments of unprepared bowel is associated with an anastomotic dehiscence rate above 20 percent. There has been interest in cleansing the proximal bowel at the time of emergency celiotomy by the technique of intraoperative colonic lavage. A catheter is inserted into the cecum through an appendicostomy, and several liters of saline solution are instilled. The solution washes out the bowel contents through a large-bore catheter (or sterile anesthesia tubing) that has been placed in the colon proximal to the cancer. The cancerous segment of bowel, including the site of insertion of the evacuating tubing, is then excised, and an anastomosis is fashioned using the prepared, cleansed colon.

(3) Primary resection with anastomosis and proximal protective colostomy or ileostomy.

(4) Subtotal colectomy with ileosigmoidostomy. This approach is particularly suited for treating obstructing lesions of the sigmoid colon. The procedure requires more dissection and takes longer than segmental resection, but it eliminates the problem of synchronous polyps or cancers and removes the entire obstructed distended segment. In addition, the excellent blood supply of the ileum helps to ensure that the anastomosis is secure and reduces the risk of anastomotic leakage.

Perforation

Cancer causing a perforation of the colonic wall represents a life- threatening surgical emergency. The diagnosis may not be obvious before celiotomy because the clinical picture can mimic a perforated peptic ulcer or diverticulitis with generalized peritonitis. Thorough exploration of the peritoneal cavity is mandatory. Often, it can be difficult to distinguish an inflammatory pelvic mass caused by diverticulitis from that caused by a perforated sigmoid colon cancer.

There are fewer surgical options for treating a perforated colon cancer than for an obstructing cancer. The goal of operation is to remove the diseased, perforated segment of bowel. It may be possible to fashion an anastomosis; however, an anastomosis of unprepared bowel fashioned in a contaminated field should always be protected by proximal colostomy or ileostomy. The temporary diverting stoma can be closed about 10 weeks after the emergency operation.

An alternative is to resect the perforated segment and exteriorize the proximal and distal limbs as described above (or to use Hartmann's operation for more distal lesions). After the patient has recovered from the initial surgery, intestinal continuity can be restored.

After the perforated segment has been resected, the peritoneal cavity should be thoroughly irrigated with saline solution to remove all visible contamination, and intravenous antibiotics should be continued postoperatively.

Every effort should be made to perform a curative operation for obstructing or perforated cancers. The prognosis is somewhat poorer for these patients, but the 5-year survival rate approaches 30 percent in patients with no obvious metastases who are treated by immediate resection of the lesion.

Staging and Prognosis

In the absence of obvious metastatic disease, the precise stage of colorectal cancer can be determined only after surgical resection and histopathologic analysis. Unlike other solid tumors, the size of the primary lesion has little influence on prognosis of colon cancer. The factors that are most closely related to outcome are (1) the depth of tumor penetration into the bowel wall, (2) the involvement of regional lymph nodes, and (3) the presence of distant metastases. Numerous staging systems have been developed over the past several decades, but the system proposed by Dukes in 1932 for description of rectal cancer has continued to be used widely because of its simplicity (Fig. [26-75](#)). The classification scheme proposed by Dukes concerned only rectal cancer and recognized three stages; however, the most often used variant of this system recognizes a fourth stage (distant metastases) and has been extended to include colon cancers. This system fails to include important prognostic information such as vascular invasion, perineural invasion, histologic differentiation, or DNA content of tumor cells. Nevertheless, the simplicity of the Dukes classification system and its consistent correlation with prognosis suggest that it will continue to be in use for some time.

The American College of Surgeons' Commission on Cancer has urged adoption of the TNM (tumor-node-metastasis) staging system (Fig. [26- 76](#)). This system separately identifies the depth of invasion of the tumor (T), the regional lymph node status (N), and the presence of distant metastases (M):

Primary Tumor (T)

TX—Primary tumor cannot be assessed.

T0—No evidence of primary tumor.

T1—Tumor invades submucosa.

T2—Tumor invades muscularis propria.

T3—Tumor invades through the muscularis propria into the subserosa or into nonperitonealized pericolic or perirectal tissues.

T4—Tumor perforates the visceral peritoneum or directly invades other organs or structures.

Regional Lymph Nodes (N)

NX—Regional lymph nodes cannot be assessed.

N0—No regional lymph node metastasis.

N1—Metastasis in 1 to 3 pericolic or perirectal lymph nodes.

N2—Metastasis in 4 or more pericolic or perirectal lymph nodes.

N3—Metastasis in any lymph node along the course of a named vascular trunk.

Distant Metastasis (M)

MX—Presence of distant metastasis cannot be assessed.

M0—No distant metastasis.

M1—Distant metastasis.

The stage of the TNM system is as follows:

The TNM system provides more detailed pathologic information and can be converted easily to the simpler Dukes system. Stage I is equivalent to Dukes' Stage A, Stage II to Dukes' Stage B, Stage III to Dukes' Stage C, and Stage IV to Dukes' Stage D.

The 5-year survival rates clearly demonstrate the prognostic importance of disease staging. The best outcome is associated with Stage I disease, with more than 90 percent of patients surviving at 5 years. For Stage II disease, this figure drops to between 60 and 80 percent, and for patients with lymph node metastasis (Stage III), 5-year survival rates range from 20 to 50 percent. When distant metastases are present, fewer than 5 percent of patients survive for 5 years.

For all but Stage IV disease, survival rates have improved since the 1940s and 1950s. Whether this improvement is attributable to diagnostic or treatment advances is not clear. More thorough and precise staging procedures may now be responsible for eliminating or reducing the misclassification of patients, which can artificially and adversely affect the observed response to treatment.

Other disease factors that have a negative prognostic influence include poor histologic differentiation of the tumor (mucin-producing tumors and “signet ring” cell tumors have a poorer prognosis) (Fig. [26-77](#)), venous or perineural invasion by the cancer, bowel perforation, elevated CEA level, and aneuploid nuclei.

Adjuvant Chemotherapy

Colorectal carcinoma has been resistant to most chemotherapeutic agents. However, chemotherapy following extirpation of all detectable disease is theoretically appealing because effective agents should have an advantage if the tumor burden is minimal. Chemotherapy is most effective when the burden of carcinoma is smallest and the fraction of malignant cells in growth phase is highest. Recently, there have been indications that the combination of 5-fluorouracil (5-FU) and levamisole or 5-FU and leucovorin is associated with increased survival when administered postoperatively to selected patients with no apparent residual disease (*adjuvant chemotherapy*).

A large intergroup clinical trial examined the effect of levamisole alone, levamisole with 5-FU, and observation only in patients with resected colon carcinoma. Levamisole alone produced no significant effect, but the combination of 5-FU and levamisole improved disease-free survival rate and overall survival rates compared with surgery alone in patients with Stage III colon cancer. Therapy with 5-FU and levamisole reduced the risk of cancer recurrence by 39 percent, cancer-related deaths by 32 percent, and the overall death rate by 31 percent in these patients. There was no survival benefit in patients with Stage II disease who received chemotherapy. A combination of 5-FU and leucovorin has improved

survival rates in patients with Stage III colon cancer.

The role of adjuvant chemotherapy for rectal cancer is less well defined. Variation in surgical technique, particularly with regard to the extent of mesorectal excision and the use of pre- or postoperative radiation, makes comparison of available data difficult. The few prospective, randomized trials evaluating postoperative adjuvant chemotherapy suggest that there may be a modest survival advantage in treating Stage II and Stage III rectal cancer.

Long-Term Follow-Up

There is some controversy about the frequency of follow-up examinations for tumor recurrence in patients treated for colorectal cancer. Some physicians have adopted a nihilistic approach (because of poor prognosis once recurrent cancer is detected) and recommend that once a patient is discharged following surgery, further treatment should be given only if symptoms of recurrent disease develop.

A small group of patients, however, definitely benefit from detection and aggressive treatment of recurrent cancer. About 70 percent of recurrent cancers become detectable within 2 years of initial treatment and 90 percent within 4 years. Patients who have been treated successfully for colon or rectal cancer have a higher incidence of metachronous colorectal cancer, and early detection and treatment of these cancers can improve the prognosis for these patients. This fact alone should provide sufficient rationale for long-term follow-up of all patients treated for colorectal cancer.

Follow-up evaluation may include physical examination, sigmoidoscopy, colonoscopy, liver function tests, CEA assay, chest radiography, barium enema, liver scan, CT, and MRI.

Physical examinations following surgical treatment for colorectal cancer may assuage both patient and physician, but they are of little value in the early detection of recurrent disease. By the time a recurrent cancer has become palpable by abdominal examination, it is invariably unresectable.

Sigmoidoscopy is of some value when following patients treated with low anterior resection and anastomosis for rectal or sigmoid cancer because it allows direct inspection of the anastomosis. *Colonoscopy* is most useful for the detection of metachronous polyps or cancers in patients with previously treated colorectal cancer.

Both *liver function tests* and *CEA assay* can indicate the presence of liver metastases from colorectal cancer. A significant elevation of any of the liver function tests or a rising CEA level calls for a CT scan of the abdomen and a chest radiograph.

The principles of managing metastatic disease are similar for cancer of the colon and the rectum and are discussed in the following section on rectal cancer.

Rectal Cancer

Pathophysiology

As with colon cancer, the vast majority of malignant neoplasms of the rectum are adenocarcinomas. The significant premalignant conditions such as adenomatous and villous polyps, familial adenomatous polyposis, and ulcerative colitis are also the same. Adenocarcinomas of the rectum must be differentiated from squamous cell cancer of the anal canal because the management is entirely different. Malignant melanoma rarely can be found in the rectum, and its management again differs from that of adenocarcinoma. The prognosis of rectal malignant melanoma is so poor that wide resective surgery is not always indicated.

Rectal cancer is considered separately from colon cancer because of its anatomic location and configuration. Because the rectum resides within the confines of the pelvis, wide excision of the cancer and surrounding structures (as is possible with colon cancer) is usually impossible. The proximity of the rectum to the anal sphincter mechanism also creates unique problems of management. In order to achieve a curative distal margin beyond the cancer, a surgeon must consider the risk of injury to the sphincter mechanism. If the sphincter cannot be preserved to function perfectly, it should be removed. It is this need to remove the sphincter mechanism that results in the need for a colostomy, and it is wrong to jeopardize cure of a cancer to avoid a colostomy. The details of colostomy construction and care will be described later, but it is critically important for the patient to understand that a colostomy is usually constructed because of the anatomic location of the cancer, not its severity. Most patients assume that the need for a colostomy indicates a hopelessly incurable cancer.

Rectal cancer also requires special consideration because the innervation of the urogenital system lies in close proximity, as in fact do the organs of the urogenital tract (see Fig. 26-8). There is a high rate of impotence in men from radical treatment of rectal cancer, and we are reminded that accurate definitions of sexual dysfunction in women are lacking. The relationship of the rectum to the genitourinary system can be of supreme importance in determining the surgical procedure itself; the posterior wall of the vagina can be resected in continuity with an invasive rectal cancer, but a cancer invading the prostate or bladder base in a man requires a much more extensive and complicated procedure. The rectum also requires special consideration because it has a dual arterial blood supply (see Fig. 26-6), and the lymphatic channels of cancer metastasis follow the routes of the arterial system (see Fig. 26-7). The primary physiologic function of the rectum is to serve as a reservoir for fecal material, which can be stored until the individual chooses to empty it. Any disturbance of the reservoir or continence mechanism can result in significant handicap to the individual. This consideration is of prime importance in determining the treatment of rectal cancer. It is because of the distensible nature of the rectal reservoir that cancers can grow to considerable size before causing any symptoms. This fact makes screening for rectal, as well as colon, cancers mandatory. The organ is accessible, and asymptomatic (more curable) cancers should be sought by routine screening or surveillance, especially in individuals known to be at significant risk.

Perhaps the most unique aspect of the anatomy of the rectum is its easy accessibility. The anus allows access to the rectum with several therapeutic and diagnostic modalities, including the physician's finger. This easy access has resulted in consideration of tumor treatment less than the radical resective procedures that often require removal of the rectum and anus, necessitating construction of a colostomy.

These unique features of the rectum must be considered for the screening, diagnosis, and treatment of patients known to have or suspected of having rectal cancer. One must consider the *local* factors related to the rectum itself. The next level of consideration involves the relationship of the cancer to the structures *adjacent* to the rectum such as lymph nodes, sphincter mechanism, and surrounding organs. Perhaps the most important prognostic consideration lies in the status of the cancer with respect to *distant* metastasis. Rectal cancer can spread through the lymphatics of the mesenteric or iliac system. It can metastasize to the liver via the portal venous system or to the lungs, or rarely to the brain or bones, via other hematogenous routes.

Diagnosis and Classification

The important aspects of complete evaluation of rectal cancer are related to the *local*, *regional*, and *systemic* manifestation of the cancer (Fig. 26-78).

Local Aspects

Local evaluation of a rectal cancer helps determine the proper modality of management and gives some insight into the prognosis of the cancer. Careful digital rectal examination must be performed to evaluate size, fixation, and ulceration of the cancer, as well as any suggestion of extension of the cancer to pararectal lymph nodes or adjacent organs. The rectal cancer needs to be visualized by the surgeon using a rigid sigmoidoscope, adequate biopsy taken, degree of fixation to surrounding tissue evaluated, risk of obstruction predicted, size and ulceration of lesion determined, and distance from the distal edge of the tumor to the dentate line measured carefully. Flexible fiberoptic instruments are not totally reliable because only tiny biopsies can be taken, and the flexibility of the instrument can give a false sense of security relative to the distance between the cancer and the mucocutaneous junction (dentate line). It is this distance that ultimately determines if a surgeon can preserve the anal sphincter mechanism.

Cancers that have an adequate "clinical" classification system allow the treating physicians to securely plan a course of treatment before operation. So far this cannot be fully accomplished with rectal cancer because the important variables either have not been defined completely or cannot be determined before initiating treatment. The previously noted classification systems (see Figs. 26-75 and 26-76) describe the depth of cancer invasion and the status of cancer spread to surrounding lymph nodes. This is a "histologic" classification that requires removal of the rectum and pararectal tissue before its application.

The biologic behavior of rectal cancer cannot be predicted by its location or size. Poorly differentiated cancers have a worse long-term prognosis than those which are well differentiated. Some very aggressive histologic patterns such as "signet ring" cell and mucinous cancers carry an especially poor prognosis. Attempts are being made to further predict the biologic aggressiveness of rectal cancers by analysis of the DNA from biopsies of the cancers (see Fig. 26-77). These biologic predictors become important when local, nonresective techniques are described for small, favorable-appearing cancers. If the cancers that look favorable but have aggressive biologic behavior can be identified, they should be treated by nothing less than radical resection.

The *depth of invasion* of a rectal cancer is known to be an important variable. In order for a rectal cancer to be classified as "invasive," it must extend at least into the muscularis mucosae, for at this level it has access to the vascular and lymphatic systems, and metastasis becomes a possibility. Morson showed long ago that even a small, mobile cancer, confined to the rectal wall but extending into the muscularis mucosae, will have a 10 percent incidence of metastasis to pararectal lymph nodes. A major, unfulfilled challenge in evaluating rectal cancer is the pretreatment determination of lymph node involvement by cancer.

The depth of cancer invasion can be determined accurately by use of *transanal ultrasonography* (see Fig. 26-65). Each layer of the rectal wall can be identified and its penetration by cancer described. Although lymph nodes can be visualized by ultrasonography, the presence of cancer within these nodes cannot be accurately predicted. CT scanning of the rectum has been extremely helpful in evaluation of the cancer itself, but nuclear magnetic resonance imaging, especially with the use of newly available transrectally positioned coils, should be very helpful in predicting depth of invasion and involvement of lymph nodes and other pararectal structures by the primary cancer. Sometimes transanal excisional biopsy of an entire small lesion is helpful for determining depth of invasion, but this will not define lymph node involvement.

The advantages of these accurate descriptions of the cancer become evident when the treatment is chosen and decisions must be made regarding extent of operation and use of adjuvant modalities such as chemotherapy or radiation therapy.

Regional Aspects

Regional evaluation of pararectal structures involves transanal ultrasound and CT scanning or MRI, as just described, plus some special considerations. A major factor early in the evaluation of a patient with rectal cancer involves determination of the possibility and extent of resectability. After a cancer grows through the wall of the rectum, it can invade any of the surrounding structures. This is usually determined by evaluation of the *fixation* or tethering of the tumor to these structures. This makes complete examination of these structures necessary. A woman should have a complete pelvic examination with care to determine if the tumor is suspected of invading the vagina or has spread to the ovaries. In a man, evaluation of extension into the prostate or bladder is important. This could require transanal ultrasonography, CT scanning, MRI, or even cystoscopy and biopsy.

It is suspected that the extent of fixation of the tumor is an important predictor of prognosis. It is also a major factor in determining the need for adjuvant radiation therapy, sometimes coupled with chemotherapy. Even if the cancer shrinks with preoperative adjuvant therapy, consideration must be given to resection of the involved structures originally suspected of invasion by the cancer. The performance of very extensive resective procedures must be tempered by the patient's age and general health. The determination of extent of fixation of the cancer is complicated by the fact that many invasive cancers are surrounded by an intense inflammatory reaction that creates the sensation of fixation rather than actual invasion by the cancer.

The surgeon also should know the status of the upper urinary tract. If, for example, one of the ureters is involved with the invasive cancer, a determination should be made before operation that the contralateral kidney is functioning. Specific plans may need to be made for a urologist to deal with the bladder and ureter at the time of surgery. Our own preference is to use intravenous contrast material at the time of CT scanning for liver metastases to evaluate kidney size and function, as well as position and possible involvement of the ureters.

One of the most frustrating aspects of regional evaluation of rectal cancer is the inability to determine cancer involvement of the adjacent, or regional, lymph nodes. This one fact would be most helpful in determining need for adjuvant treatment preoperatively. CT and transanal ultrasound, as well as distal palpation, have allowed determination of the presence of enlarged nodes but have not been helpful in determining presence of cancer within these nodes. There is hope that improved MRI techniques, use of monoclonal antibodies, and perhaps PET scanning will be helpful in this important aspect of staging. Most studies have shown the involvement of lymph nodes in the resected specimen to be the most influential predictor of prognosis. Recently, more credence is being placed on the involvement of the "tangential" margin of resection (Fig. 26-79). This is the peripheral margin, including the extent of clearance of the cancer from surrounding tissue. Interestingly, this margin is not reported by the pathologist unless special effort is made to "ink" the tissue immediately after the resection. A great deal of discussion exists in the surgical literature over what represents an adequate distal margin, the range being 2 to 5 cm. This is based on studies to evaluate the extent of microscopic spread of cancer distal to the palpable edge of the tumor. What difference does it make if the distal margin is 2 or 5 cm when the tangential margin is minimal or nonexistent? Many surgeons believe that this nonreported regional margin is the major cause of local recurrence of rectal cancer. Although the cure of rectal cancer can be very high, the cure of locally recurrent rectal cancer is exceedingly small and often requires devastatingly radical surgery. Thus the complicated evaluation of regional spread is critical.

Distant Spread

Determination of distant spread of rectal cancer is important because these patients generally are considered to be incurable. The knowledge of metastasis to the liver or lungs may significantly modify the surgical management of the primary cancer. Routine evaluation for distant metastatic disease consists of a chest radiograph, determination of CEA level, and a contrast-enhanced CT scan of the upper and lower abdomen. The latter also allows evaluation of the urinary tract, as described previously. Suspicious lesions in the liver may need further evaluation by MRI or angio-CT. CT also will help in determining metastasis to the ovaries or dissemination within the peritoneal cavity. Rectal cancer rarely metastasizes to the bones or to the brain, and in the absence of symptoms, these two areas are not included in routine surveillance. The liver must be involved extensively with cancer before liver function tests become abnormal, and most authors find limited use in these tests. PET scanning is extremely sensitive in its ability to detect metastatic or recurrent tumor and is often helpful in planning treatment.

Surgical Treatment General Principles

The surgical treatment of rectal cancer should accomplish complete removal of the cancer and involved regional tissue. The resection should encompass an adequate margin around all facets of the cancer, not just the usually reported distal margin. The extent and timing of the surgery must be planned carefully, including partial or complete resection of surrounding structures. Because most of the rectum resides within the confines of the bones of the pelvis, the technical accomplishment of an anastomosis can be difficult. This difficulty has resulted in a history of significant complications with the colorectal anastomosis and many suggestions for avoiding serious complications, including the use of temporary, protecting intestinal stomas. The ability to construct low anastomoses, deep within the pelvis, has improved in recent years with the advent of the intraluminal surgical stapling instruments. This has markedly reduced the need for abdominal perineal resection without reducing the level of local control of the rectal cancer (Fig. 26-80 and Table 26-15).

Basic principles of cancer surgery should be used, in that dissection of the rectum and surrounding tissue must be accomplished first, and then consideration should be given to restoration of intestinal continuity (Fig. 26-81). Intestinal continuity should be restored ("low anterior resection") only if the anastomosis can be performed with an excellent blood supply and no tension. In addition, the sphincter muscle must function adequately. It is a disservice to accomplish a heroically low anastomosis in an elderly person, only to leave her or him with impaired continence. After all, an opening of the colon without sphincter control is a "colostomy," and a person is much better served to have a well-positioned, well-constructed colostomy on the abdominal wall rather than in the perineal area, which is inaccessible and incapable of having secure coverage by an appliance.

If anal sphincter function cannot be preserved, or if the cancer is so low that the sphincter or anus must be encompassed in the resection, the entire rectum must be removed and egress of feces accomplished by construction of a permanent colostomy (Fig. 26-82). The operation is performed by mobilization of the left colon and rectum through a laparotomy or celiotomy, with dissection of the distal rectum from the perineum, and is referred to as an *abdominal-perineal*, or *AP*, *resection*. The precise terminology is an *abdominal-perineal proctosigmoidectomy*.

Preparation of a patient for resection of a rectal cancer should include antibiotics and mechanical bowel preparation, as described for colonic cancer, plus a few special considerations. The patient should always have potential colostomy sites identified and marked on the abdominal wall before surgery (Fig. 26-83). This should be done even if an anastomosis is anticipated, because unexpected problems can be encountered at the time of surgery, and a colostomy might be required. An unpredicted colostomy is made worse by having it in a poor location. End colostomies are usually constructed of the distal descending (not sigmoid) colon and are located in the left lower quadrant of the abdomen or in the midline, sometimes at the umbilicus. The site is chosen as illustrated in Fig. 26-83. The patient is marked in the supine position and then asked to sit and move about. The site of the stoma should be accessible to the patient for care and on the summit of a fold of fat, not in a crease or on a scar or bony prominence. Any position off the midline should bring the stoma through the body of the rectus muscle. A low anastomosis is sometimes protected with a temporary loop ileostomy, and for this reason, it is advantageous to mark the right lower quadrant before surgery (see Fig. 26-61). The best sites are marked with ink, and the next day, in the operating room, the sites are identified by scratching the skin with a needle.

The operation is performed with general anesthesia, sometimes supplemented with epidural anesthesia; the latter provides perfect relaxation and may facilitate a lower dissection of the rectum and anastomosis. It is not unusual for blood transfusion to be required during or immediately after the procedure, and preparation should be made well in advance, especially if "directed donors" are involved. If the cancer is suspected of involving the bladder or ureter, or if there has been previous pelvic surgery, cystoscopy and placement of ureteral catheters after induction of anesthesia can be extremely helpful. This does not preclude injury to the ureter but usually makes its intraoperative identification and preservation easier.

A patient who is to have surgical resection of the rectum should be positioned on the operating table so that a sterile field can be prepared for access to the perineum, as well as to the abdomen (Fig. 26-84). This allows mobilization of the left colon and rectum from above and passage of anastomotic stapling instruments or removal of the rectum from below. Before the operation is started, a Foley catheter is placed in the bladder, a nasogastric tube in the stomach, and a #34 mushroom-tipped catheter in the rectum. Irrigation of the rectum with saline solution is accomplished through the rectal catheter; and when the effluent is clear, approximately 100 mL of povidone-iodine solution is instilled in the rectum.

The abdomen is opened through a midline incision, and exploration is carried out, looking for other intraabdominal problems, especially the presence of metastatic disease to the peritoneum, the ovaries, or the liver. Small liver metastases, especially on the surface of the liver, may have been missed by CT scan. Any suspicious lesions should be biopsied and undergo frozen-section examination while the operation proceeds. The liver is most accurately evaluated for metastatic disease by use of intraoperative ultrasonography.

The entire left colon is then mobilized on its mesentery, using the embryonic fusion plane (white line of Toldt) between the mesentery and the fat of the lateral pelvic wall (Fig. 26-85 A,B). When this is done properly, the ureter and genital vessels can be identified and reflected posteriorly. Especially if low anastomosis is a possibility, full mobilization of the splenic flexure of the colon may be helpful.

The inferior mesenteric artery is ligated at the aorta (see Fig. 26-85C). This is done not so much to enhance the cure of the cancer as to allow better mobilization of the descending colon and to clearly define the parameters of the distal dissection. This is the first of five potential injury points to the innervation necessary for normal sexual function (Fig. 26-86 C). The mesentery is then incised to the junction of the descending and sigmoid colon. The ascending left colic artery and the inferior mesenteric vein will be transected as this is done (see Fig. 26-85 C). If the need for further mobility is anticipated, the inferior mesenteric vein should be transected a second time at the peripheral margin of the duodenum (see Fig. 26-85 D). This will allow the descending colon, with its blood supply through collaterals from the superior mesenteric artery, to be placed deep within the pelvis without tension.

The bowel is then transected at the junction of the descending and sigmoid colon. The proximal component is protected for construction of an anastomosis or a colostomy. If an anastomosis is anticipated, preparation can be made for using the stapling instrument by inserting a purse-string suture (see Fig. 26-84).

The distal dissection of the rectum is then completed (see Fig. 26-86), first posteriorly between the presacral fascia and the enveloping fascia of the rectum (the second and third nerve injury points), then laterally with transection of the "lateral ligaments" (the fourth nerve injury point), and finally anteriorly with care taken to avoid the seminal vesicles or vagina (the fifth nerve injury point). During this dissection, the rectum rises out of the pelvis. Before beginning the anterior dissection, the surgeon may irrigate the rectum from above with 40% ethanol in an attempt to lessen the chance of implanting viable tumor cells should the rectum be entered during the difficult anterior dissection. The extent of the distal dissection is then determined by the location of the cancer, aiming for a margin 3 to 5 cm beyond the palpable edge of the tumor. If the cancer cannot be palpated, sigmoidoscopy should be performed to precisely locate the cancer and an adequate transection site marked from the abdominal aspect. The dissection can be extended to within the levator muscles. The rectum can be transected with a linear stapling instrument in anticipation of a double- stapled anastomosis (Fig. 26-87). If anastomosis is not possible, the rectum can be removed from below by a second operating team (Fig. 26-88), or the abdominal part of the procedure can be completed, including colostomy construction, the patient turned to the prone-flexed position, and the rectum removed by perineal dissection.

If the rectum and sphincter mechanism are completely removed, a permanent colostomy must be constructed using the prepared distal descending colon. If an anastomosis is constructed but is of questionable security (suspect blood supply, tension, or surrounding sepsis), a protecting loop colostomy or ileostomy, of a temporary nature, can be constructed. This should occur only rarely, because any problems should be corrected before completing the operation. A leaked colorectal anastomosis, even with a temporary colostomy, is a serious problem and is better avoided by correcting the cause for concern.

A special situation is created when a *coloanal anastomosis* is considered (see Fig. 26-81). This involves stripping the distal rectal mucosa and constructing an anastomosis between descending colon and anus, similar to the technique of ileal pouch–anal anastomosis used after colectomy and mucosal proctectomy for mucosal ulcerative colitis or familial

polyposis. In fact, some centers are describing construction of a colonic J-pouch as part of the “coloanal” procedure. If a coloanal anastomosis is constructed, it should be protected with a temporary proximal diverting colostomy or, preferably, a protecting loop ileostomy. Some surgeons find limited use for the coloanal anastomosis because removing only the mucosa is not adequate treatment for an invasive cancer. The coloanal procedure is an excellent, albeit complex, procedure for benign neoplasms, such as “carpet-type” villous adenomas, and for management of radiation-induced rectal injury, where normal well-vascularized tissue can be pulled through the radiation-damaged rectum.

The *Hartmann resection* is a compromise type of resection used occasionally for rectal cancer (see Fig. 26-81). It involves resection of the rectum and/or sigmoid colon and construction of an end colostomy, but it avoids anastomosis and perineal dissection of the distal rectum and perineum. This procedure is used most often for treating sigmoid diverticulitis, but it can be used for management of rectal cancer if there has been too much obstruction or sepsis to allow a safe anastomosis. It is also useful for the elderly or severely unstable patient in whom an adequate distal margin can be accomplished but anastomosis is inappropriate because the procedure should not be prolonged or because the sphincter may not function adequately for a very low anastomosis. If there is no reason to anticipate distal obstruction of the rectal segment, the proximal end can be closed and allowed to reside within the peritoneal cavity. If there is an anal stricture or any other possible distal obstruction, a mucous fistula should be constructed of the distal rectum.

When the Hartmann procedure is done, it is usually anticipated to be a temporary situation with ultimate construction of a colorectal anastomosis to restore intestinal continuity. This eventual anastomosis is sometimes not safe or possible, so the Hartmann procedure becomes the definitive procedure. For this reason, the stoma site should be chosen carefully, because these “temporary” colostomies can become “permanent.”

Metastatic Disease

The management of metastatic disease from rectal cancer is the same as that for colon cancer. Suspicious lesions on or within the liver should have biopsies taken early during the operative procedure and frozen-section evaluation completed. If a small number of easily resectable lesions exist, they can be removed with some margin, once the rectal cancer has been definitely managed. If many lesions exist, biopsy should be taken to help plan subsequent treatment. If massive replacement of the liver exists, management of the primary cancer may need to be modified. Such patients have an extremely limited life expectancy, and the now “palliative” procedure should do just that: provide relief of symptoms with as little morbidity as possible. This may mean only a colostomy without attempted resection of a “fixed” cancer, or it may mean resection with a “short” distal margin to eliminate the need for a colostomy during the few remaining months of life.

If the patient can have adequate resection of local and regional aspects of the rectal cancer, significant cures are being reported with liver resection when there are fewer than four lesions and both lobes of the liver are not involved. This is determined by repeat evaluation of the liver after the patient has recovered from treatment of the primary lesion. Most oncologists will treat diffuse hepatic involvement with systemic chemotherapy, although some prefer to wait until symptoms occur. The delivery of chemotherapeutic agents directly to the liver via the portal vein or the hepatic artery has been described, but no reports have yet shown a favorable ratio of treatment benefits versus morbidity of this modality of treatment.

Metastases to the lung, the second most common site of metastasis from colorectal cancer, are also managed by resection if the primary site can be considered “cured” and the lesions within the lung appear to have the potential for “cure” by resection.

Complications

Complications resulting from the surgical management of rectal cancer overlap with those of any major abdominal procedure, including sepsis, myocardial infarction, bleeding, pulmonary embolus, and wound problems.

There are some specific problems related to rectal cancer. One is risk of injury to sexual function. The sites of these injuries have been described. It is not known precisely how to avoid these injuries and still do an adequate cancer operation. There is at least a 50 percent incidence of significant impotence in men following resection of the rectum for cancer. For this reason, it is critical to discuss the situation with the patient before surgery and to record the pretreatment status of sexual function. It is prudent to include this risk of sexual impairment as part of a preoperative informed-consent note. This and mention of the possible need for a colostomy are the two most important elements of such a note for the patient with rectal cancer.

If a man is rendered impotent by treatment of a rectal cancer, it is advisable to wait a year before seeking consultation for implantation of some type of penile prosthetic device. The wait is important to be sure that the malignant growth has been cleared from the pelvis and to give the patient a chance to overcome functional and psychological impediments to normal sexual function that can result from pelvic surgery and the change in body image caused by the colostomy.

Leakage from colorectal anastomosis has been reported in up to 20 percent of cases using current techniques. This complication can be decreased by constructing the anastomosis with well-vascularized bowel under no tension and by draining the pelvis from above. It is suspected that most “leaks” are due to sepsis in the pelvis that dissects into the anastomosis rather than to primary defects in the anastomosis. Young, muscular men have a higher incidence of anastomotic leaks. This may be due to the technical difficulty of operating within a muscular pelvis, or it may be due to the power of their sphincter mechanism, which acts as a postoperative obstruction and stresses the anastomosis. Such patients should have dilation of the anal sphincter in the operating room, at the conclusion of the procedure, and every few days in the postoperative period, until normal intestinal function returns.

A leaked colorectal anastomosis usually presents between 4 and 7 days postoperatively. The symptoms range from mild temperature elevation, to fecal fistula in the wound, to diffuse peritonitis. When such a problem is suspected, feeding is discontinued and radiologic studies are performed, at first supine and upright films of the abdomen and chest looking for extraluminal gas. CT scan can be helpful to show abscess formation, inflammation, or extraluminal gas, and a water-soluble contrast study can demonstrate the actual leak. Barium should be avoided because leakage of barium (especially mixed with stool) creates a severely destructive form of peritonitis.

The management of a leaking colorectal anastomosis depends on the severity of the problem. The majority cause minimal symptoms of sepsis and are demonstrated only with water-soluble contrast studies of the rectum. Antibiotics and bowel rest usually suffice for treatment. If the leakage has caused an abscess that drains into the rectum or a fistula that drains to the wound, the patient should be placed on antibiotics, complete bowel rest, and total parenteral nutrition for at least 2 weeks. If a large leak is demonstrated, or if the patient has symptoms of diffuse peritonitis, laparotomy should be performed with the intention of constructing a completely diverting colostomy with mucous fistula, or a loop ileostomy. Rarely the anastomosis is taken down and the proximal component used for the colostomy, but the anastomosis should never be reconstructed in the face of significant sepsis. It is important to drain the area of sepsis with the drain placed external to the rectum. Drains placed through the rectal wall perpetuate the fistula.

A special problem can occur intraoperatively during the performance of the rectal dissection: massive venous bleeding from the presacral space. This type of bleeding has been notoriously difficult to control because it actually comes from venous structures within the bone. The older control maneuvers of ligating the iliac vessels are ineffective and may be hazardous. The specimen should be removed and the pelvis packed for 24 to 48 h, or, preferably, a surgical metal “tack” should be driven into the sacrum to compress the venous space.

Malfunction of the urinary system, mainly urinary retention, is also common after resection of the rectum. This is especially true in men, in whom prostatic hypertrophy may coexist. The perineal dissection of the rectum comes very close to the membranous urethra. Foley catheters should be left in place at least 1 week after proctectomy. If urinary symptoms exist preoperatively, urologic consultation should be obtained. Sometimes a patient must be discharged from the hospital with an indwelling urinary catheter or have prostatic resection performed before discharge. Women also may have problems with bladder function postoperatively. The other problem is distortion of the vagina, which impairs sexual function in the postoperative period.

Obstructing Rectal Cancer

Obstructing rectal cancer causes a special problem because it forces the surgeon to operate on intestine that cannot be prepared, and it precludes the use of such adjuvant measures as preoperative radiation therapy. The cancers frequently are at an advanced stage and are fixed to adjacent structures; the patient may be medically compromised by the obstructed intestine as well as by the cancer. The general principles are the same as those discussed previously for obstructing cancer of the left colon.

The simplest and safest way of handling an obstructing rectal cancer is to do a limited laparotomy and construct a *decompressing* transverse colostomy (Fig. 26-89). Ileostomy may not be sufficient for decompression because a competent ileocecal valve may prevent reflux of colonic contents to the ileum. This can be done through a small upper midline incision directly over the dilated colon. The choice of location for the incision and the stoma can be aided by a preoperative x-ray of the abdomen with a coin placed at the umbilicus. Either a loop colostomy (over a small supporting rod) or “blowhole” colostomy (using only one wall of a massively dilated colon) can be constructed (Fig. 26-90). Construction of a decompressing colostomy allows subsequent complete evaluation of the patient and elective resection of the rectal cancer. If adjuvant preoperative treatment is not elected, the definitive resection can be done during the same hospitalization. If adjuvant therapy is chosen, the patient can be discharged from the hospital and the definitive, elective resection performed weeks or months after the decompressing procedure. The major disadvantage of the staged procedure with initial colonic decompression is the inability to explore the entire abdominal cavity at the time of decompression. This may carry some risk in the patient with a “competent” ileal-cecal valve, preventing decompression of the obstruction into the small intestine. Such patients may have massive cecal dilatation and be at some risk of necrosis of the cecal wall.

Some surgeons who operate for obstructing rectal cancer prefer to resect the cancer, using a Hartmann resection, as described previously. Others prefer to resect the cancer and all the dilated abdominal colon, allowing construction of an ileal-rectal anastomosis (see Fig. 26-81), which can be done safely when the cancer and dilated colon are removed. This type of anastomosis causes some, but usually not significant, increase in the number of daily bowel movements.

One also can resect an obstructing rectal cancer, construct an anastomosis, allow the bowel to decompress through the connection, and then protect the anastomosis with a temporary transverse colostomy or loop ileostomy, as described previously.

Colostomy Construction and Management

The construction of a colostomy is sometimes needed in the management of rectal cancer. The indications for its use have already been described, but it is important to emphasize several points. A colostomy is an opening of the colon to the abdominal wall, without the benefit of a sphincter to control the timing of its emptying. The need for a colostomy is determined more by the location of the cancer than by its severity. This is important for patients to understand, since they usually assume, as does all of our society, that the need for a colostomy indicates incurability of the cancer. In years past, this was in fact the case. Until Miles popularized the safe surgical removal of the rectum by the abdominal-perineal approach, most cancers continued to grow until they obstructed as late-stage lesions, and a colostomy was required to prolong a life complicated by the misery of uncontrolled growth of a rectal cancer within the pelvis. As explained previously, colostomies are also used to protect distal complicated anastomoses or perforations or to decompress a colon that is dilated because of distal obstruction. A loop ileostomy has been used increasingly to accomplish the same goal. Its construction does not require colonic dissection or any manipulation of the blood supply to the colorectal anastomosis. Its closure can be performed easily without involving the colon or its vessels.

The function and care of a colostomy are determined by the physiologic differences between the proximal and distal colon. The proximal colon functions to absorb water from the material that passes to it from the ileum. Its contents are liquid, and its peristaltic motion is frequent and irregular. The distal colon acts as a reservoir to allow fecal material to accumulate until it can be expelled. Its content is solid or semisolid, and its peristaltic motion is powerful and infrequent.

Although the functional aspects are important, colostomy construction is better defined by the surgeon according to the required *objective of the stoma* rather than its location. A surgeon must decide if a *decompressing* or a *diverting* stoma is required.

A *decompressing* colostomy is constructed to relieve the colonic distention caused by distal obstruction. A transverse loop colostomy is the most frequently employed stoma of this type (see Fig. 26-89). It is made by bringing a loop of colon to the abdominal wall and suspending it over a small supporting rod. If the colon is too massively dilated to bring through the abdominal wall, only one wall of the colon is used to form a so-called blowhole colostomy (see Fig. 26-90). The technique can be used for a dilated cecum as well, but this results in a stoma that is difficult to manage. Studies have shown that a well-constructed loop transverse colostomy also will provide *diversion* for as long as 6 weeks. After that time, the posterior wall of the colon falls below the abdominal wall, and fecal material can pass into the distal segment.

If complete diversion is required because of distal perforation of the intestine, a loop colostomy is inadequate. Complete diversion can be guaranteed only if the colon is divided, with construction of an end colostomy and mucous fistula.

All colostomies are opened primarily in the operating room, and full thickness of the intestine is sutured to full thickness of skin. This “maturation” of the colostomy allows immediate application of an appliance and avoids subsequent stricture of the stoma, a condition that causes significant difficulty for the patient, requiring ineffective dilation of the stoma and eventual revision. The construction of a colostomy by suspending a dilated loop of colon over a large glass supporting rod and opening it 24 to 48 h later should be relegated to history.

Any patient with abdominal trauma, diverticulitis, or a cancer anywhere in the left colon or rectum should have potential colostomy and ileostomy sites marked preoperatively, as described previously, and the possible need for a stoma should be discussed. For the patients who have a high probability of needing a stoma, more definitive preoperative preparation is made. They and their close family are given information about the stoma and its function, if possible by an enterostomal therapy nurse or other nurse who works closely with surgeons and their ostomy patients. If possible, a visit is arranged with another person who has had successful construction of a similar stoma. This visit can be arranged through the United Ostomy Association, a well-organized volunteer group made up of people who have had ostomy surgery.

The immediate *postoperative care* of the colostomy involves observation for viability and infection and application of an appliance. Supporting rods are removed after 1 week in most patients. Colostomies initially become edematous and then shrink considerably, requiring adjustment in the size of the appliance opening. If a colostomy becomes necrotic because of inadequate blood supply, gentle evaluation must be made into the lumen of the stoma. Necrosis deep in the abdominal wall requires laparotomy and reconstruction of the stoma. More superficial necrosis will be tolerated but usually results in eventual stricture of the stoma.

A patient with an end colostomy and a preoperative regular bowel habit may elect to “control” the colostomy with an *irrigation* regimen. This takes advantage of the infrequent, mass peristaltic motions of the left colon that can be stimulated by distending the distal colon with irrigation solution—the same effect as an enema in the rectum. These patients can begin small-volume irrigation (250 to 500 mL of saline solution or warm water) on the fifth postoperative day and increase to the normal 1000-mL irrigating volume. An individual who irrigates an end colostomy successfully often will not wear an appliance over the stoma but at most a small cover with a pad to absorb mucous and a charcoal filter to control odor.

The long-term care of a colostomy depends on the type of stoma and individual preference. A transverse colostomy requires constant wearing of an appliance that must be emptied multiple times during the day. An end colostomy on the left side can be irrigated or controlled with diet and medication, or it can be left alone to function with an appliance in place. Colostomy problems are usually related to a poor location of the stoma that makes sealing an appliance difficult and the control of gas and odor impossible. Since the colostomy has no sphincter, gas passes without control. Colostomy patients therefore must be observant of gas-forming foods and should have available medications to slow intestinal function during social occasions.

Bleeding and obstructive symptoms should be evaluated for proximal sources, especially synchronous or metachronous cancer. A person with rectal cancer is at increased risk of developing additional cancers. Surveillance endoscopy should be performed through the colostomy just as it would be through the anus. Colostomies can become blocked by impaction just as a rectum does, and the presenting symptoms may be frequent passage of liquid stool around an obstructing bolus. This should be considered before prescribing antidiarrheal agents to a patient having “diarrhea” from the colostomy.

Colostomy stricture usually results from a stoma opened in a delayed fashion in which serositis developed or from a poorly vascularized colostomy. An initially huge stoma can stricture down to a tiny, incapacitating opening. The stenosis is always at skin level, prevents irrigation, and can be the source of enormous difficulty for the patient. These strictures can be corrected easily by excising the scarred skin and resuturing well-vascularized bowel to the skin (Fig. 26-91).

Adjuvant Therapy for Rectal Cancer

The overall 5-year survival rate of rectal cancer has been no better than 50 percent in many series. Death results from distant metastatic disease or, perhaps more significantly, from recurrent disease in the pelvis. Local recurrence in the pelvis is usually reported to be between 20 and 30 percent. This pelvic recurrence is the source of severe misery and suffering. Local recurrence happens because the ability to do wide resection of the rectal cancer and surrounding tissue is limited by the confines of the pelvis. Adjuvant radiation therapy has been advocated to control tumor spread to the periphery of the margins of resection (Fig. 26-92). Some prefer to apply this postoperatively when the adequacy of cancer resection is in doubt. Others prefer to use the radiation therapy preoperatively to destroy the cancer at the periphery before surgical dissection is performed. Our statistics have shown a very low incidence of local recurrence and a high 5-year survival rate with preoperative radiation therapy (Table 26-16). Two treatment regimens were used: The shorter course is used when tumor shrinkage is not required. The biologic dose of radiation is highly dependent on the timing of administration, so 2000 cGy administered in 5 days is not very different from 4500 cGy delivery over a 5-week period. In the longer regimen, there is a 7-week waiting period after completion of radiation therapy to allow resolution of all radiation-induced inflammatory changes and to permit surgery before the fibrotic phase of radiation injury begins.

There is little complication resulting from the use of preoperative radiation therapy, whereas many have seen serious problems with the use of postoperative radiation therapy. This is usually caused by injury to the small intestine, which becomes fixed in the pelvis after surgery (see Fig. 26-92). The major disadvantage of preoperative radiation therapy is that some favorable lesions, not needing adjuvant radiation therapy, will be included. In some series, these favorable but invasive cancers have always been treated with something less than surgical resection, as will be explained shortly.

Some institutions use a compromise referred to as the “sandwich” technique, in which a short course of radiation therapy is given preoperatively, with an additional postoperative dosage used if the pathology of the cancer is unfavorable. Because of the low risk and good results from the preoperative regimens, we have not seen the need for the “sandwich” technique. A modern schema for the utilization of adjuvant radiation therapy in the management of rectal cancer is summarized in Table 26-17.

Currently, systemic adjuvant chemotherapy is used to improve long-term survival by reducing the incidence of distant metastatic disease and to improve local disease control as a radiosensitizer in combination with external-beam radiotherapy. Data from large trials indicate that patients with Stage II and Stage III rectal carcinomas achieve a modest survival benefit from adjuvant chemotherapy. Preliminary evidence suggests that combined-modality treatment with 5-fluorouracil-based chemotherapy in conjunction with radiation therapy provides additional survival benefit. Decreased local recurrence rates with chemoradiation therapy also suggest that 5-fluorouracil acts, in part, as a radiosensitizer. Another local benefit of chemoradiation therapy is its ability to downstage locally advanced rectal cancers (as demonstrated by transrectal ultrasonography) when given preoperatively, which may render a tumor resectable. Trials are underway to further define the role of chemoradiation in the treatment of rectal carcinoma.

Nonresectional Therapy

So far we have described only radical resection for invasive rectal cancer, yet it has long been known that some cancers are detected in their early phase, before onset of regional extension. In these cases, patients could be cured by effective local destruction of the tumor (Table 26-18). Adequate local palliation also becomes an issue in patients with widespread metastatic disease and a severely limited life expectancy.

Identifying the cancer appropriate for something less than resection is made difficult because no adequate clinical staging system exists for rectal cancer in the pretreatment state. Morson warned that even small presumed Dukes' A lesions will have spread to regional lymph nodes 10 percent of the time. New diagnostic modalities described previously, such as transanal ultrasound, CT scanning, MRI, monoclonal antibody, nuclear scan, and PET scanning, and accurate tumor markers of biologic activity should soon allow more accurate selection of favorable yet invasive rectal cancer that is amenable to treatment by something less than radical proctectomy.

These cancers have been treated by local excision, excision and deep electrocoagulation, and most recently, *endocavitary radiation* (Fig. 26-93). This latter technique applies high-dose, superficial, low-voltage radiation directly to the cancer. This accurately directed high dosage avoids injury to normal surrounding tissue, the dose-limiting factor of external radiation treatment.

We have advanced from excision and deep coagulation of these favorable but invasive cancers, to endocavitary radiation, and finally, to external combined with endocavitary radiation treatment. We have had good results with selected cancers but have learned that initially “aggressive” cancers require resection, even in high-risk patients and those with metastatic disease. The future probably will show that external radiation, combined with any adequate form of local destruction, will constitute adequate, definitive treatment for some rectal cancers. Endocavity radiation is not available in all centers.

Incurable Cancer

The management of hepatic and pulmonary metastases has been described previously. The question arises as to what to do with the rectal cancer itself in the patient who presents with incurable, massive metastatic disease. If the life expectancy is longer than 6 months, it is probably appropriate to give 4500 cGy external radiation (perhaps combined with systemic chemotherapy) to shrink the tumor and then to perform a palliative resection. This takes away a significant amount of “out-of-hospital” time from the patient whose time has become precious, but it is clearly the best form of palliation, even if it requires a colostomy. An alternative recently described is to give a course of external radiation, with or without chemotherapy, and maintain a lumen through the cancer by laser destruction, dilation, or stenting of the tumor, thus avoiding resective surgery and the need for a colostomy. This choice must be used in highly selected patients. Since both plans start with external radiation therapy, the ultimate choice of treatment can be delayed until response to the external radiation is observed. It also allows the surgeon to better understand the patient, his or her desires, and his or her general state of health during the process.

Recurrent Cancer

A special and complicated problem exists for local recurrence of rectal cancer. It had been thought previously that the cancers recurred at the distal margin of the anastomosis and grew into the lumen. This surely happens with an inadequate margin, but most surgeons in the field now believe that there is residual cancer on the pelvic wall at the time of initial resection. It takes approximately 18 months for this cancer to grow large enough to present through the anastomosis. These cancers, by their nature, are fixed to the pelvic wall and surrounding viscera. They cause significant symptoms of pain secondary to sacral nerve roots, obstruction of the ureters, or rectal obstruction after low anterior resection are difficult to palliate and almost impossible to cure. This low salvage for locally recurrent rectal cancer has been the major impetus for investigating adjuvant modalities of treatment, such as preoperative radiation therapy.

Locally recurrent disease can be detected by biopsy after a low anterior resection and in women through the vagina after an abdominal-perineal resection. A more difficult problem exists in men after an abdominal- perineal resection, when no orifice exists to allow access to the suspicious area. Any postoperative patient with symptoms in the pelvis or a rising CEA titer should be suspected of recurrent disease. PET scanning is becoming most helpful in the diagnosis of these patients. Once the suspicious area is identified, CT-guided biopsy becomes helpful in confirming the diagnosis.

Once the diagnosis is made, therapy must be planned carefully. External radiation usually combined with chemotherapy is especially important if it has not been used before. The tissue damage from radiation is cumulative, and repeat courses are impossible. Resection should be planned to include all involved organs, such as vagina, bladder, sacrum, or adherent small intestine. Preoperative cystoscopy with biopsy and careful pelvic examination are mandatory. Removal of the rectum and urinary bladder with surrounding lymphatic tissue is called a *pelvic exenteration* (Fig. 26- 94). This requires the patient to have two permanent stomas, one for feces and one for urine. Some but not all of the sacrum can be removed, as described previously in the section on anatomy.

Consideration has been given to intraoperative radiation therapy to the tumor using a linear accelerator; this requires a combined operating room–radiation therapy suite, and the results so far have not been favorable. Techniques are being described for implanting high-dose afterload radiation therapy catheters to the areas suspicious for unclear margins.

Despite such heroic measures, the palliation and salvage rate of recurrent rectal cancer is very low (10 to 20 percent). The mode of death from locally recurrent rectal cancer involves much more intense suffering than that which results from the progressive malnutrition of distant metastatic disease. For these reasons, it is important that a surgeon managing rectal cancer know all the management options and that the first therapy applied be the best “curative” procedure for a given patient.

Other Rectal Tumors Lymphoma

Most gastrointestinal lymphomas represent metastatic disease. Primary, non-Hodgkin's, colonic lymphomas account for approximately 10 percent of all gastrointestinal lymphomas. The gastrointestinal tract is a common site of HIV-associated non-Hodgkin's lymphomas. For the diagnosis of primary lymphoma of the gastrointestinal tract to be made, there should be no palpable peripheral adenopathy, no involvement of the liver or spleen, a normal white blood cell count, and no involvement of other sites. The cecum and rectum are the most common sites of colonic involvement.

Lymphoma of the intestinal tract may appear as annular thickening, bulky exophytic growth, or thickened dilatations of the bowel wall. The cut surface has the appearance of “fish flesh.” Colonic lymphomas are classified by cell type and in order of frequency are histiocytic, lymphocytic, mixed cell, and Hodgkin's diseases.

Diagnosis

Patients are typically 50 to 60 years old, with colonic lymphoma more common in men and rectal involvement more common in women. Abdominal or rectal pain, altered bowel habits, weight loss, and rectal bleeding are the most common presenting symptoms. Approximately 20 percent of patients will present with obstructive symptoms.

An accurate history and physical examination with complete blood work, chest x-ray, barium enema, abdominal CT scan, and bone marrow biopsy are used for preoperative staging. Most rectal lymphomas are palpable by digital examination or visible by biopsy through a sigmoidoscope. Patients with colonic lymphoma may have tender, palpable abdominal masses. Barium enema examination may reveal multiple lesions consisting of filling defects and infiltrating or ulcerating lesions. The appearance of a lymphoma on barium enema examination can be similar to that of a carcinoma. The lesions often are submucosal; consequently, superficial colonoscopic biopsies are not diagnostic. CT scan may help to characterize primary colonic lymphomas and to identify tumor spread or invasion.

Treatment

Complete excision is the treatment of choice for these tumors. Resection of primary gastrointestinal lymphomas permits accurate staging, better local control of the tumor, and decreased complications of bleeding and perforation. Liver and lymph node biopsies, as well as splenectomies, have been recommended for complete operative staging.

Postoperative radiation therapy and chemotherapy are given according to the stage of the tumor. Radiation therapy is indicated for palliation of unresectable disease and in cases when the surgical resection is not complete. Chemotherapy is recommended for treatment of systemic disease.

Prognosis depends on the stage at diagnosis. The 5-year survival rate following curative resection is 50 percent. The best prognosis exists in patients with intraluminal lymphomas less than 5 cm in diameter with no lymph node metastases. Patients with AIDS and those with widespread disease have a poor prognosis.

Retrorectal Tumors

Pathophysiology

The retrorectal space is bound anteriorly by the rectum, posteriorly by the presacral fascia, superiorly by the peritoneal reflection, inferiorly by the rectosacral fascia, and laterally by the lateral ligaments, ureters, and iliac vessels (see Fig. 26-37). More than 50 percent of these lesions are congenital (Table 26-19). The remainder are classified as neurogenic, osseous, or miscellaneous lesions. Inflammatory lesions and metastatic carcinoma also should be considered in the differential diagnosis of lesions in this space. Malignancy is more common in lesions found in the postnatal pediatric population than in adults. In the adult population, 10 percent of cystic lesions are malignant, whereas solid lesions have a 60 percent malignancy rate.

Developmental cysts constitute the majority of congenital lesions and may arise from any of the germ cell layers. Dermoid and epidermoid cysts are benign lesions that have an ectodermal origin. When these cysts communicate with the skin surface, a postanal dimple is found. These cysts can become infected and may be misdiagnosed as a fistula-in-ano. Enterogenous cysts are endodermal in origin and are benign, thin-walled, and lined by columnar epithelium. They are usually asymptomatic, but they have a tendency to become infected. Teratomas are true neoplasms and contain tissue from each germ cell layer. They are more common in women and are considered tumors of infancy and childhood. Teratomas are uncommon in adults, but when they occur, 30 percent become malignant.

Chordomas are the most common malignant tumors found in this region and are believed to arise from remnants of the notochord. These cancers are slow-growing, invasive lesions that destroy adjacent bone. Plain radiographs show the expansion and destruction of bone, trabeculation, and calcification.

Anterior sacral *meningoceles* contain cerebrospinal fluid. Patients present with headaches, often associated with constipation and low back pain. The “scimitar” sign (a sacrum with a rounded, concave border without bony destruction) is the pathognomonic x-ray appearance of anterior sacral meningoceles. Meningitis can occur, and aspiration of these lesions should be avoided to prevent this complication.

Diagnosis

Many patients are symptom-free and the tumor is found on routine pelvic or rectal examination. Symptomatic patients with retrorectal tumors often complain of postural pain in the lower back, perianal, or rectal region, with the pain often referred to the legs or buttocks. Other complaints include recurrent infections, constipation, diarrhea, urinary symptoms, or obstructed labor. More than 90 percent of these lesions are palpable by digital rectal examination.

Plain x-ray or CT scans are used to confirm the diagnosis and to aid in surgical planning. CT scans help to define the limits of the lesion and to determine whether it is cystic or solid. Bony destruction indicates malignancy. Barium enema will show anterior displacement of the rectum. Myelograms may be indicated if central nervous system involvement is suspected. Biopsy is not recommended if the lesion is deemed operable because tumor seeding along biopsy tracts and septic complications after needle biopsies have been reported.

Treatment

Surgical resection is the treatment of choice, even in asymptomatic patients, because of the risk that the lesion may become infected, cause dystocia, be malignant, or have malignant potential. The abdominal approach is recommended for high retrorectal and extraspinal neurogenic lesions. The posterior (transsacral) approach is recommended for low lesions and infected cysts. For large lesions, an abdominal-sacral approach has been most useful. The coccyx needs to be removed to gain better exposure and to decrease the risk of recurrence, since it is frequently the source of the cysts. If needle biopsy was performed, the tract must be included in the resection to prevent tumor implantation in the needle tract. En bloc resections of sacrum, neural structures, and rectum are sometimes needed for curative resection of chordomas. Anterior sacral meningoceles can be treated by posterior laminectomy, anterior transabdominal, or a combination anteroposterior approach.

Benign tumors and cysts have a good prognosis and are treated adequately by surgical excision. Chordomas have a high local recurrence rate. They are slow-growing tumors, and 10-year survival rates vary from 20 to 70 percent. High-dose radiotherapy may provide palliation in inoperable or inadequately resected chordomas. Chemotherapy has no role in the management of tumors in this region.

Colonic Carcinoid

Carcinoid tumors arise from neuroectodermal tissue. In the gastrointestinal tract they originate in the enterochromaffin cells at the base of the crypts of Lieberkühn. These cells contain neurosecretory granules and are part of the amine precursor uptake and decarboxylation (APUD) system. Approximately 2 percent of gastrointestinal carcinoids occur in the colon and 15 percent in the rectum. Colonic carcinoids occur most frequently in the cecum, followed by the transverse and sigmoid colon. The male-to-female ratio is equal, and the peak incidence is in the sixth to seventh decades of life. There is a high incidence of synchronous or metachronous lesions in patients with colonic or rectal carcinoids.

Diagnosis

Colonic carcinoids smaller than 1 cm are usually asymptomatic and almost never malignant. They are submucosal and rarely seen on barium enema examination or colonoscopy.

They are usually found incidentally during operation for other disease.

Because most midgut and hindgut carcinoids are clinically silent, the diagnosis is often made late in the course of the disease. Symptomatic colonic carcinoids are often bulky, advanced lesions. Many patients present with weight loss, anorexia, abdominal discomfort, and occasionally a palpable mass. Barium enema may reveal polypoid or circumferential lesions, simulating findings seen with adenocarcinomas. Carcinoid tumors evoke a desmoplastic reaction that may distort surrounding tissues. Less than 5 percent of colonic carcinoids cause the carcinoid syndrome; if the syndrome is present, it usually indicates liver metastases.

Invasive malignancy correlates with size of the lesions. Lesions smaller than 2 cm rarely metastasize, but symptomatic carcinoids larger than 2 cm are usually aggressive malignancies. Eighty percent of patients with lesions larger than 2 cm have metastases at the time of diagnosis, 45 percent have local spread, and 38 percent have distant metastases, with the liver being the most common site of distant spread.

Lesions greater than 2 cm that have invaded the muscularis propria or show positive nodes have a very poor prognosis. Median survival for these lesions is less than 12 months. The overall 5-year survival rate for colonic carcinoid is less than 50 percent.

Treatment

Small lesions found incidentally during exploration may be removed by wedge or segmental resection. Local excision can be curative for lesions less than 2 cm discovered before invasion of the muscularis propria, if lymph nodes are negative. The surgical treatment for larger lesions with no liver metastasis is similar to that for colonic adenocarcinoma. Single liver metastases should be resected. Resection of the primary tumor can be palliative in patients with otherwise incurable metastatic disease. There is evidence that a-interferon used in the treatment of malignant carcinoids resulted in decreased symptoms and stabilization of disease. Somatostatin has been used for management of symptomatic carcinoid syndrome. The prognosis of patients with colonic carcinoids is worse than that of patients with carcinoids of the rectum or appendix.

Rectal Carcinoid

Diagnosis

Rectal carcinoids are yellow-gray submucosal nodules covered by intact mucosa. They are usually small and asymptomatic and may be seen on proctosigmoidoscopy. Larger lesions can ulcerate, causing rectal bleeding and pain. Malignancy correlates with the size of the lesion and is based on invasion into the muscularis propria, not histologic differentiation. There is a 4 percent chance of malignancy in lesions smaller than 2 cm, but it is unsafe to assume solely on the basis of its size that a lesion is benign. More than 60 percent of rectal carcinoids larger than 2 cm have liver metastases, and 90 percent have local lymph node metastases at the time of diagnosis.

Treatment

Transanal excision is indicated for small, asymptomatic rectal lesions. The depth of invasion should be evaluated, and a more extensive cancer operation should be performed for those lesions with invasion of the muscularis propria. Malignant lesions should be treated by anterior resection or abdominal perineal resection. Radical surgery, however, has not been beneficial for rectal carcinoids with evidence of distant metastatic disease. The 5-year survival rate among persons with no metastases is 92 percent. The 5-year survival rate is 44 percent if there are local lymph node metastases and less than 10 percent if distant metastases are present.

Composite Carcinoid-Carcinomas

Composite carcinoid-carcinomas (adenocarcinoids) are uncommon tumors that have histologic features of both carcinoid and adenocarcinoma. Neoplastic change involving a common precursor cell is more likely than coincidental malignant change of two mature cell types.

These tumors behave more like adenocarcinomas than carcinoids. They have been reported in various locations along the gastrointestinal tract; 80 percent are found in the appendix and 10 percent in the colon and rectum. Metastases to regional lymph nodes are found in 69 percent and to the liver in 23 percent of patients at the time of diagnosis.

Surgical treatment should be planned as if the lesions were adenocarcinomas. Somatostatin may control symptoms by preventing release of pharmacologically active tumor products.

Anal Neoplasms

Tumors of the Anal Margin Squamous Cell Carcinoma

A major consideration in managing anal cancers involves the precise anatomic location of the lesion. Squamous cancer of the skin at the anal *margin* behaves similarly to other cutaneous squamous cell cancers, whereas squamous (or transitional) cancer of the anal *canal* is a high-grade cancer with significant risk of metastasis and death.

Squamous cell carcinomas of the anal margin arise between the dentate line and the outer limit of perianal skin (defined as 5 cm from the anal verge). These tumors are well-differentiated keratinizing lesions similar to squamous carcinoma of the skin.

Diagnosis

These cancers are more common in men. Patients frequently present with complaints of a palpable mass, bleeding, itching, pain, or tenesmus. The lesions have rolled, everted edges and a central ulceration. Associated anal condylomata or chronic fistula-in-ano may be found in up to 15 percent of patients. Any chronic, nonhealing ulceration in the perianal area should be biopsied to rule out squamous carcinoma. These cancers are slow-growing and metastasize late, and most patients present 12 to 24 months after the onset of symptoms.

Treatment

Small lesions are treated by wide local excision with a 2-cm margin. Combined chemotherapy and radiation have been used successfully to induce regression of large lesions. Abdominal-perineal resection should be reserved for patients with fecal incontinence and those who develop local recurrence after conservative treatment. Synchronous metastases to the inguinal nodes occur in 8 to 40 percent of patients and are treated by inguinal node dissection once positive nodes have been identified. Lesions less than 5 cm have a better prognosis, and the overall 5-year survival rate is 60 to 80 percent.

Basal Cell Carcinoma

Basal cell carcinoma is a rare tumor of the anal margin and behaves like basal cell carcinoma elsewhere on the skin. It is more common in men and presents as a lesion with raised edges and central ulceration. Basal cell carcinoma is treated by wide local excision with adequate margins and primary closure or skin grafting. Large lesions may require abdominal- perineal resection. Although these cancers rarely metastasize, there is a 30 percent local recurrence rate. Recurrences are treated by reexcision.

Bowen's Disease

Bowen's disease is an intraepidermal squamous cell carcinoma (carcinoma in situ) that is a rare, slow-growing cancer. It usually presents in the elderly but has been reported in association with anal condyloma in young, sexually active patients.

Diagnosis

Patients often complain of perianal itching, burning, or bleeding. Many lesions go unnoticed but may be found incidentally during histologic examination of tissues removed during minor anal procedures. Discrete, erythematous, scaly plaques are found on examination. Lesions may have the appearance of eczema or chronic irritation due to pruritus ani.

The presence of ulceration may indicate the development of an invasive carcinoma. Less than 10 percent of untreated lesions go on to develop invasive squamous cell carcinoma, and 35 percent of these develop metastases. Associated malignancies have been reported in the past with high percentage of patients with Bowen's disease, but recent studies have proved this to be incorrect.

Treatment

Treatment consists of wide local excision with adequate margins, as determined by frozen section. Multiple biopsies in all four quadrants of the perineal skin are then taken to ensure complete resection. Skin grafts may be necessary for coverage if the excision is extensive. Recurrent lesions, without invasion, can be treated by reexcision.

Paget's Disease

Extramammary, perianal Paget's disease is a rare pathologic entity. It arises from the intraepidermal portion of the apocrine glands. Paget's disease is frequently but not always associated with an underlying carcinoma. If left untreated, adenocarcinoma of the apocrine glands may develop after a long preinvasive phase.

Diagnosis

Patients usually are 60 to 70 years old, and there is a female predominance. Severe, intractable pruritus is characteristic of this disease. Examination reveals an erythematous, eczematous rash. Diagnosis often is delayed because lesions frequently are treated with a variety of topical medications without success before biopsy. A biopsy should be obtained from any nonhealing, pruritic anal lesion in order to exclude this diagnosis. Histologic examination reveals Paget cells, which are large pale cells with eccentric nuclei. A detailed physical examination and complete evaluation of the gastrointestinal tract should be performed; patients with extramammary Paget's disease are at increased risk of developing noncontiguous malignancies.

Treatment

Wide local excision with clear margins, confirmed by frozen section, and with multiple biopsies of the perianal skin is the treatment of choice. Abdominal-perineal resection may be necessary for the very few advanced lesions associated with invasive carcinoma. Lymph node dissection is indicated only for patients with proven positive inguinal nodes.

The most frequent sites of metastatic disease are the inguinal and pelvic lymph nodes, followed by liver, bone, lung, brain, bladder, prostate, and adrenal glands. Patients with metastatic disease do poorly despite aggressive treatment.

Anal Canal Cancer Epidermoid Carcinoma

The cloacogenic or transitional zone of the anal canal is located 6 to 12 mm above the dentate line. Carcinomas of this area may be referred to as *nonkeratinizing squamous, basaloid, cloacogenic, or transitional*. Basaloid carcinomas have the characteristic histologic appearance of palisading nuclei. Squamous carcinomas in this region are generally

nonkeratinizing. Although these cancers have different histologic features, they exhibit similar biologic behavior and are grouped together as epidermoid carcinomas of the anal canal.

Diagnosis

These cancers are more common in women. Many patients have only minor symptoms and are often erroneously diagnosed as having benign anorectal disorders. These lesions can present as an indurated, bleeding mass that may itch or cause discomfort; however, some patients present with a painless mass lesion.

Digital rectal examination reveals the size, location, and degree of fixation of these tumors. Proctosigmoidoscopy verifies the exact location of the tumor relative to the dentate line. Biopsies of the mass are necessary for diagnosis. Evaluation of the remainder of the colon should be done with colonoscopy or barium enema. Examination of both inguinal regions determines if there is any suspicious nodal enlargement, and a chest radiograph and CT scan of the liver are performed to search for distant metastatic disease.

Between 30 and 40 percent of these patients have metastatic disease at the time of diagnosis. The risk of metastasis increases with the depth of invasion, the size of lesion, and the histologic grade of the lesion. When located below the dentate line, these cancers metastasize to the inguinal nodes. Those above the dentate line will spread to the superior rectal, internal pudendal, hypogastric, or obturator lymph nodes. The most common site of distant metastatic disease is the liver.

Treatment

Small, early, node-negative lesions that are confined to the submucosa may be treated by local excision. The treatment for larger, more deeply invasive lesions has changed in the past two decades. In the past, abdominal-perineal resection was the standard treatment for squamous carcinoma of the anal canal. Local recurrence after abdominal-perineal resection was 25 to 50 percent, and the 5-year survival rate ranged from 50 to 70 percent.

Radiation therapy along with surgical treatment for residual disease has been used successfully. A 5-year survival rate of 79 percent has been reported, but there is a high complication rate associated with the high doses of radiation used for treatment.

In 1972, Nigro began using combined chemoradiation in an effort to reduce the radiation dose necessary to adequately treat these tumors. In his initial protocol, 3000 cGy external radiation, combined with 5-fluorouracil and mitomycin C, was followed by abdominal-perineal removal of the rectum. No residual carcinoma was found in 60 to 90 percent of patients, and the protocol was modified to reserve surgery for those patients with residual disease after chemoradiation.

Simultaneous chemoradiation is the preferred method of treatment for squamous carcinoma of the anus (Table 26-20). Local recurrence is 15 percent and is treated by abdominal-perineal resection. Treatment toxicity is increased for those patients with a history of pelvic surgery and those patients requiring abdominal-perineal resection after chemoradiation. Several salvage chemoradiation protocols have been used with variable success. Negative prognostic factors include the presence of involved nodes, lesions greater than 5 cm, and invasion of contiguous organs other than the rectum and skin. The overall 5-year survival rate is 83 percent, and as expected, patients with visceral metastases have a poor prognosis.

The presence of involved inguinal lymph nodes at the time of diagnosis of the primary carcinoma is a bad prognostic sign. Involved inguinal nodes are included in the radiation field, but lymph node dissection is reserved only for patients with positive nodes after chemoradiation. There is no indication for prophylactic lymph node dissection.

Squamous cell carcinoma has been reported in the chronic perineal wounds and sinus tracts seen in Crohn's disease. Malignancy should be considered in any patient with continuous anal pain or a nonhealing perineal wound. Biopsy of such sinus tracts is mandatory and may be diagnostic.

Adenocarcinoma

Adenocarcinoma of the anal canal is usually the distal extension of a primary rectal cancer. It may rarely arise from the columnar epithelium of the anal glands or within a long-standing anorectal fistula. Patients complain of perianal pain, swelling, and frequently have an abscess or fistula-in-ano. Abdominal-perineal resection is indicated in these carcinomas, sometimes after preoperative radiation therapy. Many of these patients have distant metastases at the time of diagnosis, and the long- term outlook is poor.

Melanoma

Melanoma represents 1 to 3 percent of anal canal cancers. It is the third most common site for melanoma, following the skin and the eye. The majority arise from the epidermoid lining of the anal canal adjacent to the dentate line. They spread submucosally, and by the time they cause symptoms, the extent of invasion is usually beyond surgical cure.

Diagnosis

The average age at presentation is between the fifth and sixth decades of life, the male-to-female ratio is equal, and the majority of patients reported in the literature are Caucasian.

Patients commonly present with bleeding, pain, or an anal mass. They may complain of alteration in bowel habits and are often misdiagnosed as having a thrombosed hemorrhoid. Most melanomas in this region are nonpigmented or lightly pigmented polypoid lesions.

Metastases occur via lymphatic and hematogenous routes, with lymphatic spread to the mesenteric nodes being more common than to inguinal nodes. Thirty-eight percent of patients have metastatic disease at the time of diagnosis, with lung, liver, and bone the most frequent sites of hematogenous spread.

Treatment

There has been no difference in survival in patients treated by wide local excision or abdominal-perineal resection; however, the radical resection is more effective for control of local disease and is the only procedure that removes clinically undetected lymph nodes.

Prophylactic lymph node dissection is of no value, and therapeutic lymph node dissection is advised only for those patients with positive inguinal nodes. Current chemotherapy has been ineffective. Overall 5-year survival rate is less than 20 percent, with few reported long-term survivors. Five- year survival rates in patients with lesions larger than 1.6 mm have not been reported.

PSYCHIATRIC DISORDERS

Many psychiatric disorders include symptoms of anorectal pathology as manifestations of the psychiatric disease. Surgeons must be aware of this and limit surgical therapy to specifically diagnosed surgically remediable problems. It is a serious error to operate for severe anorectal symptoms without definite findings of a primary disease process. For example, there has been extensive work seeking a psychiatric basis for the irritable bowel syndrome.

Patients who present with severe constipation or pelvic floor malfunction (such as nonrelaxation of the anal sphincter mechanism) have been shown to have a high incidence of psychological abnormalities, but no specific psychiatric diagnosis has been consistent.

The problem of unexplained anal pain is a difficult diagnostic dilemma. One must consider psychological overlay in most patients who continue to complain of anal pain despite no abnormal findings on examination. The full evaluation of these patients must include physical examination, CT scan, defecography, anal manometry, and electromyography. Cancer, inflammation, abscess, and nerve root compression are all considered possibilities. If the pain persists despite a negative evaluation, one must consider a psychological causation and refer the patient for behavioral evaluation. It is possible that spasm of the levator floor and the proctalgia fugax syndrome are related to stress, much as are migraine headaches. For these patients, treatment of the underlying problem is appropriate after cancer and other surgical problems are eliminated as possibilities.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 27 - The Appendix

Rosemary A. Kozar
Joel J. Roslyn

ANATOMY AND FUNCTION

The appendix first becomes visible during embryologic development in the eighth week of life as a protuberance off the terminal portion of the cecum. During both antenatal and postnatal development, the growth rate of the cecum exceeds that of the appendix, displacing the appendix medially toward the ileocecal valve. The relationship of the base of the appendix to the cecum remains constant, whereas the tip can be found in a retrocecal, pelvic, subcecal, preileal, or right pericolic position. These anatomic considerations have significant clinical importance in the context of acute appendicitis. The three taenia coli converge at the junction of the cecum with the appendix and can be a useful landmark to identify the appendix. The appendix can vary in length from less than 1 to greater than 30 cm; most appendices are 6 to 9 cm in length. Appendiceal absence, duplication, and diverticula have all been described.

For many years, the appendix was erroneously viewed as a vestigial organ with no known function. It is now well recognized that the appendix is an immunologic organ which actively participates in the secretion of immunoglobulins, particularly IgA. Though the appendix is an integral component of the *gut-associated lymphoid tissue (GALT)* system, its function is not essential and appendectomy has not been associated with any predisposition to sepsis or any other manifestation of immune compromise. Lymphoid tissue first appears in the appendix about 2 weeks after birth. The amount of lymphoid tissue increases throughout puberty, remains steady for the next decade, and then begins a steady decrease with age. After the age of 60, virtually no lymphoid tissue remains within the appendix and complete obliteration of the appendiceal lumen is common.

INFLAMMATION OF THE APPENDIX

Acute Appendicitis

Historical Background

There is evidence in the literature that alchemists and physicians in the 1500s recognized the existence of a clinical entity associated with severe inflammation of the cecal region, known as “perityphlitis.” Although the first successful appendectomy was reported in 1736, it was not until 1886 that Reginald Fitz helped establish the role of surgical removal of the inflamed appendix as curative therapy for this disease, which was once thought to be fatal. In 1889, Charles McBurney presented his classic report before the New York Surgical Society on the importance of early operative intervention for acute appendicitis in which he described the point of maximal abdominal tenderness to be determined by the pressure of one finger placed one-third of the distance between the anterior superior iliac spine and the umbilicus. Five years later he devised the muscle-splitting incision which today bears his name.

Incidence

Appendicitis remains one of the most common acute surgical diseases. The incidence of acute appendicitis roughly parallels that of lymphoid development, with the peak incidence in early adulthood. Appendicitis occurs more frequently in males especially at the time of puberty. A review of over 2000 patients with appendicitis demonstrated an overall 1.3:1 male predominance.

A decline from 100 cases per 100,000 population to 52 cases per 100,000 population was demonstrated over a study period from 1975 to 1991. This degree of change does not seem to be explained by improved diagnosis, and the explanation for this phenomenon remains elusive. Currently, 84 percent of all appendectomies are performed for acute pathology. The rate of normal appendectomy averages 16 percent, with females comprising 68 percent of those patients found to have a normal appendix at exploration.

Etiology and Pathogenesis

Obstruction of the lumen is the dominant causal factor in acute appendicitis. Fecaliths are the usual cause of appendiceal obstruction. Less common are hypertrophy of lymphoid tissue; inspissated barium from previous x-ray studies; vegetable and fruit seeds; and intestinal worms, particularly ascarids.

The frequency of obstruction rises with the severity of the inflammatory process. Fecaliths are found in about 40 percent of cases of simple acute appendicitis, about 65 percent of cases of gangrenous appendicitis without rupture, and about 90 percent of cases of gangrenous appendicitis with rupture.

The probable sequence of events following occlusion of the lumen is as follows. A closed-loop obstruction is produced by the proximal block, and continuing normal secretion of the appendiceal mucosa rapidly produces distention. The luminal capacity of the normal appendix is only about 0.1 mL—there is no real lumen. Secretion of as little as 0.5 mL distal to a block raises the intraluminal pressure to about 60 cmH₂O. The human being is one of the few animals with an appendix capable of secreting at pressures high enough to lead to gangrene and perforation. Distention stimulates nerve endings of visceral afferent pain fibers, producing vague, dull, diffuse pain in the midabdomen or lower epigastrium. Peristalsis is also stimulated by the rather sudden distention, so that some cramping may be superimposed on the visceral pain early in the course of appendicitis.

Distention continues, not only from continued mucosal secretion, but also from rapid multiplication of the resident bacteria of the appendix. As pressure in the organ increases, venous pressure is exceeded. Capillaries and venules are occluded, but arteriolar inflow continues, resulting in engorgement and vascular congestion. Distension of this magnitude usually causes reflex nausea and vomiting, and the diffuse visceral pain becomes more severe. The inflammatory process soon involves the serosa of the appendix and in turn parietal peritoneum in the region, producing the characteristic shift in pain to the right lower quadrant.

The mucosa of the gastrointestinal tract, including the appendix, is very susceptible to impairment of blood supply. Thus its integrity is compromised early in the process, allowing bacterial invasion of the deeper coats. As progressive distention encroaches on the arteriolar pressure, the area with the poorest blood supply suffers most: ellipsoidal infarcts develop in the antimesenteric border. As distention, bacterial invasion, compromise of vascular supply, and infarction progress, perforation occurs, usually through one of the infarcted areas on the antimesenteric border.

This sequence is not inevitable; some episodes of acute appendicitis apparently subside spontaneously. Many patients who are found at operation to have acute appendicitis give a history of previous similar but less severe attacks of right lower quadrant pain. Pathologic examination of the appendices removed from these patients often reveals thickening and scarring, suggesting old healed acute inflammation.

Bacteriology

A variety of anaerobes, aerobes, or facultative bacteria have been cultured from peritoneal fluid, abscess contents, and appendiceal tissue in patients with gangrenous or perforated appendicitis. An average of 10 different organisms were recovered per specimen. *Bacteroides fragilis* and *Escherichia coli* were isolated from almost all specimens. Other frequent isolates were *Peptostreptococcus* (80 percent), *Pseudomonas* (40 percent), *Bacteroides splanchnicus* (40 percent), and *Lactobacillus* (37 percent).

A quantitative bacteriologic study of the appendix wall of children showed no significant differences between the flora of normal and of acutely inflamed appendices. *Bacteroides*, *E. coli*, and streptococci were the most common organisms isolated.

Cytomegalovirus-associated appendicitis has been reported in a patient with acquired immunodeficiency syndrome.

Clinical Manifestations

Symptoms

Abdominal pain is the prime symptom of acute appendicitis. Classically the pain is initially diffusely centered in the lower epigastrium or umbilical area, is moderately severe, and is steady, sometimes with intermittent cramping superimposed. After a period varying from 1 to 12 h, but usually within 4 to 6 h, the pain localizes in the right lower quadrant. This classic pain sequence, though usual, is not invariable. In some patients the pain of appendicitis begins in the right lower quadrant and remains there. Variations in the anatomic location of the appendix account for many of the variations in the principal locus of the somatic phase of the pain. For example, a long appendix with the inflamed tip in the left lower quadrant causes pain in that area; a retrocecal appendix may cause principally flank or back pain; a pelvic appendix, principally suprapubic pain; and a retroileal appendix may cause testicular pain, presumably from irritation of the spermatic artery and ureter. Malrotation is also responsible for puzzling pain patterns. The visceral component is in the normal location, but the somatic component is felt in that part of the abdomen where the cecum has been arrested in rotation.

Anorexia nearly always accompanies appendicitis. It is so constant that the diagnosis should be questioned if the patient is not anorectic. Vomiting occurs in about 75 percent of patients, but is not prominent or prolonged, and most patients vomit only once or twice.

Most patients give a history of obstipation from before the onset of abdominal pain, and many feel that defecation would relieve their abdominal pain. However, diarrhea occurs in some patients, particularly children, so that the pattern of bowel function is of little differential diagnostic value.

The sequence of symptom appearance has great differential diagnostic significance. In over 95 percent of patients with acute appendicitis, anorexia is the first symptom, followed by abdominal pain, which is followed in turn by vomiting (if vomiting occurs). If vomiting precedes the onset of pain, the diagnosis should be questioned.

Signs

Physical findings are determined principally by the anatomic position of the inflamed appendix as well as by whether the organ has already ruptured when the patient is first examined.

Vital signs are not changed very much by uncomplicated appendicitis. Temperature elevation is rarely more than 1°C; the pulse rate is normal or slightly elevated. Changes of greater magnitude usually mean that a complication has occurred or that another diagnosis should be considered.

Patients with appendicitis usually prefer to lie supine, with the thighs, particularly the right thigh, drawn up, because any motion increases pain. If asked to move, they do so slowly and gingerly.

The classic right lower quadrant physical signs are present when the inflamed appendix lies in the anterior position. Tenderness is often maximal at or near the point described by McBurney as being “located exactly between an inch and a half and two inches from the anterior spinous process of the ileum on a straight line drawn from that process to the

umbilicus.” Direct rebound tenderness is usually present, and referred or indirect rebound tenderness is frequently present, and the tenderness is felt maximally in the right lower quadrant, indicating peritoneal irritation. Rovsing’s sign—pain in the right lower quadrant when palpatory pressure is exerted in the left lower quadrant—also indicates the site of peritoneal irritation. Cutaneous hyperesthesia in the area supplied by the spinal nerves on the right at T10, T11, and T12 frequently but not always accompanies acute appendicitis. In patients with obvious appendicitis, this sign is superfluous, but in some early cases it may be the first positive sign. It is elicited either by needle prick or, better, by gently picking up the skin between the forefinger and thumb. This ordinarily is not unpleasant but is painful in areas of cutaneous hyperesthesia.

Muscular resistance to palpation of the abdominal wall roughly parallels the severity of the inflammatory process. Early in the disease, resistance, if present, consists mainly of voluntary guarding. As peritoneal irritation progresses, muscle spasm increases and becomes largely involuntary—true reflex rigidity as opposed to voluntary guarding.

Variations in the position of the inflamed appendix produce variations from the usual in physical findings. With a retrocecal appendix, the anterior abdominal findings are less striking, and tenderness may be most marked in the flank. When the inflamed appendix hangs into the pelvis, abdominal findings may be entirely absent, and the diagnosis may be missed unless the rectum is examined. As the examining finger exerts pressure on the peritoneum of the cul-de-sac of Douglas, pain is felt in the suprapubic area as well as locally. Signs of localized muscle irritation may also be present. The psoas sign indicates an irritative focus in proximity to that muscle. The test is performed by having patients lie on their left side; the examiner then slowly extends the right thigh, thus stretching the iliopsoas muscle. The test is positive if extension produces pain. Similarly, a positive obturator sign of hypogastric pain on stretching the obturator internus indicates irritation at that locus. The test is performed by passive internal rotation of the flexed right thigh with the patient supine.

Laboratory Findings

Mild leukocytosis, ranging from 10,000 to 18,000/mm³, is usually present in patients with acute, uncomplicated appendicitis and is often accompanied by a moderate polymorphonuclear predominance. If a normal white blood cell count with no left shift is present, the diagnosis of acute appendicitis should be reconsidered. It is unusual for the white blood cell count to be greater than 18,000/mm³ in uncomplicated appendicitis. White blood cell counts above this level raise the possibility of a perforated appendix with or without an abscess. Urinalysis can be useful to rule out the urinary tract as the source of infection. Although several white or red blood cells can be present from ureteral or bladder irritation as a result of an inflamed appendix, bacteriuria in a catheterized urine specimen will not be seen with acute appendicitis.

Plain films of the abdomen, although frequently obtained as part of the general evaluation of a patient with an acute abdomen, are rarely helpful in diagnosing acute appendicitis, although they can be of significant benefit in ruling out additional pathology. In patients with acute appendicitis, one often sees an abnormal bowel gas pattern, which is a nonspecific finding. The presence of a fecalith is rarely noted on plain films, but when present is highly suggestive of the diagnosis. A chest x-ray is sometimes indicated to rule out referred pain from a right lower lobe pneumonic process.

Graded compression sonography has been suggested as an accurate way to establish the diagnosis of appendicitis. The appendix is identified as a blind-ending, nonperistaltic bowel loop originating from the cecum. With maximal compression, the diameter of the appendix is measured in the anteroposterior dimension. A scan is considered positive if a noncompressible appendix 6 mm or greater in the anteroposterior direction is demonstrated (Fig. 27-1). The presence of an appendicolith establishes the diagnosis. Sonographic demonstration of a normal appendix, which is an easily compressible blind-ending tubular structure measuring 5 mm or less, excludes the diagnosis of acute appendicitis. The study is considered negative if the appendix is not visualized and there is no pericecal fluid or mass. When the diagnosis of acute appendicitis is excluded by sonography, a brief survey of the remainder of the abdominal cavity should be performed to establish an alternative diagnosis. In females of childbearing age, the pelvic organs must be adequately visualized either transabdominally or by endovaginal examination in order to exclude gynecologic pathology as a possible cause of acute abdominal pain. The sonographic diagnosis of acute appendicitis has a reported sensitivity of 78 to 96 percent and a specificity of 85 to 98 percent. Sonography is similarly effective in children and pregnant women, although its application is somewhat limited in late pregnancy.

Sonography has definite limitations and results are user-dependent. A false-positive scan can occur in the presence of periappendicitis from surrounding inflammation, a dilated fallopian tube can be mistaken for an inflamed appendix, inspissated stool can mimic an appendicolith, and, in obese patients, the appendix may not be compressible not because of an acutely inflamed appendix but because of overlying fat. False-negative sonograms can occur if appendicitis is confined to the appendiceal tip, the cecum is retrocecal in location, the appendix is markedly enlarged and mistaken for small bowel, or if the appendix is perforated and therefore compressible.

Additional radiographic techniques include computed tomography, barium enema, and radioisotope-labeled leukocyte scans. Although CT has been reported to be as, or more, accurate than sonography, it is significantly more expensive. Because of the cost and added radiation exposure, CT should be used primarily when an appendiceal abscess is suspected to ascertain the feasibility of percutaneous drainage. Diagnosis based on barium enema depends on the nonspecific findings of extrinsic mass effect on the cecum and nonfilling of the appendix, and is associated with an accuracy ranging from 50 to 84 percent. Radiographic evaluation of patients with suspected appendicitis should be reserved for patients in whom the diagnosis is in doubt and should not delay, or substitute for, prompt operative intervention when clinically indicated.

Laparoscopy can serve as both a diagnostic and therapeutic maneuver for patients with acute abdominal pain and suspected acute appendicitis. Laparoscopy is probably most useful for the evaluation of females with lower abdominal complaints, since appendectomy is performed on a normal appendix in as many as 30 to 40 percent of such patients. Differentiating acute gynecologic pathology from acute appendicitis is easily accomplished using the laparoscope.

Appendiceal Rupture

Immediate appendectomy has long been the recommended treatment of acute appendicitis because of the known progression to rupture. A study on the natural history of appendicitis demonstrated a rate of 14 percent normal, 70 percent inflamed, and 16 percent perforated appendices. This study suggested that delays in presentation were responsible for the majority of perforated appendices. There is no accurate way of determining when and if an appendix will rupture prior to resolution of the inflammatory process. Although it has been suggested that observation and antibiotic therapy alone may be an appropriate treatment for acute appendicitis, nonoperative treatment risks the morbidity and mortality associated with a ruptured appendix.

Appendiceal rupture occurs most frequently distal to the point of luminal obstruction along the antimesenteric border of the appendix. Rupture should be suspected in the presence of fever greater than 39°C (102°F) and a white blood cell count greater than 18,000/mm³. In the majority of cases rupture is contained and patients display localized rebound tenderness. Generalized peritonitis will be present if the walling-off process is ineffective in containing the rupture.

In 2 to 6 percent of cases, an ill-defined mass will be detected on physical examination. This could represent a phlegmon, which is matted loops of bowel adherent to the adjacent inflamed appendix, or a periappendiceal abscess. Patients who present with a mass have a longer duration of symptoms, usually at least 5 to 7 days. The ability to distinguish acute, uncomplicated appendicitis from perforation on the basis of clinical findings is often difficult, but it is important to make the distinction since their treatment differs. CT scan may be very beneficial in guiding therapy. Phlegmons and small abscesses can be treated conservatively with intravenous antibiotics; well-localized abscesses may be managed with percutaneous drainage; and complex abscesses should be considered for surgical drainage. If operative drainage is required, it should be performed using an extraperitoneal approach, with appendectomy reserved for cases in which the appendix is easily accessible. Interval appendectomy performed at least 6 weeks following the acute event has classically been recommended for all patients treated either nonoperatively or with simple drainage of an abscess. The reported incidence of recurrent appendicitis in patients not undergoing interval appendectomy ranges from 0 to 37 percent, and is highest during the first year. Because a perforated cecal carcinoma can be mistaken for an appendiceal abscess, all patients over the age of 50 years should undergo either barium enema or colonoscopic examination prior to interval appendectomy.

Differential Diagnosis

The differential diagnosis of acute appendicitis is essentially the diagnosis of the “acute abdomen” (see Chap. 22). This is because clinical manifestations are not specific for a given disease but are specific for disturbance of a physiologic function or functions. Thus an essentially identical clinical picture can result from a wide variety of acute processes within or near the peritoneal cavity that produce the same alterations of function as acute appendicitis.

Accuracy of preoperative diagnosis should be about 85 percent. If it is consistently less, some unnecessary operations are probably being done, and a more rigorous preoperative differential diagnosis is in order. On the other hand, an accuracy consistently greater than 90 percent should also cause concern, since this may mean that some patients with atypical but bona fide acute appendicitis are being “observed” when they should have prompt surgical intervention. The Haller group has shown, however, that this is not invariably true. Before their study, the perforation rate at the hospital where the study took place was 26.7 percent, and acute appendicitis was found at 80 percent of the operations. By a policy of intensive in-hospital observation when the diagnosis of appendicitis was unclear, the group raised the rate of acute appendicitis found at operation to 94 percent, while the perforation rate remained unchanged at 27.5 percent.

There are a few conditions in which operation is contraindicated, but in general the disease processes that are confused with appendicitis are also surgical problems or, if not, are not made worse by surgical intervention. The more frequent error is to make a preoperative diagnosis of acute appendicitis only to find some other condition (or nothing) at operation; much less frequently, acute appendicitis is found after a preoperative diagnosis of another condition. The most common erroneous preoperative diagnoses—accounting for more than 75 percent—in descending order of frequency are acute mesenteric lymphadenitis, no organic pathologic condition, acute pelvic inflammatory disease, twisted ovarian cyst or ruptured graafian follicle, and acute gastroenteritis.

Differential diagnosis of appendicitis depends upon three major factors: the anatomic location of the inflamed appendix; the stage of the process, that is, whether simple or ruptured; and the age and sex of the patient.

Acute Mesenteric Adenitis

This is the disease most often confused with acute appendicitis in children. Almost invariably an upper respiratory infection is present or has recently subsided. The pain is usually less or more diffuse, and tenderness is not as sharply localized as in appendicitis. Voluntary guarding is sometimes present, but true rigidity is rare. Generalized lymphadenopathy may be noted. Laboratory procedures are of little help in arriving at the correct diagnosis, although a relative lymphocytosis, when present, suggests mesenteric adenitis. Observation for several hours is in order if the diagnosis of mesenteric adenitis seems likely, since mesenteric adenitis is a self-limited disease, but if the differentiation remains in doubt, immediate operation is the only safe course.

Acute Gastroenteritis

This is very common in childhood but can usually be easily differentiated from appendicitis. Viral gastroenteritis, an acute self-limited infection of diverse causes, is characterized by profuse watery diarrhea, nausea, and vomiting. Hyperperistaltic abdominal cramps precede the watery stools. The abdomen is relaxed between cramps, and there are no localizing signs. Laboratory values are normal.

Salmonella gastroenteritis results from ingestion of contaminated food. Abdominal findings are usually similar to those in viral gastroenteritis, but in some cases the abdominal pain is intense, localized, and associated with rebound tenderness. Chills and fever are common. The leukocyte count is usually normal. The causative organisms can be isolated from essentially 100 percent of patients, but culturing may take too long to help the clinician in making a timely differential diagnosis in cases of abdominal pain. Similar attacks in other persons eating the same food as the patient greatly strengthen the presumptive diagnosis of salmonella gastroenteritis.

Typhoid fever is now a rare disease. This probably accounts for the frequency of missed diagnosis—it is rarely seen and rarely thought of. The onset is less acute than in appendicitis, with a prodrome of several days. Differentiation is usually possible because of the prostration, maculopapular rash, inappropriate bradycardia, and leukopenia. Diagnosis is confirmed by culture of *Salmonella typhosa* from stool or blood. Intestinal perforation, usually in the lower ileum, develops in about 1 percent of cases and requires immediate surgical therapy.

Diseases of the Male

Diseases of males must be considered in differential diagnosis of appendicitis, including torsion of the testis and acute epididymitis, since epigastric pain may overshadow local symptoms early in these diseases. Seminal vesiculitis may also mimic appendicitis but can be diagnosed by palpating the enlarged, tender seminal vesicle on rectal examination.

Meckel's Diverticulitis

This causes a clinical picture very similar to that of acute appendicitis. Preoperative differentiation is academic and unnecessary, since Meckel's diverticulitis is associated with the same complications as appendicitis and requires the same treatment—prompt surgical intervention. Diverticulectomy can nearly always be done through a McBurney incision, extended if necessary. If the base of the diverticulum is broad, so that removal would compromise the lumen of the ileum, then resection of the segment of ileum bearing the diverticulum with end-to-end anastomosis is done.

Intussusception

In contrast to Meckel's diverticulitis, it is extremely important to differentiate intussusception from acute appendicitis, because the treatment is quite different. The age of the patients is important: appendicitis is very uncommon under age 2, whereas nearly all idiopathic intussusceptions occur under age 2. Intussusception occurs typically in a well-nourished infant who is suddenly doubled up by apparent colicky pain. Between attacks of pain the infant appears quite well. After several hours, the patient usually passes a bloody mucoid stool. A sausage-shaped mass may be palpable in the right lower quadrant. Later, as the intussusception progresses distad, the right lower quadrant feels abnormally empty. The preferred treatment of intussusception, if seen before signs of peritonitis supervene, is reduction by barium enema, but treatment of acute appendicitis by barium enema may be catastrophic.

Regional Enteritis

The manifestations of acute regional enteritis—fever, right lower quadrant pain and tenderness, and leukocytosis—often simulate acute appendicitis. Diarrhea and the infrequency of anorexia, nausea, and vomiting favor a diagnosis of enteritis but are not sufficient to exclude acute appendicitis without celiotomy. In an appreciable percentage of patients with chronic regional enteritis, the diagnosis has been first made at the time of operation for presumed acute appendicitis. Acute ileitis should be distinguished from Crohn's disease. In the face of an acutely inflamed distal ileum with no cecal involvement and a normal appendix, appendectomy is indicated. Progression to Crohn's ileitis is uncommon.

Perforated Peptic Ulcer

Perforated peptic ulcer closely simulates appendicitis if the spilled gastroduodenal contents gravitate down the right gutter to the cecal area and if the perforation spontaneously seals fairly soon, thus minimizing upper abdominal findings.

Other Lesions

Diverticulitis or perforating carcinoma of the cecum or of that portion of the sigmoid that lies on the right side may be impossible to distinguish from appendicitis. Extensive diagnostic studies in an attempt to make a preoperative differentiation are not warranted.

Epiplaic Appendagitis

Epiplaic appendagitis probably results from infarction of the appendage(s) secondary to torsion. Symptoms may be minimal, or there may be continuous abdominal pain in an area corresponding to the contour of the colon, lasting several days. Pain shift is unusual, and there is no diagnostic sequence of symptoms. The patient does not look ill, nausea and vomiting are unusual, and appetite is commonly unaffected. Localized tenderness over the site is usual and is often marked on rebound but without rigidity. In 25 percent of reported cases, pain has persisted or recurred until the infarcted epiplaic appendages were removed.

Urinary Tract Infection

Acute pyelonephritis, on the right side particularly, may mimic a retroileal acute appendicitis. Chills, right costovertebral angle tenderness, pus cells, and particularly bacteria in the urine usually suffice to differentiate the two.

Ureteral Stone

If the calculus is lodged near the appendix, it may simulate a retrocecal appendicitis. Pain referred to the labia, scrotum, or penis; hematuria; and/or absence of fever or leukocytosis suggest stone. Pyelography usually confirms the diagnosis.

Primary Peritonitis

Primary peritonitis rarely mimics simple acute appendicitis but presents a picture very similar to diffuse peritonitis secondary to a ruptured appendix. The diagnosis is made by peritoneal aspiration. If nothing but cocci are seen on the Gram's-stained smear, peritonitis is primary and treated medically; if the flora are mixed, secondary peritonitis is indicated.

Henoch-Schönlein Purpura

This syndrome usually occurs 2 to 3 weeks after a streptococcal infection. Abdominal pain may be prominent, but joint pains, purpura, and nephritis are nearly always present also.

Yersiniosis

Human infection with *Yersinia enterocolitica* or *Y. pseudotuberculosis* is probably transmitted through food contaminated by feces or urine. *Yersinia* infections cause a variety of clinical syndromes, including mesenteric adenitis, ileitis, colitis, and acute appendicitis. Many of the infections are mild and self-limited, but some lead to a systemic septic course with a high fatality rate if untreated. The organisms are usually sensitive to tetracyclines, streptomycin, ampicillin, and kanamycin. A preoperative suspicion of the diagnosis should not delay operative intervention, since appendicitis caused by *Yersinia* cannot be clinically distinguished from appendicitis of other causation. About 6 percent of cases of mesenteric adenitis and 5 percent of cases of acute appendicitis are caused by *Yersinia* infection.

Campylobacter jejuni causes diarrhea and pain that mimics the pain of appendicitis. The organism can be cultured from stool.

Gynecologic Disorders

The rate of erroneous diagnosis of acute appendicitis is highest in young adult females. Rates of appendectomy being performed on a normal appendix of 32 to 45 percent have been reported in women 15 to 45 years old. Diseases of the female internal generative organs that may be erroneously diagnosed as appendicitis are, in approximate descending order of frequency, pelvic inflammatory disease, ruptured graafian follicle, twisted ovarian cyst or tumor, endometriosis, and ruptured ectopic pregnancy. Laparoscopy plays a significant role in establishing the diagnosis.

Pelvic Inflammatory Disease

The infection is usually bilateral but if confined to the right tube may mimic acute appendicitis. Nausea and vomiting are nearly always present in patients with appendicitis, but only in approximately half of those with pelvic inflammatory disease. The greatest value of these symptoms for establishing a diagnosis of pelvic inflammatory disease is their absence. Pain and tenderness are usually lower, and motion of the cervix is exquisitely painful. Intracellular diplococci may be demonstrable on smear of the purulent vaginal discharge. The ratio of appendicitis to pelvic inflammatory disease is low in the early phase of the menstrual cycle and high during the luteal phase. The clinical use of all the above-mentioned distinctions has resulted in a reduction of the incidence of negative findings on laparotomy in young women to 15 percent.

Ruptured Graafian Follicle

Not uncommonly, ovulation results in the spill of sufficient blood and follicular fluid to produce brief, mild lower abdominal pain. If the fluid is unusually copious and from the right ovary, appendicitis may be simulated. Pain and tenderness are rather diffuse. Leukocytosis and fever are minimal or absent. Since this pain occurs at the midpoint of the menstrual cycle, it is often called mittelschmerz.

Other Diseases

Diseases not mentioned in the previous sections that occur in patients of all ages and both sexes and that must be considered in the differential diagnosis of appendicitis are foreign body perforations of the bowel; closed-loop intestinal obstruction; mesenteric vascular occlusion; pleuritis of the right lower chest; acute cholecystitis; acute pancreatitis; hematoma of the abdominal wall; and numerous conditions too rare to mention.

Appendicitis in the Young

The establishment of a diagnosis of acute appendicitis in the young is even more difficult than in the adult. The inability of young children to give an accurate history, diagnostic delays by both parents and physicians, and the frequency of gastrointestinal upset in children are all contributing factors. The more rapid progression to rupture and the inability of the underdeveloped greater omentum to contain a rupture lead to significant morbidity rates in children. In a study of 1366 specimens from pediatric patients undergoing appendectomies, histologic analysis revealed 12 percent normal appendices, 68 percent nonperforated appendices, and 20 percent perforated appendices. Children under 8 years old had a twofold increase in the rate of perforation as compared to older children. The incidence of major complications was related to rupture, with an incidence of 1.2 percent in simple appendicitis compared to 6.4 percent in perforated appendicitis. The treatment regimen for perforated appendicitis generally includes immediate appendectomy, antibiotic irrigation of the peritoneal cavity, transperitoneal drainage through the wound, and a 10-day antibiotic treatment regimen. Laparoscopic appendectomy has also been shown to be safe and effective in children.

Appendicitis in the Elderly

Although the incidence of appendicitis in the elderly is lower than in younger patients, the morbidity and mortality are significantly increased in this patient population. Delays in diagnosis, a more rapid progression to perforation, and concomitant disease are all contributing factors. The diagnosis of appendicitis may be more subtle and less typical than in younger individuals, and a high index of suspicion needs to be maintained. In patients over 80 years old, perforation rates of 49 percent and mortality rates of 21 percent have been cited.

Appendicitis During Pregnancy

Appendicitis is the most frequently encountered extrauterine disease requiring surgical treatment during pregnancy. The incidence is approximately 1 in 2000 pregnancies. Acute appendicitis can occur at any time during pregnancy but is more frequent during the first two trimesters. As fetal gestation progresses, the diagnosis of appendicitis becomes more difficult as the appendix is displaced laterally and superiorly (Fig. 27-2). Nausea and vomiting after the first trimester or new-onset nausea and vomiting should raise the consideration of appendicitis. Abdominal pain and tenderness will be present, although rebound and guarding are less frequent due to laxity of the abdominal wall. Elevation of the white blood cell count above the normal pregnancy levels of 15,000 to 20,000/mm³ with a predominance of polymorphonuclear cells is usually present. When the diagnosis is in doubt, abdominal ultrasound may be beneficial. Laparoscopy may be indicated in equivocal cases, especially early in pregnancy. The performance of any operation during pregnancy carries a risk of premature labor of 10 to 15 percent, and the risk is similar for both negative laparotomy and appendectomy for simple appendicitis. The most significant factor associated with both fetal and maternal death is appendiceal perforation. Fetal mortality increases from 3 to 5 percent with early appendicitis to 20 percent with perforation. The suspicion of appendicitis during pregnancy should prompt rapid diagnosis and surgical intervention.

Appendicitis in Patients with AIDS or HIV Infection

The cause of appendicitis in HIV-positive patients and patients with AIDS is similar to the cause of appendicitis in the general population. In some cases, however, an underlying opportunistic infection, usually secondary to cytomegalovirus or Kaposi's sarcoma, may be responsible for the development of appendiceal inflammation. Although the clinical presentation of appendicitis in patients who have AIDS or are HIV-positive does not differ significantly from patients without AIDS or HIV infection, the leukocytosis usually noted in acute appendicitis may not be present. Cytomegalovirus enteritis and tuberculosis or lymphoma involving the distal ileum can mimic appendicitis. Diagnostic laparoscopy can be helpful, with laparoscopic appendectomy being indicated when an inflamed appendix is found. If appendectomy is performed prior to perforation, there is no increase in morbidity or mortality.

Treatment

Despite the advent of more sophisticated diagnostic modalities, the importance of early operative intervention should not be minimized. Once the decision to operate for presumed acute appendicitis has been made, the patient should be prepared for the operating room. Adequate hydration should be insured, electrolyte abnormalities corrected, and preexisting cardiac, pulmonary, and renal conditions should be addressed. Many trials have demonstrated the efficacy of preoperative antibiotics in lowering the infectious complications in appendicitis. It is common practice by most surgeons to routinely administer antibiotics to all patients with suspected appendicitis. If simple acute appendicitis is encountered, there is no benefit in extending antibiotic coverage beyond 24 h. If perforated or gangrenous appendicitis is found, antibiotics are continued until the patient is afebrile and has a normal white blood cell count. For intraabdominal infections of gastrointestinal tract origin of mild to moderate severity, the Surgical Infection Society has recommended single-agent therapy with cefoxitin, cefotetan, or ticarcillin-clavulanic acid. For more severe infections, single-agent therapy with carbapenems or combination therapy with a third-generation cephalosporin, monobactam, or aminoglycoside, plus anaerobic coverage with clindamycin and metronidazole, is indicated.

Open Appendectomy (Fig. 27-3)

Most surgeons employ either a McBurney (oblique) or Rocky-Davis (transverse) right lower quadrant muscle-splitting incision in patients with suspected appendicitis. The incision should be centered over the point of maximal tenderness or a palpable mass. If an abscess is suspected, a laterally placed incision is imperative to avoid generalized contamination of the peritoneal cavity. If the diagnosis is in doubt, a lower midline incision is recommended by some to allow a more extensive examination of the peritoneal cavity.

Several techniques can be used to locate the appendix. Since the cecum is usually visible within the incision, the convergence of the taeniae can be followed to the base of the appendix. A sweeping lateral to medial motion can aid in delivering the appendiceal tip into the operative field. Occasionally limited mobilization of the cecum is needed to aid in adequate visualization. Once identified, the appendix is mobilized by dividing the mesoappendix, taking care to ligate the appendiceal artery securely. The appendiceal stump can be managed by simple ligation or by ligation and inversion with either a purse-string or Z stitch as tradition dictates. As long as the stump is clearly viable and the base of the cecum not involved with the inflammatory process, the stump can be safely ligated with a nonabsorbable suture. The mucosa is frequently obliterated to avoid the development of a mucocele. The peritoneal cavity is irrigated and the wound closed in layers. If perforation or gangrene is found, the skin and subcutaneous tissue should be left open and allowed to heal by secondary intent or be closed in 4 to 5 days as a delayed primary closure. In children, who generally have little subcutaneous fat, primary wound closure has not led to an increased incidence of wound infection.

If appendicitis is not found, a methodical search for an alternative diagnosis must be performed. The cecum and mesentery should first be inspected. Next, the small bowel is examined in a retrograde fashion beginning at the ileocecal valve. In females, special attention should be paid to the pelvic organs. An attempt is also made to examine the upper abdominal contents. Peritoneal fluid should be sent for Gram's stain and culture. If purulent fluid is encountered, it is imperative that the source be identified. A medial extension of the incision (Fowler-Weir), with division of the anterior and posterior rectus sheath, is acceptable if further evaluation of the lower abdomen is indicated. If upper abdominal pathology is encountered, the right lower quadrant incision is closed and an appropriate upper midline incision performed.

Laparoscopy

Laparoscopy has emerged as a new technique for both the diagnosis and treatment of acute appendicitis. Laparoscopy can be of enormous assistance in the evaluation of young women of childbearing age in whom acute appendicitis is suspected. Laparoscopic appendectomy may be preferable in obese patients who would require a large incision for the open approach or in individuals who are particularly concerned about cosmesis. There are a number of theoretical advantages for the laparoscopic approach over the traditional open operation described over 100 years ago. These include decreased incidence of wound infection, less pain for the patient, reduced hospitalization, and a more rapid return to employment. The data have not demonstrated any significant economic benefit to laparoscopic appendectomy for the general population. Both operative time and supply costs are increased while duration of hospital stay is not significantly shortened. Nonetheless, laparoscopic appendectomy has been demonstrated to be a safe and effective treatment. Recently reported prospective randomized trials comparing open and laparoscopic techniques (Table 27-1) indicate that operative time is generally greater using the laparoscopic approach, but the time between the operation and return to activity, and the rate of wound infection, are decreased.

The superiority of laparoscopic appendectomy over the conventional approach continues to be an issue of some debate. Laparoscopy's most important role is in diagnostic evaluation of young females whose symptoms are not the classical symptoms of appendicitis and in patients in whom the diagnosis remains in question. It does provide visualization the entire abdominal cavity, unlike the limited exposure provided by a right lower quadrant incision. For properly trained personnel, laparoscopic appendectomy may be appropriate for most cases of appendicitis, with a conversion rate of approximately 10 percent, mostly in patients with ruptured appendices.

The principles behind laparoscopic appendectomies are similar to those for other laparoscopic procedures (Fig. 27-4). All patients should have an indwelling urinary catheter and nasogastric tube inserted prior to trochar insertion. The surgeon frequently stands to the patient's left with the video monitor at the foot of the table, or to the patient's right. Pneumoperitoneum is established and a 10-mm trochar cannula is inserted through the umbilicus. A 10-mm forward-viewing laparoscope is placed through the cannula and the peritoneal cavity is inspected. Next, a 10-mm trochar is introduced into the suprapubic region in the midline and additional 5-mm ports placed in either the right upper or lower quadrant. Exposure is facilitated by placing the patient in the Trendelenburg position, right side up. Generally the cecum is easily visualized and the appendix easily identified. Gentle traction can be applied to the mesoappendix by retracting the tip of the appendix with an atraumatic grasper placed through the right upper quadrant trochar. The mesoappendix is divided with a stapling device or by using electrocautery for dissection and clips or a ligating loop to secure the appendiceal artery. Division of the mesoappendix should be done as close to the appendix as possible. After the base of the appendix is adequately exposed, two ligating loops are placed proximally and one distally over the base. The appendix is divided with scissors or electrocauterization. Alternatively, the appendix could be divided using a stapling device. Invagination of the appendiceal stump is not routinely performed. The appendix is removed via the suprapubic trocar site.

The role of laparoscopic appendectomy for removal of the normal appendix has not been well defined. As for an open procedure, an appendectomy should be considered in patients explored for right lower quadrant pain with no identifiable pathology. Complication rates of 1 to 3 percent have been reported under these circumstances. If the appendix is normal, the same thorough search for an alternative diagnosis should be performed as with an open procedure.

Prognosis

Mortality

The mortality from appendicitis in the United States has steadily decreased from a rate of 9.9 per 100,000 in 1939 to 0.2 per 100,000 in 1986. Among the factors responsible are the significantly decreasing incidence of appendicitis; better diagnosis and treatment, attributable to the now- available antibiotics, intravenous fluids, blood, and plasma; and a higher percentage of patients receiving definitive treatment before rupture.

Principal factors in mortality are (1) whether rupture occurs before surgical treatment and (2) the age of the patient. The overall mortality rate in unruptured acute appendicitis is a little higher than the rate for a general anesthetic, which is 0.06 percent. The overall mortality rate in ruptured acute appendicitis is about 3 percent—a 30-fold increase. The mortality rate of ruptured appendicitis in the elderly is about 15 percent—a fivefold increase from the overall rate.

Death is usually attributable to uncontrolled sepsis—peritonitis, intraabdominal abscesses, or gram-negative septicemia. Sepsis may impose metabolic demands of such magnitude on the cardiovascular or respiratory systems that they cannot be met, in which case cardiac or respiratory insufficiency is the direct cause of death. Pulmonary embolism continues to account for some deaths. Aspiration causing the patient to drown in his or her own vomitus is a significant cause of death in the older age group.

Morbidity

Morbidity rates parallel mortality rates, being precipitously increased by rupture of the appendix and to a lesser extent by old age. In one report, complications occurred in 3 percent of patients with nonperforated appendicitis and in 47 percent of patients with perforations. Most of the serious early complications are septic and include abscess and wound infection. Wound infection is common but is nearly always confined to the subcutaneous tissues and promptly responds to wound drainage, which is accomplished by reopening the skin incision. Wound infection predisposes to wound dehiscence also. The type of incision is relevant; complete dehiscence rarely occurs in a McBurney incision. The efficacy of systemic antibiotics in reducing the incidence of wound infections has not been agreed on. But a significant reduction in morbidity has been shown in patients receiving metronidazole or the combination of systemic clindamycin and topical ampicillin.

The incidence of intraabdominal abscesses secondary to peritoneal contamination from gangrenous or perforated appendicitis has decreased markedly since the introduction of potent antibiotics. The sites of predilection for abscesses are the appendiceal fossa, pouch of Douglas, subhepatic space, and between loops of intestine. The latter are usually

multiple. Transrectal drainage is preferred for an abscess that bulges into the rectum.

Fecal fistula is an annoying but not particularly dangerous complication of appendectomy that may be caused by sloughing of that portion of the cecum inside a constricting purse-string suture, by the ligature's slipping off a tied but not inverted appendiceal stump, or by necrosis from an abscess encroaching on the cecum.

Intestinal obstruction, initially paralytic but sometimes progressing to mechanical obstruction, may occur with slowly resolving peritonitis with loculated abscesses and exuberant adhesion formation.

Late complications are quite uncommon. Adhesive band intestinal obstruction after appendectomy does occur but much less frequently than after pelvic surgical therapy. The incidence of inguinal hernia is three times greater in patients who have had an appendectomy. Incisional hernia is like wound dehiscence in that infection predisposes to it, it rarely occurs in a McBurney incision, and it is not uncommon in a lower right paramedian incision.

TUMORS

Appendiceal malignancies are rare, occurring in 0.5 percent of all appendectomies. The tumors are usually discovered at the time of laparotomy either as an incidental finding or in association with acute inflammation of the appendix. In a recent review of over 2000 appendectomy specimens, histologic confirmation of appendiceal neoplasms included carcinoid (0.27 percent), adenocarcinoma (0.14 percent), malignant mucocele (0.005 percent), and lymphoma (0.005 percent).

Carcinoid

The finding of a firm, yellow, bulbar mass in the appendix should raise the suspicion of an appendiceal carcinoid. The appendix is the most common site of gastrointestinal carcinoid followed by the small bowel and rectum. Carcinoid syndrome is rarely associated with appendiceal carcinoid unless widespread metastases are present, which occurs in 2.9 percent of cases. Symptoms attributable directly to the carcinoid are rare, though the tumor can occasionally obstruct the appendiceal lumen much like a fecalith and result in acute appendicitis.

The majority of carcinoids are located in the tip of the appendix. Malignant potential is related to size, with tumors less than 2 cm rarely resulting in extension outside of the appendix. In one report, 78 percent of appendiceal carcinoids were less than 1 cm, 17 percent were 1 to 2 cm, and only 5 percent were greater than 2 cm. Treatment rarely requires more than simple appendectomy. For tumors less than 2 cm with extension into the mesoappendix, and for all tumors greater than 2 cm, a right hemicolectomy should be performed.

Adenocarcinoma

Primary adenocarcinoma of the appendix is a rare neoplasm of three major histologic subtypes: mucinous adenocarcinoma, colonic adenocarcinoma, and adenocarcinoid. The most common mode of presentation for appendiceal carcinoma is that of acute appendicitis. Patients may also present with ascites or a palpable mass, or the neoplasm may be discovered during an operative procedure for an unrelated cause. The recommended treatment for all patients with adenocarcinoma of the appendix is a formal right hemicolectomy. Appendiceal adenocarcinomas have a propensity for early perforation, though not clearly associated with a worsened prognosis. The presence of pseudomyxoma peritonei secondary to mucinous adenocarcinoma does not adversely affect prognosis. Overall 5-year survival is 55 percent and varies with stage and grade. Patients with appendiceal adenocarcinoma are at significant risk for both synchronous and metachronous neoplasms, approximately half of which will originate from the gastrointestinal tract.

Mucocele

An appendiceal mucocele leads to progressive enlargement of the appendix from the intraluminal accumulation of a mucoid substance. Mucoceles are of four histologic types, and the type dictates the course of the disease and prognosis: retention cysts, mucosal hyperplasia, cystadenomas, and cystadenocarcinomas. A mucocele of benign etiology is adequately treated by a simple appendectomy.

In a small percentage of patients, the mucocele occurs along with gelatinous ascites known as pseudomyxoma peritonei. Pseudomyxoma peritonei can be associated with either ovarian or appendiceal mucinous tumors, usually of a malignant nature. When pseudomyxoma peritonei is present, survival is significantly decreased probably due to the association with the ascites and malignancy. The initial surgical procedure should include cytoreductive surgery along with a right hemicolectomy for cystadenocarcinoma of the appendix and bilateral oophorectomy, hysterectomy, and appendectomy for ovarian cystadenocarcinoma. A recent review found that even with aggressive surgical treatment, there is a 57 percent incidence of local recurrence of pseudomyxoma peritonei from an appendiceal primary site. Once a recurrence is detected, death usually ensues from progressive bowel obstruction and renal failure. Adjuvant therapy including radiation and intraperitoneal and systemic chemotherapy has been recommended, but its efficacy and role are unclear.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 28 - Liver

Seymour I. Schwartz

ANATOMY

The liver constitutes approximately one-fiftieth of total body weight. Its size reflects the complexity of its functions. True division into right and left lobes (hemilivers) is in line with the fossa for the inferior vena cava posteriorly and the gallbladder fossa anteroinferiorly (Cantlie's line) (Fig. 28-1). Based on the distribution of the intrahepatic branches of the hepatic artery, portal vein, and bile ducts, the right lobe is divided into an anterior section and a posterior section, and the left lobe is divided into a medial section and a lateral section. Couinad proposed a functional division of the liver related to the hepatic venous drainage (Fig. 28-2). The liver is divided into two "livers" (lobes) by the portal scissura in which the middle hepatic vein courses. The right lobe is divided into two "sectors" by the right hepatic vein. The right posterolateral "sector" contains segment VI anteriorly and segment VII posteriorly. The right anterolateral "sector" contains segment V anteriorly and segment VIII posteriorly. The left lobe is divided by a line containing the left hepatic vein. The left anterior sector is divided by the umbilical fissure into segment IV, the anterior part of which is the quadrate lobe, segment III, which is the anterior part of the left lobe. The posterior portion is segment II. The dorsal segment I, in regard to its vascularization, is independent of the portal division and the three main hepatic veins (Fig. 28-3).

Biliary Drainage

Each sector is drained by a major segmental duct formed by the confluence of subsegmental draining structures. The anterior and posterior sectoral ducts in the right lobe join to form the right hepatic duct, while the medial and lateral segmental ducts in the left lobe terminate in the left hepatic duct, which joins the right duct to form a common hepatic duct in the porta hepatis. This lies anteriorly in relation to other structures in the area.

Blood Supply

The afferent blood supply to the liver arises from two sources: (1) the hepatic artery, which carries oxygenated blood and accounts for approximately 25 percent of hepatic blood flow, and (2) the portal vein, which accounts for approximately 75 percent of hepatic blood flow and drains the splanchnic circulation. The common hepatic artery originates from the celiac axis and, after contributing the gastroduodenal and right gastric artery, ascends in the hepatoduodenal ligament to the left of the common bile duct and anterior in the portal vein. It bifurcates into a right and left branch to the left of the line dividing the right and left lobes. The major right hepatic artery originates from the superior mesenteric artery in 17 percent of people. This vessel usually courses to the liver in the hepatoduodenal ligament to the right of the common duct and anterior to the portal vein. Intrahepatic anastomoses between the right and left hepatic arteries do not occur. The cystic artery is usually an extrahepatic branch of the right hepatic artery.

The portal venous system contains no valves (Fig. 28-4). It returns to the liver the blood that the celiac, superior mesenteric, and inferior mesenteric arteries supply to the gastrointestinal tract, pancreas, and spleen. The vessel is formed behind the pancreas, at the level of L 1 to L2, by the confluence of the superior mesenteric and splenic veins and, at times, the inferior mesenteric vein. The portal vein resides posteriorly in relation to the hepatic artery and bile duct in the hepatoduodenal ligament but, in rare instances, is located anterior to the pancreas and first portion of the duodenum, in which circumstance it is frequently associated with a partial or complete situs inversus and is subject to injury during cholecystectomy or gastrectomy. In the porta hepatis, the vein divides into two branches, which course to each lobe. The average length of the main portal vein is 6.5 cm, and the average diameter is 0.8 cm.

The hepatic venous system (Fig. 28-5) begins in the liver lobules as a central vein into which the sinusoids empty. The central veins unite to form sublobular veins, which in turn fuse to form collecting veins. The collecting veins gradually increase in size by joining other large intrahepatic collecting channels, which coalesce to form the three major hepatic veins. The hepatic venous tributaries are intersectoral in position. The major hepatic veins are classified as right, left, and middle. The right hepatic vein drains the entire posterior portion as well as the superior area of the anterior portion of the right lobe. The left hepatic vein drains the entire area to the left of the umbilical fissure. The inferior areas of the medial and anterior portions of the two lobes are drained by the middle vein. In human beings there are no valves in the hepatic venous system. Total hepatic blood flow can be measured by means of hepatic vein catheterization and use of the Fick principle. The average value is 1500 mL/min/1.73 m 2 of body surface.

LIVER FUNCTION

The liver consists of four physiologic-anatomic units that are interrelated:

- (1) The circulatory system. A dual blood supply nourishes the liver and acts as a vehicle for material absorbed from the intestinal tract to be utilized in the metabolic pool. Blood vessels are accompanied by lymphatics and nerve fibers that contribute to the regulation of blood flow and intrasinuosoidal pressure.
- (2) Biliary passages. These serve as channels of exit for materials secreted by liver cells, including bilirubin, cholesterol, and detoxified drugs. This system originates with the Golgi apparatus adjacent to the microvilli of the bile canaliculi and eventually terminates in the common bile duct.
- (3) The reticuloendothelial system. This system has 60 percent of its cellular elements in the liver and includes the phagocytic Kupffer cells and endothelial cells.
- (4) The functioning liver cells (hepatocytes), which are capable of wide variation of activity. The metabolic pool in the liver serves the needs of the entire body. The cell performs both anabolic and catabolic activities, secretes, and stores. The large amount of energies required for these transformations result from the conversion of adenosine triphosphate (ATP) to adenosine diphosphate (ADP). A second source is the aerobic oxygenation in the metabolic pool via the tricarboxylic acid cycle of Krebs.

Function Tests

The so-called liver function tests (Table 28-1) evaluate liver activity by assessing the degree of functional impairment. They do not provide a pathologic diagnosis, and the extreme functional reserve of the organ occasionally produces normal results in the face of significant lesions. Many of these tests do not measure a specific function of the liver, and other organ systems may be implicated. False-positive results for each of the tests are found in about 10 percent of hospital controls. False-negative tests also occur in about 10 percent of most tests.

Proteins

Hepatic cells are responsible for the synthesis of albumin, fibrinogen, prothrombin, and other factors involved in blood clotting. A reduction of serum albumin is one of the most accurate reflections of the extent of liver disease and the effects of medical therapy. Because the half-life of albumin is 21 days, impairment of hepatic synthesis must be present for over 3 weeks before abnormalities are noted. The correlation between total protein and disease of the liver is not as close as that between the serum albumin level and liver disease, since albumin is produced only by hepatic cells and a reduction is frequently compensated for by an increase in the level of globulin.

Carbohydrates and Lipids

Glycogenesis, glycogen storage, glycogenolysis, and the conversion of galactose into glucose all represent hepatic functions. Hypoglycemia is a rare accompaniment of extensive hepatic disease, but the amelioration of diabetes in patients with hemochromatosis is considered an indication of neoplastic change. The more common effect of hepatic disease is a deficiency of glycogenesis with resulting hyperglycemia. A hepatic enzyme system is responsible for the conversion of galactose into glucose, and abnormal galactose tolerance tests are seen in hepatitis and active cirrhosis. In rare instances, a familial deficiency in this enzyme system accounts for spontaneous galactosemia accompanied by an obstructive type of jaundice that appears after the first week of life and subsides when lactose is removed from the diet.

Synthesis of both phospholipid and cholesterol takes place in the liver, and the latter serves as a standard for the determination of lipid metabolism. The liver is the major organ involved in the synthesis, esterification, and excretion of cholesterol. In the presence of parenchymal damage, both the total cholesterol and percentage of esterified fraction decrease. Biliary obstruction results in a rise in cholesterol, and the most pronounced elevations are noted with primary biliary cirrhosis and the cholangiolitis accompanying toxic reactions to phenothiazine derivatives.

Enzymes

The three enzymes that achieve abnormal serum levels in hepatic disease and have been studied widely are alkaline phosphatase, serum glutamic oxalacetic transaminase (SGOT), and serum glutamic pyruvic transaminase (SGPT). SGOT is present in the liver, myocardium, skeletal muscles, kidney, and pancreas. Cellular damage in any of the above-mentioned tissues results in elevation of the serum level. In reference to the liver, the most marked increases accompany acute cellular damage regardless of cause, and extremely high levels are noted in patients with hepatitis. SGOT is only moderately increased in cirrhosis and biliary obstruction. SGPT is more particularly applicable to the evaluation of liver disease, since the hepatic content greatly exceeds myocardial concentration. Elevations accompany acute hepatocellular damage. Lactic acid dehydrogenase (LDH) levels also may be elevated.

Serum alkaline phosphatase provides an elevation of the patency of the bile channels at all levels, intrahepatic and extrahepatic. Elevation is demonstrated in 94 percent of patients with obstruction of the extrahepatic biliary tract due to neoplasm and 76 percent of those in whom the obstruction is caused by calculi. Intrahepatic biliary obstruction and cholestasis also cause a rise in the enzyme level. In the presence of space- occupying lesions such as metastases, primary hepatic carcinoma, and abscesses, the alkaline phosphatase level is also increased. The overall correlation between metastatic carcinoma of the liver and an elevated enzyme level is as high as 92 percent. Sixty percent of patients with primary hepatic carcinoma also demonstrate a significant increase. Granulomatous and infiltrative lesions such as sarcoidosis, tuberculosis, and lymphoma are irregularly associated with mild to moderate increases in the alkaline phosphatase. Elevation of the serum level of this enzyme is also associated with diseases that have as a common denominator increased osteoblastic activity. 5[pr]-Nucleotidase is a phosphatase that catalyzes hydrolysis of nucleotides. Levels are elevated in hepatobiliary disease.

Dye Excretion

The hepatic removal of dyes from the circulation depends on hepatic blood flow, hepatocellular function, and biliary excretion. The presence of jaundice produces a disproportionate indocyanine green or sulfobromophthalein (Bromsulphalein) retention, and fever, shock, hemorrhage, and recent surgical treatment may all result in increased levels. An increased retention is associated with acute cellular damage and is also noted in patients with cirrhosis, carcinoma, and chronic passive congestion. Since the rate of disappearance from the blood is constant, hepatic blood flow can be determined by injecting the dye at a rate that will maintain a constant blood level and applying Fick's principle to the blood removed from a catheterized hepatic vein. An intestinal xenon technique has been shown to provide an accurate method of measuring portal vein and total hepatic blood flow. Determination of the fecal excretion of the radioactive material has been applied to the differential diagnosis of obstructive jaundice, particularly in the establishment of the diagnosis of congenital atresia of the bile ducts, in which case HIDA or DISIDA fails to appear in the stool.

Coagulation Factors

In liver disease, multiple coagulation defects may occur. Two mechanisms contribute to the deficiency of coagulation factors: (1) in obstructive jaundice, the bile source required for the absorption of the fat-soluble vitamin K results in a decreased synthesis of prothrombin, and (2) hepatocellular dysfunction is accompanied by an inability of the liver to synthesize prothrombin. Abnormal values for prothrombin time have been noted in a variety of hepatic diseases with parenchymal damage, and determination is particularly applicable in the evaluation of patients undergoing liver biopsy or surgical procedures. An increase in prothrombin time subsequent to the injection of parenteral vitamin K is used as an indication of hepatic function and suggests obstructive jaundice. Decreases in factors V, VII, and IX and fibrinogen also have been noted in hepatic disease. Cirrhosis also has been associated with an increased fibrinolysis due to defective synthesis of fibrinolytic inhibitors and delayed removal of plasminogen activators.

Quantitative Tests

Maximal rate of urea synthesis after an oral challenge with casein or an intravenous bolus of amino acids provides an assessment of hepatic function, as does the determination of galactose elimination capacity.

Bile Pigment Metabolism

See the section Jaundice in Chap. 22.

SPECIAL STUDIES

Needle Biopsy of the Liver

This is the one preoperative study that provides a pathologic diagnosis. It depends on an area of tissue measuring 1 to 4 cm in length and containing approximately 5 to 20 lobules representing a general anatomic change. Close to 100 percent accuracy has been demonstrated for both posthepatic and postnecrotic cirrhosis. Intrahepatic cholestasis, hepatitis, and cellular degeneration resulting from toxicity are all diffuse lesions and readily diagnosed. Focal lesions, such as neoplasms, granulomas, and abscesses, may be missed, but the correlation between needle biopsy and operative and autopsy findings is high. In the case of focal nodular hyperplasia, a needle biopsy often reports normal structure made up of normal hepatocytes and fails to diagnose the lesion because the classic central scar is missed. Occasionally, exsanguination has been reported, particularly in patients with hemangioma, and pain, pneumothorax, hemorrhage, and bile peritonitis are the complications to be considered. If hemangioma is a considered diagnosis, needle biopsy is contraindicated.

Ultrasonography, Computed Tomography, Magnetic Resonance Imaging, and Scintigrams

Preoperative *ultrasonography (US)* has its highest yield in defining hepatic abscesses, cystic lesions, and most hemangiomas. Intraoperative US detects lesions unidentified by preoperative studies and demonstrates the intraparenchymal vascular anatomy, thus facilitating resection. US and duplex scanning can determine the patency of a portasystemic shunt noninvasively.

Computed tomography (CT) is used with or without vascular enhancement. Intravenous contrast material improves the diagnostic yield over plain CT, but the best results in definition of parenchymal lesions are obtained with intraarterial infusion scans. A catheter is placed in the common hepatic artery, and the scan is taken while the artery is infused. The tumor lights up. The catheter is then positioned in the superior mesenteric artery, and a scan is taken while it is infused and again during the venous phase. Lesions present as radiolucencies. For most lesions, CT provides the best results.

Magnetic resonance imaging (MRI) is particularly applicable for assessing vascular lesions. The T₂ phase provides a characteristic demonstration of a hemangioma. Scintillation scanning also can detect intrahepatic focal lesions. Using tagged red cells or technetium (^{99m}Tc)-sulfur colloid, the diagnosis of hemangioma can be established.

A prospective analysis of laboratory tests and imaging studies has shown greater than 65 percent accuracy in the detection of hepatic lesions. No combination of laboratory tests increased this accuracy. If the laboratory tests were used with one of the imaging studies, the accuracy was increased to 76 percent. The use of all the liver imaging tests and laboratory tests lowers the accuracy and needlessly increases the expense.

Angiography

Since hepatic tumors, both primary and metastatic, depend on an arterial circulation, unusual vascular patterns are also detected by injection of the hepatic artery with radiopaque material. Unusual arrangements of the arteries and “tumor staining,” analogous to that found in cerebral and osseous neoplasms, may be noted. Angiography also demonstrates extrahepatic vascular anatomy and provides a “road map” for the surgeon. Angiography is currently used less frequently.

Measurements of Portal Pressure and Evaluation of Portal Circulation

See Portal Hypertension later in this chapter.

TRAUMA

(See Chap. 6, Trauma.)

HEPATIC ABSCESSES

Hepatic abscesses are related to two distinct groups of pathogens, pyogenic bacteria and *Entamoeba histolytica*. Distinctive features in the clinical manifestations and therapy of these two variations necessitate separate consideration.

Pyogenic Abscesses

Incidence

The lesion is present in 0.36 percent of autopsies. The highest percentage of cases occur in the sixth and seventh decades, and there is no predilection for either sex.

Etiology

Pyogenic abscesses of the liver result from (1) ascending biliary infection, (2) hematogenous spread via the portal venous system, (3) generalized septicemia with involvement of the liver by way of the hepatic arterial circulation, (4) direct extension from intraperitoneal infection, or (5) other causes, including hepatic trauma. Recently, the most frequent antecedent cause has been cholangitis secondary to calculi or carcinoma in the extrahepatic biliary duct system. In one series, over 80 percent of patients with a pyogenic liver abscess had an underlying hepatobiliary or pancreatic cancer. The second most common cause is related to generalized septicemia, while the portal venous route of infection has decreased in importance. Pylephlebitis occurs in 0.05 percent of cases of acute appendicitis and 3 percent of patients with perforated appendicitis. No segment of the intestine drained by the portal venous system can be excluded as a possible cause, and the incidence associated with acute diverticulitis is as high as that of appendicitis. There is an increased incidence in immunocompromised patients. There has been an increase in the percentage in which no cause is apparent. These account for about 20 percent of cases.

Cultures are positive in over 90 percent of cases in which they were obtained. *Escherichia coli*, *Klebsiella*, and *Streptococcus* are the organisms most commonly isolated. *Staphylococcus* and *Pseudomonas* are occurring more frequently, and mixed bacterial and fungal abscesses are noted in about 25 percent of patients.

Pyogenic abscesses may be solitary, multiple, and multilocular. Single and multiple abscesses occur with equal frequency. When a single abscess is present, it is usually located in the right lobe.

Clinical Manifestations

Since most pyogenic hepatic abscesses are secondary to other significant infections, it is difficult to delineate a pathognomonic symptom. Fever is the most common symptom, and a “picket fence” configuration of the temperature chart generally has been noted. Fever is frequently accompanied by chills, profuse sweating, nausea, vomiting, and anorexia. Pain is a late symptom and is more common with large, solitary abscesses. Liver enlargement is noted in 30 to 60 percent of cases. Hepatic tenderness is present in half the patients. Jaundice is related to the causative pathology.

Diagnostic Studies

Leukocytosis with white blood cell counts ranging between 18,000 and 20,000/mm³ is usual. Half the patients are anemic. Positive blood cultures are demonstrated in approximately 40 percent of patients, the most significant yields accompanying abscesses secondary to systemic septicemia. Liver function tests are not diagnostic, but elevation of the alkaline phosphatase level is the most frequent abnormality. Hypoalbuminemia is inconsistent. Characteristically, radiographs reveal an elevation and immobility or restriction of motion of the right leaf of the diaphragm. There is also obliteration of the right cardiophrenic angle on the posteroanterior chest film and the anterior costophrenic angle on the lateral film. Abscesses produced by gas-forming microorganisms are associated with air-fluid levels in the liver. The *CT* is the most accurate radiographic study (over 90 percent) as contrasted with ultrasound (80 percent) and radionuclide (70 percent) (Figs. 28-6 and 28-7).

Treatment

This is based on appropriate antibiotic therapy combined with drainage in selected cases. Intravenous antibiotics are usually administered for 2 weeks, followed by 1 month of oral therapy. The abscesses may be drained percutaneously under *US* or CT control. Several series have reported success rates of 80 percent using this technique. Equivalent results have been reported for percutaneous and surgical drainage. The route of surgical access depends on the position of the abscess and may be transthoracic or transabdominal (Figs. 28-8 and 28-9). Because of the availability of antibacterial agents, transperitoneal drainage is no longer associated with prohibitive morbidity or mortality. In a small group of patients with multiple abscesses confined to a lobe, treatment is best managed by resection (Fig. 28-10).

Prognosis and Complications

Percutaneous and surgical drainage have been associated with mortality rates ranging from 7.5 to 20 percent. With multiple abscesses, the rate is significantly increased. The likelihood of death for patients treated with antibiotics alone is approximately 50 percent.

Amebic Abscesses

Incidence

E. histolytica has been found wherever surveys have been made on the human population from northern Canada to the Straits of Magellan. Amebic abscess of the liver is a disease of the middle-aged adult and predominates in males with a 9:1 ratio. The concept of racial immunity is invalid.

Pathology

Amebas reach the liver by way of the portal venous system from a focus of ulceration in the bowel wall. Hepatic involvement is usually a large, single abscess containing liquefied material with a characteristic reddish brown “anchovy paste” fluid. The lesions are usually single and occur in the right lobe of the liver, either near the dome or on the inferior surface in juxtaposition to the hepatic flexure. The wall is only a few millimeters thick and consists of granulation tissue with little or no fibrosis. Microscopically, three zones are recognized: a necrotic center, a middle zone with destruction of parenchymal cells, and an outer zone of relatively normal hepatic tissue in which amebas may be demonstrated.

Clinical Manifestations

Abscesses become evident when they cause generalized systemic disturbances coupled with symptoms and signs of hepatic involvement. The chief complaints are fever and liver pain. Pain is present in 88 percent of patients, and the pattern is related to the location of the hepatic abscess. With pain and tenderness over the right lower intercostal spaces, there may be associated bulging and pitting edema of the subcutaneous tissue. Superior surface abscesses result in pain referred to the right shoulder, whereas abscesses in the bare area, which have no contact with the serosal surface, are latent as far as pain is concerned. Left lobe abscesses present as a painful epigastric swelling.

Fever accompanied by chills and sweating is present in over three-quarters of patients, but the temperature does not reach the levels resulting from pyogenic abscesses unless there is secondary infection. One-third to one- half of adults offer a history of antecedent diarrhea, whereas in children grossly bloody, mucous stool occurs more frequently. Tender hepatomegaly is an almost constant feature. Clinical jaundice is relatively rare.

Diagnostic Studies

Patients with acute disease show no anemia but an appreciable degree of leukocytosis, whereas those with prolonged illness have anemia with less marked leukocytosis. Amebas are found in the stool of only 15 percent of cases collected from the literature. Liver function tests are not helpful in establishing the diagnosis. The indirect hemagglutination test is almost always positive. Radiographic findings are similar to those described for pyogenic abscesses. Scintillography, *US*, and angiography also have helped to localize the lesion.

Diagnosis is frequently established by aspiration of the abscess cavity, a relatively innocuous procedure. Although the “anchovy paste” aspirate is considered pathognomonic, the abscess content may be creamy white, even though there is no secondary bacterial infection. Amebic trophozoites are demonstrated in the aspirate of fewer than one-third of patients.

Complications

The most common complication is secondary infection, which occurs in approximately 22 percent of patients. Rupture of the amebic abscess accounts for the next most common group of complications. The direction of rupture is reproduced in Fig. [28-11](#). Pleuropulmonary complications occur in 20 percent of patients. This is usually the result of direct extension of the hepatic process. The most serious route of rupture is into the pericardial cavity, and this is usually secondary to extension of an abscess in the left lobe. Rupture into the peritoneal cavity or into an intraabdominal viscus occurs in 6 to 9 percent of the patients.

Treatment

This consists of administration of amebicidal drugs combined with aspiration or surgical drainage when indicated. The initial approach is usually conservative and directed toward eradicating the parasite from the intestinal tract, liver, and abscess itself. In general, the patient is not considered for surgical treatment until the intestinal phase is controlled. Metronidazole, which acts in both the hepatic and intestinal sites, has replaced emetine and chloroquine. Both the hepatic and intestinal infections generally have been cured by 400 mg three times a day for 4 days, occasionally combined with closed aspiration. In endemic areas, most patients are managed with drugs alone and rarely require aspiration. A single dose of 2.5 g combined with aspiration also has had dramatic results.

Surgical Procedures

The indications for aspiration are (1) the persistence of clinical manifestations following a course of amebicidal drugs, (2) clinical or radiographic evidence of a hepatic abscess, and (3) absence of findings that would suggest secondary infection of a liver abscess. Drug therapy should be instituted several days prior to aspiration. In most instances, aspiration is not required, and the abscess will be resolved by the medication. There is no indication for injection of any drug directly into the abscess cavity. In the absence of localizing signs, the preferred route for percutaneous drainage is through the ninth or tenth interspace between the anterior and posterior axillary lines. Once an abscess has been demonstrated to be secondarily infected, open drainage is the treatment of choice.

Prognosis

This depends on the relative virulence of the organism and the resistance of the host, the stage of infection, the multiplicity of abscesses, and the presence of complications. In uncomplicated cases, the mortality rate is less than 5 percent, whereas with complications a 43 percent mortality has been reported.

CYSTS AND BENIGN TUMORS

Nonparasitic Cysts

These lesions may be single, multiple, diffuse, localized, unilocular, or multilocular. They include (1) blood and degenerative cysts, (2) dermoid cysts, (3) lymphatic cysts, (4) endothelial cysts, (5) retention cysts, consisting of (*a*) solitary retention cysts and (*b*) multiple retention cysts (polycystic disease), and (6) proliferative cysts (cystadenomas). Autopsy incidences of approximately 0.15 percent have been reported. One percent of *CT* scans reveal intrahepatic cysts. The clinically apparent nonparasitic solitary cysts occur more frequently in women in the fourth, fifth, and sixth decades, at an average age of 52 years. Polycystic hepatic disease also occurs much more frequently in women.

Pathology

Solitary nonparasitic cysts are usually located in the right lobe of the liver. The cyst content is a clear, watery material, and characteristically, the cysts have a low internal pressure, in contrast to the high tension in parasitic cysts. Occasionally, the fluid is yellowish brown, suggesting necrosis of adjacent parenchyma. Polycystic disease of the liver has a honeycomb appearance with multiple cavities, and the lesions commonly are distributed throughout the entire liver. At times, however, one lobe, more frequently the right, is preferentially involved. Unlike the solitary nonparasitic cyst, polycystic disease of the liver (Fig. [28-12](#)) frequently is associated with cystic involvement of other organs; 51.6 percent of polycystic livers are associated with polycystic kidneys. Conversely, the incidence of hepatic cysts in patients with known polycystic renal disease varies between 19 and 34 percent. Polycystic livers have been implicated as a rare cause of portal hypertension and also have been associated with atresia of the bile ducts, cholangitis, and hemangiomas.

Traumatic cysts are usually single, are filled with bile, and contain no epithelial lining. Cystadenomas are grossly smooth, encapsulated, and lobular and contain a mucoïd material. They are lined by a proliferative columnar epithelium.

Clinical Manifestations

Both solitary and polycystic lesions grow slowly and are relatively asymptomatic. A painless right upper quadrant mass is the most frequent complaint, and when symptoms occur, they are usually related to pressure on adjacent viscera, such as early satiety. Acute abdominal pain may accompany the complications of torsion, intracystic hemorrhage, or intraperitoneal rupture. Physical examination may reveal the mass, and the kidneys may be palpable. Jaundice is rare. Liver function tests usually demonstrate no abnormality. Scintillography, *CT* scan, *US*, and arteriography have been used to define the intrahepatic position of the mass, and peritoneoscopy may be diagnostic (Fig. [28-13](#)).

Treatment

Asymptomatic solitary nonparasitic cysts and polycystic disease of the liver require no treatment. Large, solitary symptomatic cysts usually can be managed electively unless there is rupture, intracystic hemorrhage, or torsion. Patients have been managed successfully with radiographically controlled percutaneous catheter drainage, at times reinforced with injection of a sclerosing solution such as alcohol. This procedure is frequently associated with recurrence. Permanent resolution can be effected by a simple operation in which the cyst is widely unroofed back to its junction with normal hepatic parenchyma. This can be accomplished laparoscopically. In the case of significant intracystic bleeding, cystectomy might be required. Internal drainage into the intestine is indicated only if there is erosion into a major hepatic duct that cannot be repaired.

In patients with significant symptoms attributed to the mass effect of polycystic liver, palliation can be achieved by nonanatomic resection and wide fenestration of the larger cysts.

Prognosis

The prognosis of polycystic disease is essentially that of the accompanying renal disease. Hepatic failure, jaundice, and the manifestations of portal hypertension are rare. The mortality rate for surgically treated nonparasitic cysts of the liver approaches zero.

Hydatid Cysts

Hydatid disease (*echinococcosis*) is characterized by worldwide distribution and frequent hepatic involvement. The incidence among human beings depends on the incidence in intermediate hosts, including sheep, pigs, and cattle. The southern half of South America, Iceland, Australia, New Zealand, and southern parts of Africa are regarded as intensive endemic areas. Most cases reported in the United States have occurred in immigrants from Greece and Italy.

Pathology

The most common unilocular hydatid cyst is caused by *Echinococcus granulosus*, whereas the alveolar type is caused by *E. multilocularis*. Approximately 70 percent of hydatid cysts are located in the liver, and in one-quarter to one-third of these cases there are multiple cysts. The right lobe is affected in 85 percent of patients. Cysts are usually superficial and are composed of a two-layer laminated wall, an inner germinative membrane, and an outer adventitia. The two membranes are in close contact with each other but are not linked. The fluid in the hydatid cyst has a high pressure of approximately 300 mL of water and is colorless, opalescent, and slightly alkaline. Inside the main hydatid vesicle, daughter cysts are usually found. Extension is commonly into the peritoneal cavity, but progressive intrahepatic expansion may result in the replacement of liver parenchyma.

In contrast to the unilocular hydatid cysts, the alveolar hydatid is a growth without a capsule and with a tendency toward multiple metastases. As growth progresses, the center becomes necrotic, and the periphery invades the blood vessels and lymph channels. The causative agent of this lesion is found more frequently in the colder regions of Alaska, Russia, and the Alps.

Complications

Intrabiliary rupture represents the most common complication and occurs in 5 to 10 percent of cases. Suppuration, the second most common complication, is caused by bacteria from the biliary tract. The formation of the purulent material results in death of the parasite and conversion into a pyogenic abscess. Intraperitoneal rupture results in the showering of hydatid fluid, brood capsules, and scolices into the peritoneum, leading to transient peritoneal irritation of varying intensity. Usually, the reproductive elements survive and initiate the formation of new cysts (*secondary echinococcosis of the peritoneum*). Cysts located in the superior portion of the liver tend to grow craniad into the pleural cavity and become intrathoracic. These can be differentiated from primary pulmonary cysts by the presence of daughter cysts and bile pigments. Empyema and bronchopleural fistula must result.

Clinical Manifestations

Patients with simple or uncomplicated multivesicular cysts are usually asymptomatic. When symptoms occur, they are caused by pressure on adjacent organs. Abdominal pain and tenderness are the most common complaints, followed by a palpable mass. A tumor, which is palpable in 70 percent of the patients, or diffuse hepatic enlargement in a patient who has lived in an endemic region is cause for suspicion. The so-called hydatid thrill and fremitus are quite rare. Jaundice and ascites are uncommon. With secondary infection, tender hepatomegaly, chills, and spiking temperatures occur. Urticaria and erythema offer evidence of a generalized anaphylactic reaction. With biliary rupture, the classic triad of biliary colic, jaundice, and urticaria may be noted. Vomiting with passage of hydatid membranes in the emesis (*hydatidemesia*) and passage of membranes in the stool (*hydatidenteria*) also may occur. The complication of intraperitoneal rupture is heralded by abdominal pain and signs of anaphylactic shock. Intrathoracic rupture is associated with shoulder pain and cough initially productive of a frothy blood-stained fluid that subsequently becomes stained with bile. Membranes are intermittently expectorated in 80 percent of these patients.

Diagnostic Studies

Radiographically, an unruptured cyst presents as a round, reticulated, calcified shadow in the liver (Fig. 28-14). Secondary infection with gas-producing organisms might be confused with daughter cysts. With intrabiliary rupture, gas is noted in the remaining cyst cavity. CT scans furnish useful information and correlate well with operative findings (Fig. 28-15). Eosinophilia is the least reliable of immunologic responses, being present in only 25 percent of all patients. The indirect agglutination test is positive in about 85 percent of patients; the complement fixation test is slightly less sensitive. This reaction becomes negative 2 to 6 months after removal of the cyst. Casoni's skin test is positive in approximately 90 percent of patients, and the reaction can be obtained years after surgical removal of the cyst or after the parasite has died.

Treatment

Small calcified cysts in patients with negative serologic test results need no treatment. Treatment is surgical, since there is no response to drug administration. Therapy consists of removal of the cyst contents without contaminating the patient, followed by appropriate management of any remaining cavity. Since the hydatid fluid is under high tension, evacuation and sterilization are carried out initially with a scolical agent such as hibitane, alcohol, or hypertonic saline. Following evacuation and irrigation, primary closure may effect cure. External drainage and marsupialization are accompanied by high complication rates and prolonged drainage. Removal of the parasite is accomplished by excision of the hydatid vesicle using the natural cleavage plane that exists between the germinative layer and adventitia. Omentoplasty provides a method of managing the cavity successfully (Fig. 28-16). Total removal of the cysts, including the adventitial layer, also may be performed. Partial hepatectomy with controlled hepatic resection has been advised for larger and multiple cysts. Marsupialization and partial hepatectomy are the alternatives for large or injected cysts. In uncomplicated cases, the results of surgical treatment are excellent, and the postoperative mortality is less than 5 percent. With intrabiliary rupture, marsupialization should be accompanied by drainage of the bile duct if there is associated obstruction. Rupture into the peritoneal cavity is treated by laparotomy and thorough cleansing, although it is frequently impossible to prevent secondary contamination. Intrathoracic rupture generally can be controlled by evacuating and draining the hepatic cysts. Alveolar disease of the liver was inevitably fatal, but more recently, satisfactory results have been obtained with extensive hepatic resection.

Benign Tumors

Hamartoma

Hamartomas are composed of tissues normally present in the organ but arranged in a disorderly fashion. The lesions vary from minute nodules to large tumors and are rarely of clinical significance. Large mesenchymal hamartomas have presented as rapidly growing abdominal masses in infants and children. US and CT are the most useful diagnostic tests. Grossly, the tumors are firm, nodular, and located immediately beneath the surface of the liver, and they may be solitary or multiple. They are generally well encapsulated and often cystic. With lesions of clinical significance, surgical excision is generally indicated. Deeply located lesions should be left alone after histologic diagnosis has been established because they do not grow rapidly and do not undergo malignant transformation.

Adenoma

Hepatic adenoma is a benign tumor that was rarely noted before the introduction of oral contraceptives in the 1960s. In an accumulation of data from many series, more than 60 percent of patients were exposed only to mestranol; an additional 20 percent of patients were exposed to a product that included mestranol. More than half the patients used oral contraceptives continuously for durations longer than 5 years; lesions have become manifest even after discontinuation of the drug. These lesions also develop during pregnancy, in patients with diabetes mellitus or glycogen storage disease, and rarely, as adenomatosis, which is defined as more than 10 adenomas in an otherwise normal hepatic parenchyma. Adenomatosis occurs with equal frequency in men and women.

In patients who were contraceptive users, hepatic adenomas tend to be larger and have higher rates of intratumoral and intraperitoneal hemorrhage. Bleeding also has occurred during pregnancy. Transformation of the adenomas into hepatocellular carcinoma has been documented; this occurs more frequently in patients with adenomatosis.

The diagnosis typically is made in women of childbearing age. Eighty percent are symptomatic, with pain or mass effect generally related to intratumoral or intraperitoneal bleeding. The latter may cause shock. The hepatic lesion usually can be identified by US or CT (Fig. 28-17). At times, a small lesion is not detected, but the diagnosis is suggested by a subglissonian hematoma. Percutaneous biopsy is contraindicated because it is associated with a high risk of bleeding.

Although regression and disappearance of hepatic adenomas have been reported after the discontinuation of contraceptives, the potential for bleeding and malignant transformation, along with the current safety of removing the lesions, favors routine resection. Deaths associated with elective resections are rare; an 8 percent mortality is reported for emergency resection. The lesions are readily identified by their pale yellow color and homogeneous appearance. They often can be removed by enucleation with a narrow rim of normal hepatic parenchyma. Because of the diffuse nature of adenomatosis and its potential for malignancy, transplantation has been used.

Focal Nodular Hyperplasia (FNH)

This is a benign lesion that usually occurs in women of reproductive age. Many believe that it is neither a neoplasm nor a hamartoma but a reaction to injury or a hyperplastic response to a preexisting spider-like vascular malformation. There is no consistent relationship to oral contraceptive use. US and CT frequently fail to define the lesion because it is isodense. The angiogram demonstrates a typical sunburst, hypervascular pattern. The patients are almost always asymptomatic, and the lesions frequently are detected at celiotomy. Spontaneous bleeding rarely occurs. The cut surfaces are tan and demonstrate a characteristic central stellate scar. To make a histologic diagnosis, a deep biopsy is necessary to define the scar because the hepatic cell structure is otherwise normal (Fig. 28-18). FNH was regarded as a possible precursor to fibrolamellar carcinoma, but unlike the latter, it contains no neurotensin. Resection is occasionally indicated for significant pain ascribed to the lesion and rarely for spontaneous rupture.

Hemangioma

Hemangioma is the most common nodule in the liver, and the liver is the internal organ most frequently affected with this lesion. The tumor occurs five times more frequently in women than in men. These are occasionally associated with FNH and cysts of the liver and pancreas. Malignant degeneration does not occur. The hemangioma must be distinguished from hemangioendothelioma or diffuse hemangiomatosis. The latter consists of widespread multicentric lesions accompanied by vascular involvement of the skin and occurs in children with clinical manifestations in the first week of life.

Most lesions are detected radiographically, are less than 5 cm in diameter, and are asymptomatic. Larger lesions can be painful and palpable. Platelet trapping in giant hemangiomas can result in thrombocytopenia. Growth of the lesions as a consequence of pregnancy or estrogen therapy has been varied. Most small lesions followed ultrasonographically during pregnancy grow minimally, but occasionally, significant increases in size occur. Rupture of a hemangioma is a rare event. In infants, large hemangiomas often are associated with high-output cardiac failure.

Hemangiomas can present a characteristic pattern on US and CT (Fig. 28-19 A). The diagnosis can be made on a T₂-weighted MRI or scintigram with ^{99m}Tc-labeled red blood cells. The arteriogram demonstrates pooling in small peripheral vessels (Fig. 28-19 B). Percutaneous biopsy often provides the histologic diagnosis but is rarely indicated because it is associated with the complication of bleeding.

In infants with high-output congestive heart failure, supportive therapy, at times with the addition of steroids, is effective, and the lesion regresses. In rare cases, hepatic artery embolization or ligation may be required to reverse cardiac failure. In adults, most hepatic hemangiomas should not be excised. Even large lesions followed for long periods show no notable increase in size or clinical manifestations. The potential for rupture is minimal and should not constitute an indication for excision. Pain, mass effect, significant growth, platelet trapping, and early rupture are indications for surgical excision. Few reports have documented reduction in size with radiation therapy, and hepatic arterial ligation is rarely effective in adults. Resection is the therapy of choice when indicated. Large lesions may necessitate anatomic resection, but enucleation is often feasible.

MALIGNANT TUMORS

Primary Carcinoma

Incidence

Although the disease is rare in people of western Europe and North America, primary carcinoma is remarkably common among the aboriginal inhabitants of Africa and in many parts of Asia. Postmortem rates in the United States average 0.27 percent, whereas in Africa the postmortem rate is 1.1 percent, hepatic carcinomas representing 17 to 53 percent of all cancers. In some Asian countries, the autopsy incidence is 2.5 percent.

Primary carcinoma of the liver occurs with greater frequency in males. In whites it is rare before age 40, whereas in Africans and Asians the affliction usually occurs before age 40. American blacks do not exhibit a predisposition toward the disease. By contrast, a higher incidence is present in Chinese subjects even after they have changed their habitation. In children, the first appearance of the neoplasm is usually before age 2, and primary carcinoma of the liver represents the most common carcinoma in the first few years of life. Hepatoblastoma usually affects children less than 2 years old; the male-to-female ratio is 6:1. This lesion also occurs in adults. Fibrolamellar carcinoma, a variant of hepatocellular carcinoma, has a propensity for adolescents and young adults, with an equal sex incidence.

Etiology

A number of etiologic factors have been implicated. Aflatoxins of the mold *Aspergillus flavus* contaminate the diet in African and Asian communities with a high incidence of hepatocellular carcinoma. Low protein intake and consequent kwashiorkor also may be factors. Just as almost every type of experimentally induced cirrhosis may be followed by carcinoma of the liver, so a definite association between cirrhosis and primary carcinoma has been noted in human beings. Postnecrotic cirrhosis is the type most commonly preceding hepatocellular carcinoma; cirrhosis is present in 60 percent of patients. Serologic markers for hepatitis B virus or antibodies to hepatitis C virus are detected in most patients with hepatocellular carcinoma in endemic areas. Hepatic malignant tumors occur in 4.5 percent of cirrhotic patients, and the incidence is increased in patients with hemochromatosis. Parasite infestation with the liver fluke *Clonorchis sinensis* has been considered a factor in the development of cholangiocarcinoma, but this is open to question. There is no increased risk for hepatic carcinoma following infectious hepatitis that does not progress to cirrhosis. In the pediatric age group, the tumor is rarely related to cirrhosis. Another factor associated with hepatocellular carcinoma is aberrant α_1 -antitrypsin-piZ.

Pathology

Liver cell (*hepatocellular*) carcinoma is the most common type; the tumor cells resemble the parenchymal cell. Bile duct carcinoma (*cholangiocarcinoma*) is apparently derived from bile duct epithelium. *Hepatoblastoma* represents an immature variant of hepatocellular carcinoma.

Grossly, each of these types may present as a single, large nodule, as extensive nodularity, or as a diffuse permeation throughout the organ. The anatomic distribution of fibrolamellar carcinoma is unusual in that 75 percent present as solitary, large left lobe tumors. They have a prominent central scar. The hepatocellular carcinomas have a trabecular structure, and vascularity is a prominent feature. These lesions frequently invade branches of the portal vein and occasionally the hepatic veins. The formation of giant cells is a feature of hepatocellular carcinoma and aids in distinguishing this lesion from secondary carcinoma of the liver. Fibrolamellar carcinoma is characterized by eosinophilic hepatocytes and abundant fibrous stroma arranged in parallel bands around tumor cells. The cell type of bile duct carcinoma is columnar, and its microscopic appearance may be impossible to distinguish from that of carcinoma of the gallbladder or extrahepatic biliary duct system. Bile is never seen in the acini or the cells, whereas mucus formation is common.

Hepatic tumors extend by four methods:

- (1) Centrifugal growth, which indicates nodular expansion leading to compression of the surrounding hepatic tissue.
- (2) Parasinusoidal extension, which refers to tumor invasion into the surrounding parenchyma, either through the parasinusoidal spaces or through the sinusoids themselves.
- (3) Venous spread or extension of tumor from small branches of the portal system in a retrograde fashion into larger branches and eventually into the portal vein. Invasion of hepatic vein tributaries is less common but may extend up to the inferior vena cava or right atrium.
- (4) Distant metastases, the result of invasion of lymph channels and vascular systems. The most frequently involved locations are regional lymph nodes and lungs.

Metastases occur in 48 to 73 percent of patients.

Clinical Manifestations

Weight loss and weakness occur in 80 percent of patients, whereas abdominal pain is present in half (75 percent of patients with fibrolamellar carcinoma). The pain is usually dull and persistent, but dramatic sudden onset may occur in patients with intraperitoneal hemorrhage secondary to rupture of a necrotic nodule or erosion of a blood vessel. Manifestations of portal hypertension, such as bleeding varices, are infrequent but foreboding symptoms.

The liver is almost always enlarged but not tender. Splenic enlargement is present in one-third of patients, as are other signs of portal hypertension. The incidence of jaundice varies from 20 to 58 percent. Ascites develops in one-half to three-quarters of patients. A rapid increase in the symptoms and signs associated with cirrhosis or hemachromatosis is highly suggestive of superimposed hepatic carcinoma. In these patients, the amelioration of diabetes and occasional hypoglycemic intervals also indicate neoplastic change.

In over half of pediatric patients, the first evidence is an abdominal mass. Hemihypertrophy and sexual precocity occur in an occasional pediatric patient with hepatoblastoma.

Diagnostic Studies

The most consistently altered liver function test is the alkaline phosphatase determination. The serum bilirubin level is usually normal; 5[pr]- nucleotidase is usually elevated.

The demonstration of *alpha-fetoprotein (AFP)* in the serum by immunodiffusion, immunoelectrophoresis, and immunoassay techniques is useful in the differential diagnosis and epidemiologic studies. This protein is normally present in the fetus but disappears a few weeks after birth. Positive AFP tests are noted in about 75 percent of Africans but in only 30 percent of patients in the United States and Europe. False-positive results occur with embryonic tumors of the ovary and testis. Resection of the tumor converts the test to "negative"; recurrence may be detected by the reappearance of AFP in the serum.

Selective hepatic arteriography has been used to demonstrate an arterial pattern within the tumor, characterized by pooling and increased vascularity (Fig. 28-20). Scintillation scanning also may identify the space- occupying lesion within the liver, but *CT* and *MRI* are more sensitive. Percutaneous needle biopsy can provide a definitive diagnosis. *US* is particularly helpful in differentiating cystic from solid tumors. Intraoperative US has added an important surgical refinement and permitted more limited resections.

Treatment

The only curative therapy for hepatocellular carcinoma is surgical excision. In most instances, this entails a lobectomy, but with an appreciation of segmental anatomy, lesser "anatomic" resections are being performed more frequently. A major resection is compromised by cirrhosis because of increased vascularity, increased morbidity, and an inability for the cirrhotic liver to regenerate. The use of intraoperative US has permitted limited resections particularly in cirrhotic patients. Limited resections, including segmentectomies and bisegmentectomies, have achieved success in cirrhotic patients. Resection has resulted in survival rates of over 40 percent in patients with small, unifocal, well-differentiated tumors without vascular invasion. Transarterial chemoembolization using gelatin sponges, iodized oil, and chemotherapeutic agents has achieved 1-, 2-, and 5-year survival rates of 51, 24, and 6 percent, respectively, while decreasing the tumor size in almost a third of patients. Unresectable hepatocellular carcinomas may be converted to resectable lesions with combined radiation chemotherapy. Percutaneous Ethanol injection has resulted in the disappearance of tumors less than 4.5 cm in diameter. Cryosurgery, using an intraoperatively placed probe that delivers liquid nitrogen, also has effected tumor necrosis and destruction. Transplantation and resection yield equivalent 5-year survival rates of approximately 50 percent, but the recurrent rate is lower in patients who were transplanted. Transplantation achieved the best results in patients with small uninodular or binodular lesions. Resection of extrahepatic recurrences of hepatocellular carcinomas has resulted in several 5-year survivals.

In the case of hepatoblastoma, particularly in children, lesions often deemed not resectable can be converted into resectable tumors that have potential for cure. A combination of preoperative chemotherapy followed by resection resulted in a cure rate of 90 percent in children.

Prognosis

The outlook for untreated primary carcinoma of the liver is extremely poor, and the duration of the disease is rarely longer than 4 months from the time of onset of symptoms. Death is the result of cachexia, hepatic failure, sequelae or portal vein thrombosis, intraperitoneal hemorrhage, and metastases. In the Mayo Clinic experience, the 5-year survival after curative resection of a primary hepatic malignancy was 27 percent. Cholangiocarcinoma, nodal metastases, cirrhosis, hypocalcemia, and increased alkaline phosphatase level and prothrombin time were associated with decreased survival. In Pittsburgh, the 5-year survival was 32 percent, with a more favorable prognosis noted for patients with fibrolamellar carcinoma. Similar survival rates were achieved with partial hepatic resection and by hepatectomy and orthotopic transplantation. In children under age 2 with hepatoblastoma, 21 of 27 who survived operation were alive and well with no evidence of disease for a mean of 53 months. By contrast, in children with hepatocellular carcinoma, the prognosis is poorer. Most tumors are not resectable, and 5-year cures are anecdotal.

Other Primary Neoplasms

The three major lesions are sarcoma, mesenchymoma, and infantile hemangioendothelioma. All hepatic mesenchymal lesions are considered malignant. Angiosarcoma is the most common primary sarcoma of the liver. Exposure to vinyl chloride and Thorotrast injection have been implicated as etiologic factors.

Angiosarcoma is characterized by a short illness, jaundice, and coma progressing rapidly to death. Infantile hemangioendotheliomas occur in children under the age of 5 and are associated with skin lesions and cardiac failure secondary to arteriovenous shunts within the tumor. Although most of these pediatric lesions are fatal, spontaneous regression has been recorded, as has success with partial hepatectomy.

Metastatic Neoplasms

Metastatic neoplasms represent the most common malignant tumors of the liver. The relative proportion of primary to secondary neoplasms is estimated to be 120, and there is no statistical difference between those with and those without cirrhosis. The liver is second only to regional lymph nodes as a site of metastases for tumors, and 25 to 50 percent of all patients dying of cancer have been found to have hepatic metastases. Fifty percent of patients with gastrointestinal tumors have hepatic metastases when autopsied.

Metastatic neoplasms reach the liver by four routes: (1) portal venous circulation, (2) lymphatic spread, (3) hepatic arterial system, and (4) direct extension.

Metastases appear in the liver at varying times in relation to primary lesions: (1) Precocious metastasis is evident when the primary lesion is not suspected (carcinoid of the ileum). (2) Synchronous metastases occur when the hepatic neoplasm is detected at the same time as the primary lesion. (3) Metachronous metastasis is one in which appearance is delayed following the successful removal of a primary tumor (ocular melanoma). The growth pattern of the metastatic tumor is frequently more rapid than the original lesion, and the mitotic count of metastatic hepatic neoplasms has been shown to be five times greater than that of the extrahepatic primary lesion.

Clinical Manifestations

Symptoms referable to the liver are present in 67 percent of patients with proven metastases. These include hepatic pain, ascites, jaundice, anorexia, and weight loss. On examination, hepatic nodularity is apparent in half the patients, and a friction rub is audible in 10 percent. Jaundice, ascites, and the signs of portal hypertension are present in approximately one-quarter to one-third of patients. With carcinoid tumors, hepatic metastases are of major importance in the pathogenesis of the flushing syndrome.

Diagnostic Studies

The alkaline phosphatase level is increased in over 80 percent of patients. The *SGOT* level is elevated in approximately two-thirds of patients, but the serum *AFP* determination is negative. The carcinoembryonic antigen (CEA) level may provide a marker for metastatic colon carcinoma. Enhanced *CT*, particularly that performed with selective arterial infusion via the hepatic artery, is the most sensitive means of defining intrahepatic metastases. *MRI*, scintigraphy, and arteriography also have been used. In the case of patients who have undergone resection of colorectal carcinomas, surveillance generally relies on sequential determinations of the CEA level. If elevation occurs, CT is performed. Intraoperative *US* has aided in the definition of resectable lesions.

Treatment

Surgical treatment of hepatic metastases should be considered only if (1) control of the primary tumor is accomplished or anticipated, (2) there are no systemic or intraabdominal metastases, (3) the patient's condition will tolerate the major operative procedure, and (4) the extent of hepatic involvement is such that resection and total extirpation of the metastasis is feasible. Resection of segments of the liver containing metastases has effected reasonable long-term survival without recurrence for patients with primaries in the colon and rectum and for those with Wilms' tumor. Rare survivors have been reported for other primaries.

A metastasis noted during a colon resection should be removed at that time if it is readily removable without anticipated blood loss. However, if it appears that a major hepatic resection will be required, the operation is delayed for months and is preceded by CT scan and angiography to assess resectability. Although 20 percent of patients with colorectal cancer have hepatic metastasis, only one-quarter of these are potentially resectable; half of these, in turn, have other metastases that would negate the value of resection.

A multi-institutional study of hepatic resection for colorectal metastases reported a 5-year actuarial survival of 33 percent and a 5-year actuarial disease-free survival of 21 percent. Single metastases and unilobar disease were associated with improved survival. Hepatic resection of greater than four metastatic lesions rarely was associated with prolonged survival. Forty percent of patients with recurrent disease after hepatic resection for metastatic colorectal tumors initially have only liver metastases. Reresection is followed by long-term survival in about one-third of carefully selected patients. Bismuth and associates have treated patients with colorectal hepatic metastases that were initially considered to be unresectable with doxorubicin, and this resulted in regression to a size that permitted resection. Cure rates equivalent to series of more favorable patients were achieved.

An implantable refillable pump has been used to intraarterially infuse floxuridine and other chemotherapeutic agents, but most series could demonstrate no improvement in survival or reduction in toxicity compared with intravenous administration of the drugs.

Patients with Wilms' tumor metastatic to the liver frequently can be cured by hepatic resection and adjuvant chemotherapy and radiation therapy. Resection of metastases from ocular melanoma, breast, stomach, uterus, cervix, ovary, and renal carcinomas rarely has achieved a cure (Table [28-2](#)).

Palliative cytoreductive surgical measures are indicated for marked pain associated with hepatic neoplasm and for the excision of metastases in patients with the flushing syndrome of carcinoid tumor. Resection of the major portion of the hepatic metastases (debulking procedure), even if residual tumor remains, has resulted in significant symptomatic improvement and reduction of the 5-hydroxyindoleacetic acid levels to normal. These subjective and objective changes have persisted for several years. Dearterialization and radiographically controlled embolization have achieved similar results.

HEPATIC RESECTION

The indications for hepatic resection include (1) trauma with resulting devascularization of hepatic tissue, (2) cysts, (3) granulomas, (4) primary neoplasms of the liver, and (5) secondary malignant tumors that involve the liver either by direct extension or as metastatic lesions.

Removal of up to 80 percent of the liver is compatible with life. Following excision of this amount, patients maintain normal blood ammonia levels and normal prothrombin times. Fibrinogen production is insignificantly impaired; clinical jaundice is a transient phenomenon. Following major resections, by the fifth postoperative day 95 percent of patients show clinical improvement in function, with the bilirubin and alkaline phosphatase levels returning to normal by the end of the third week. The most profound changes are noted in the serum albumin, which, by the third week, is usually restored to normal.

Regeneration results from marked hypertrophy of the remaining tissue. The remaining portion of the liver responds as rapidly and completely after second and third partial hepatectomies as after an initial insult. There is experimental evidence for hepatotropic substances in portal venous blood. Insulin may represent the major anabolic factor. Very little restoration occurs after partial hepatectomy of the cirrhotic liver.

Management of the Patient

Preoperative therapy is directed at maintaining optimal liver function and correcting any defects that may be present. A diet high in calories, proteins, and carbohydrates is utilized, and the administration of albumin may be required to achieve normal levels. Vitamin K is given routinely until a normal prothrombin time results. In the presence of jaundice, other fat- soluble vitamins are added. Fresh-frozen plasma will rapidly replenish coagulation factors. Since many patients have a reduced hematocrit, transfusion with fresh whole blood rich in platelets and coagulation factors is indicated. Major hepatic resection is attended by a prohibitive mortality rate in the patient with sulfobromophthalein retention greater than 35 percent, a serum albumin level lower than 2.0 g, and an increased prothrombin time that does not respond to parenteral vitamin K.

Postoperatively, infusion of 10% glucose is continued until the patient maintains an adequate oral intake to obviate severe hypoglycemia, which has been reported. Following more intensive resections, daily administration of 25 to 50 g albumin is usually required for 7 to 10 days to maintain the serum level above 3 g/dL. Antibiotics are administered prophylactically. Analgesics and hypnotics that are detoxified by the liver are used only sparingly. Intraabdominal abscess formation and sepsis are the most common complications of major hepatic resection, occurring in 20 to 30 percent of cases. Subphrenic abscesses usually can be managed with percutaneous drainage.

Operative Procedures

Control of Bleeding

This may be accomplished by (1) ligation or compression of blood vessels within the substance of the remaining liver segment, (2) efforts directed at the raw surface, including cautery and argon beam coagulation, and (3) control of the main blood vessels entering the porta hepatis.

Omental grafts, peritoneal grafts, Gelfoam, micronized collagen, and rapidly polymerizing adhesives have been applied to the raw surface as local hemostatic agents. Compression of the main vessels entering the liver facilitates the demonstration of bleeding sites along the raw surface. The hepatic artery and portal vein may be compressed for over an hour without affecting hepatic structure or function.

In order to obviate uncontrollable bleeding, total vascular occlusion can be helpful for lesions located close to the major hepatic veins and the inferior vena cava. After a vascular clamp is applied to the hepatoduodenal ligament, the vena cava above and below the liver is occluded. If the patient is loaded with fluids, 1 hour of total occlusion can be tolerated.

Techniques of Resection

On the basis of new concepts of segmental anatomy, the following classification of hepatic resection is applicable (Fig. [28-21](#)): (1) *Subsegmental*, or *wedge*, *resection* is removal of an area of the liver that is less than a segment and without an anatomic dissection plane. (2) *Left lateral segmentectomy* ("left lobectomy" in the old nomenclature) is excision of the liver mass to the left of the left segmental fissure along an anatomic plane. (3) *Left medial segmentectomy* is resection between the main interlobar fissure and the left segmental fissure. (4) *Left lobectomy* ("left hepatectomy") is excision of all hepatic tissue to the left of the main lobar fissure. (5) *Right lobectomy* ("right hepatectomy") is removal of the liver to the right of the main lobar fissure. (6) *Extended right lobectomy* is excision of the entire lobe plus the medial segment of the left lobe (*trisegmentectomy*), i.e., excision of all tissue to the right of the umbilical fossa, fossa for the ligamentum venosum, and the ligamentum teres.

Based on portal distribution, Couinaud has defined eight hepatic segments (see Fig. [28-2](#)). Segmental resection of one or two contiguous segments can be performed by a transparenchymatous approach following the anatomic scissura to the vascular pedicle (Fig. [28-22](#)).

The liver is mobilized initially by dividing the appropriate ligamentous attachments, i.e., ligamentum teres, falciform, and triangular and coronary ligaments. Dissection of the porta hepatis identifies the branches of the hepatic artery, portal vein, and biliary duct system supplying the segment or lobe to be removed (Fig. [28-23](#)). These are individually temporarily occluded. By rotating the liver, the hepatic veins may be isolated at their junctions with the inferior vena cava and ligated (Fig. [28-24](#)). Glisson's capsule is then incised along the surgical plane (Fig. [28-25 A](#)), and the cleavage plane of the hepatic parenchyma itself is best established by means of a scalpel handle or finger to permit exposure of the larger ducts and vessels, which may be individually clamped and ligated as they are encountered. This incision is continued posteriorly until the major hepatic vein or veins are identified (Fig. [28-25 B](#)), double ligated, and transected. The specimen is removed. The previously occluded structures in the porta hepatis are unclamped, and if bleeding or biliary drainage from the raw surface persists, the appropriate structure is ligated. The remaining raw surface may be covered with omentum. The blood flow to and from the remaining segments of the liver must be preserved carefully.

The majority of lobar resections, even right lobe resections, can be carried out transabdominally, and it is not necessary to proceed along the outlined sequence of events for surgical excision. Finger fracture is employed for trauma but is also applied in many instances for tumor. The sequence is summarized in Fig. [28-26](#). One can reduce the blood flow into the liver by temporarily cross-clamping the hepatoduodenal ligament; this procedure can be carried out for 60 min using a vascular clamp intermittently. Glisson's capsule is then incised anteroinferiorly, and the incision is carried down to the region of the porta hepatis. The vessels in the porta hepatis are then picked up in the parenchyma as they enter the liver. The parenchymal dissection is continued along anatomic planes, picking up vessels and ducts as they traverse the liver, until the hepatic venous structures are also picked up in the hepatic parenchyma. Intraoperative *US* is helpful in locating the intraparenchymal major vessels. The operative time is significantly reduced by this technique, and blood loss is only moderately increased. An ultrasonic disruption of the parenchyma can be used to isolate vessels as they traverse the resection plane. Trisegmentectomy necessitates anatomic dissection in order to avoid interrupting veins from the remaining segment.

PORTAL HYPERTENSION

Hypertension with the portal vein and its tributaries may accompany hepatic disease or disturbance in the anatomy of the extrahepatic vascular system. As a consequence of this elevated pressure, congestion of collateral pathways is established and may be manifested by esophagogastric varices, ascites, hypersplenism, or encephalopathy.

Etiology

The etiologic factors implicated in portal hypertension are listed in Table 28-3. Increased hepatopetal flow is an infrequent cause of portal hypertension. Hepatic arterial-portal venous fistula has been reported rarely, and the diagnosis can be established by CT, MRI, or angiography. Successful treatment has been effected by ligation of the hepatic artery or by direct closure of the fistula through an arteriotomy. Splenic arteriovenous fistula is also a relatively uncommon lesion. This has a predilection for women between the ages of 20 and 50 and may become asymptomatic during pregnancy. Calcification in the left upper quadrant is suggestive, and aortography may be diagnostic. Resection of the fistula or splenic artery and splenectomy are therapeutic.

An increase in forward blood flow in the portal venous system also has been proposed as the cause of portal hypertension in patients with tropical splenomegaly and myeloid metaplasia. In patients with increased hepatopetal flow related to splenomegaly, splenectomy alone may be therapeutic.

Since the hepatic veins constitute the sole efferent vascular drainage of the liver, obstruction or increased pressure within these vessels or their radicals results in an increased sinusoidal and portal pressure. This outflow obstruction (Budd-Chiari) syndrome is associated most frequently with an endophlebitis of the hepatic veins, which may be isolated or part of a generalized thrombophlebitis process. A web in the suprahepatic vena cava has been reported to cause the syndrome in Japanese people.

The clinical picture depends on the rapidity and degree of venous obstruction. With sudden and complete obstruction, the presentation is that of an abdominal catastrophe with severe abdominal pain, nausea, vomiting, and rapid enlargement of the abdomen by ascites. This rarely occurs. More commonly, obstruction to the hepatic venous system appears to be gradual and is associated with mild to moderate abdominal discomfort and ascites. Patients should be evaluated with an inferior vena cavagram to demonstrate occluded hepatic veins. If liver biopsy defines advanced cirrhosis, transplantation should be considered. If cirrhosis is absent, a mesenteric systemic venous shunt is the treatment of choice. A peritoneovenous shunt has provided temporary relief. Definitive treatment demands a side-to-side portal systemic anastomosis, since the portal vein must act as an efferent hepatic conduit. If the inferior vena cava is obstructed, a mesenteric-atrial shunt with a conduit may be required. In some patients, the manifestations have disappeared spontaneously.

Portal hypertension secondary to impaired flow in the extrahepatic portal venous system is unique in that the hypertension usually is not complicated by hepatocellular dysfunction. Congenital atresia or hypoplasia, as an extension of the obliterative process of the umbilical vein and ductus venosus, is rare. More commonly, there is a cavernomatous transformation of the portal vein that probably represents organization and recanalization of thrombi within the vessel. The most common etiologic factor in the development of extrahepatic portal venous obstruction in childhood may be some form of infection. Bacteria may be transmitted via a patent umbilical vein, but a history of neonatal omphalitis is rarely obtainable.

Extrahepatic obstruction also may be secondary to trauma and extrinsic compression caused by adhesions, inflammatory processes, and tumors. Isolated splenic vein thrombosis, usually a consequence of alcoholic pancreatitis, may cause esophagogastric varices. In this case, splenectomy cures the portal hypertension.

The overwhelming majority of cases of portal hypertension are related to an intrahepatic obstruction. This accounts for over 90 percent of patients with portal hypertension in most large series. A number of hepatic diseases have been implicated, but no single explanation of pathogenesis has proved totally satisfactory. The pathogenetic factors include (1) hepatic fibrosis with compression of portal venules, (2) compression by regenerative nodules, (3) increased arterial blood flow, (4) fatty infiltration and acute inflammation, and (5) intrahepatic vascular obstruction. The hepatic diseases associated with portal hypertension include nutritional cirrhosis, postnecrotic cirrhosis, schistosomiasis, biliary cirrhosis, hemochromatosis, Wilson's disease (hepatolenticular degeneration), congenital hepatic fibrosis, and infiltrative lesions.

Nutritional cirrhosis is the most common and has a worldwide distribution. In Western countries it is frequently associated with chronic alcoholism. As is true with all intrahepatic lesions, with the exception of congenital hepatic fibrosis and schistosomiasis, the major resistance to the flow of portal blood is located on the hepatic venous side of the sinusoid (postsinusoidal). Postnecrotic cirrhosis accounts for 5 to 12 percent of cases and represents progression of an acute viral hepatitis or toxic hepatic injury. Frequently, a history of viral infection is not obtainable. In the Orient, in most cases cirrhosis is a consequence of hepatitis. Both hepatitis B and C have been indicted. Biliary cirrhosis may be due to extrahepatic obstructions and secondary cirrhosis or to a primary hepatic lesion. Advanced portal cirrhosis is an almost invariable feature of hemochromatosis. Wilson's disease is characterized by alteration of hepatic function and structure and by mental deterioration. Congenital hepatic fibrosis may be related to dilatation of the intrahepatic bile ducts. It is usually an autosomal recessive disease. Clinical features include gross enlargement and firm consistency of the liver, accompanied by manifestations of portal hypertension and cholangitis. Hepatic function is not disturbed. As hepatic infestation with *Schistosoma mansoni* results in presinusoidal obstruction, there is no impairment of hepatic function until late in the course of the disease.

Pathophysiology

Portal hypertension refers to an elevated pressure within the portal venous system. This pressure reflects a dynamic, constantly fluctuating force. In addition to diurnal fluctuations, the pressure varies with changes of position, phases of respiration, and intraabdominal pressure. The normal portal pressure is less than 250 mmH₂O, with a mean value of 215 mmH₂O.

Portal pressure can be assessed by a variety of techniques. During an operative procedure, cannulation of an omental vein or the portal vein itself provides a direct recording. The pressure also can be determined by occlusive catheterization of the hepatic venule. This procedure is analogous to determination of pulmonary capillary pressure, in that it is based on the assumption that the occluding catheter creates a static column of blood extending from the hepatic vein to the junction of the hepatic arterial and portal venous streams as they converge in the sinusoidal bed. The procedure is carried out by cardiac catheterization technique and is particularly valuable in the diagnosis of extrahepatic portal obstruction. In this situation, the presinusoidal obstruction is associated with a normal occlusive hepatic venule pressure and an elevated splenic pulp pressure.

In all instances of portal hypertension, splenic pulp pressure is elevated. Intrasplenic pressure is essentially uniform throughout the pulp and is unrelated to the size of the organ. The pressure is usually 2 to 6 mmHg higher than the pressure within the portal vein per se, a function of the direction of venous flow. Splenic pulp manometry is carried out under local anesthesia but is contraindicated in patients with a bleeding tendency, thrombocytopenia, or severe jaundice.

Splenoportography or the venous phase of celiac and superior mesenteric angiography defines the pathologic features of the portal circulation. The studies provide a demonstration of collateral veins, particularly esophagogastric varices. They also provide graphic demonstration of the site of obstruction, i.e., intrahepatic or extrahepatic. Under normal circumstances, no collaterals are visualized and a good arborization is noted in the liver (Fig. 28-27). Whenever collaterals are apparent, the diagnosis of portal hypertension is suggested. Usually one can define the coronary vein contributing to esophageal varices by this technique (Fig. 28-28).

The umbilical vein also has been used to outline the portal system. In 80 percent of these cases, the obliterated vein can be isolated and dilated to permit passage of a catheter and injection of radiopaque material into the left portal vein. The portal and hepatic veins also can be visualized by percutaneous transhepatic cannulation.

Pathologic Anatomy

The collateral vessels (Fig. 28-29) that become functional in cases of portal hypertension are classified in two groups:

(1) Hepatopetal circulation occurs only when the intrahepatic vasculature is normal and obstruction is limited to the portal vein. In this situation, the accessory veins of Sappey, the deep cystic veins, the epiploic veins, the hepatocolic and hepatorenal veins, the diaphragmatic veins, and the veins of the suspensory ligaments carry a limited amount of portal venous blood to the liver.

(2) Hepatofugal flow is the type most commonly provided by the collateral circulation.

The vessels of the hepatofugal circulation include

(1) The coronary vein, which courses to the esophageal veins and then to the azygos and hemiazygos veins with eventual termination in the superior vena cava.

(2) The superior hemorrhoidal veins, which communicate by way of the hemorrhoidal plexus with hemorrhoidal branches of the middle and inferior hemorrhoidal veins and ultimately drain into the inferior vena cava.

(3) The umbilical and paraumbilical veins, which communicate with superficial veins of the abdominal wall and anastomose freely with the superior and inferior epigastric veins. Dilatation occurs in 22 percent of patients with portal cirrhosis, and the advanced stage is known as the *caput medusae*. The cephalad portion of the obliterated umbilical vein may remain patent in adult life or become recanalized, contributing to the Cruveilhier-Baumgarten syndrome.

(4) Retroperitoneally, the veins of Retzius, which form an anastomosis between the mesenteric and peritoneal veins and empty directly into the inferior vena cava.

In general, the collateral circulation does not effectively decompress the portal system, and the amount of blood shunted is relatively insignificant. Assuming the cross-sectional diameter of the normal portal vein to be 2 cm, then, according to Poiseuille's law, over 4000 collateral veins 0.5 cm in diameter will be needed to provide equivalent flow. The highest values of portal pressure are recorded in the group in which collateralization is more marked. In rare instances, spontaneous portal systemic shunts have effectively decompressed the portal system.

Esophagogastric Varices

As the veins become engorged, vessels in the submucosal plexus of the esophagus increase in size and become dilated. In the later stages, the overlying submucosa may disappear, and the walls of the vein actually may form a lining of the esophagus. The submucosal veins in the fundus and subfundal regions of the stomach also become varicose. Gastric varices occur predominantly in the cardiac end of the stomach but also have been found along the lesser curvature. Varices also have been demonstrated in the duodenum and ileum.

Although the presence of esophagogastric varices is, in itself, of minor consequence, rupture and bleeding from these vessels constitute the most alarming and serious complication of portal hypertension. The varices are almost always associated with portal hypertension but infrequently have occurred in patients with normal pressure. Over 90 percent of adult patients demonstrate intrahepatic disease, whereas in childhood the varices are usually related to extrahepatic portal obstruction. Precipitation of the bleeding episode has been ascribed to two factors: increased pressure within the varix and ulceration secondary to esophagitis. Regurgitation of gastric juice into the esophagus has been implicated, and ulcers of the esophagus have been demonstrated in 25 percent of nonintubated and 50 percent of intubated patients. The frequency and severity of bleeding also

are related to the degree of hepatocellular dysfunction.

Natural Course

Bleeding is to be anticipated in approximately 30 percent of cirrhotic patients with demonstrable varices. The elapsed time from the diagnosis of varices to the first hemorrhage varies between 1 and 187 weeks. Almost all hemorrhages occur within 2 years of the initial observation. It is difficult to predict bleeding in a patient with known varices. A prospective multicenter study using univariate analysis of endoscopic variables was able to identify patients at high risk. Etiology is a prime consideration. Varices secondary to extrahepatic portal obstruction must be considered separately, since it is rare for these patients to die of hemorrhage. By contrast, the mortality risk of repeated hemorrhage in patients with esophageal varices secondary to cirrhosis is extremely high. Approximately 70 percent of these patients die within 1 year of the first hemorrhage. Sixty percent of cirrhotic patients who have hemorrhaged once rebleed massively within 1 year.

Prophylactic sclerotherapy, propranolol, and a combination of the two modalities were of no benefit in preventing bleeding from varices in patients with alcoholic cirrhosis. A prophylactic shunt is not advised for a patient with varices that have not bled, since one cannot predict which patients will bleed; the survival is not improved, and encephalopathy may be induced. A cooperative Japanese study group has presented data suggesting that a prophylactic devascularization procedure or selective shunt can prevent bleeding without leading to a significant increase in mortality or morbidity in Child's A and B patients. There is also evidence of improved survival with prophylactic nondecompressive devascularizing procedures.

Acute Bleeding

In children, massive hematemesis almost always emanates from bleeding varices. Acute hemorrhage is usually the first manifestation of portal hypertension in children. Seventy percent of patients experience their first bleeding episode before age 7, and almost 90 percent hemorrhage before age 10.

In the adult, bleeding varices comprise one-quarter to one-third of cases of massive upper gastrointestinal tract bleeding. In cirrhotic patients, varices are the source of bleeding in approximately 50 percent, whereas gastritis is implicated in 30 percent and duodenal ulcers in 9 percent. It is now believed that peptic ulcer does not occur more frequently in cirrhotic patients. Correlation of the lesion with the severity of bleeding reveals that in the majority of cases bleeding from varices is severe hemorrhage, whereas bleeding from gastritis involves only mild to moderate blood loss.

Since the management of bleeding varices differs significantly from that of bleeding due to other causes, it is important to establish a diagnosis on an emergency basis. Physical examination may reveal the stigmata of cirrhosis. Splenomegaly is particularly suggestive of portal hypertension. Tests of hepatic function have been used but do not have uniform reliability. Barium swallow has a significantly high percentage of false-negative results. In a series of patients with proved varices, radiographic demonstration was present in only half. Celiac or superior mesenteric arteriography will rule out an arterial bleeding site, and the venous phase of the arteriogram will demonstrate collateral venous circulation. Bleeding from a varix is not visualized. Esophageal balloon tamponade has been used as a diagnostic measure, but varices are controlled in only two-thirds of the patients, and moreover, peptic ulcer may stop bleeding after the gastric balloon is inflated. Esophagoscopy represents the single most reliable technique, since it alone defines the bleeding point. On the other hand, esophagoscopy may fail to reveal varices because of variations in transvariceal blood flow. In addition, there is a significant observer variation in the endoscopic evaluation of varices.

Treatment

The therapeutic regimen is directed at promptly controlling bleeding without further disturbing an already impaired hepatic function. Rapid control is critical in order to avoid the injurious effects of shock on hepatic function as well as the toxic effects of absorption of blood from the gastrointestinal tract. The therapeutic approaches may be divided into nonoperative and operative methods that directly approach the bleeding site and techniques that act indirectly by decreasing portal pressure (Table 28-4).

Balloon tamponade has reduced the mortality and morbidity from bleeding varices in good-risk patients, particularly those in whom the varices were secondary to extrahepatic portal hypertension or compensated cirrhosis. Little change has been noted in the mortality rate for poor-risk patients, and reports have indicated failure to control hemorrhage in 25 to 55 percent of patients. Increasing awareness of the complications associated with this technique, including aspiration, asphyxiation, and ulceration at the site of the tamponade, has reduced its use. Either a four-lumen tube should be used or a small nasogastric tube should be positioned proximal to the esophageal balloon to provide suction and prevent aspiration.

Endoscopic injection of a sclerosing solution into varices also has controlled bleeding successfully. In a large series, bleeding was controlled in 93 percent of patients. There are significant risk factors associated with sclerotherapy. Endoscopic ligation of esophageal varices has resulted in control of bleeding equivalent to that of sclerotherapy, with fewer complications.

Drug therapy to reduce portal hypertension has employed surgical vasopressin, which acts by constricting the splanchnic arterial circulation and consequently reducing portal pressure and flow by approximately 40 percent. The drug is contraindicated in patients with angina, since generalized vasoconstriction results. Effective control has accompanied direct infusion of vasopressin, 0.2 unit/mL per minute. Isoproterenol may be given simultaneously to reduce the hemodynamic hazards of vasopressin related to its potential effect on the cardiac output. Somatostatin has been proved as effective as vasopressin in controlling acute bleeding and has less significant side effects. Paracentesis in a patient with bleeding varices and tense ascites will reduce portal pressure immediately.

Transjugular intrahepatic portal-systemic shunts (TIPS) using a self- expanding stent have reduced the portal-systemic gradient to about 10 mmHg and successfully stopped bleeding from esophagogastric varices. The procedure can be performed successfully in over 90 percent of patients. TIPS is followed by variceal rebleeding in 10 to 20 percent of patients. The incidence of shunt dysfunction ranges between 15 to 60 percent over a 6- to 12-month period. In the vast majority of cases, patency can be reestablished. TIPS is particularly applicable to control bleeding in patients with minimal hepatic reserve, in whom a transplant is indicated. The procedure obviates a portacaval shunt that compromises the transplant operation.

Surgical therapy includes transesophageal ligation and emergency portal- systemic shunt. The results of transesophageal ligation have improved significantly with stapling techniques, and results better than those with sclerotherapy have been reported. Devascularization procedures such as the Sugiura procedure have had variable success. Success rates from Western nations have not matched the Japanese success.

A more liberal use of emergency portacaval shunts to stop bleeding has been advised for the cirrhotic patients whose bleeding cannot be controlled by nonoperative measures. Emergency portal-systemic shunt remains an important option for selected patients with acute variceal bleeding. Recent experience of emergency shunt performed on 400 unselected patients within 8 h of initial contact indicated survival rates of 85, 78, and 71 percent at 30 days, 5 years, and 10 years, respectively. In most series, the mortality rate for patients with Child's class C disease remains high. The base figure that serves as a frame of reference for comparison is the mortality for patients with bleeding varices not subjected to emergency portacaval shunts, and this ranges between 66 and 73 percent. There is little question that an effective portal-systemic decompressive procedure almost always stops bleeding. No significance could be attributed to the presence or absence of jaundice, but ascites, when present, was associated with a marked reduction of survival rate.

In the pediatric age group and in adults with extrahepatic portal venous obstruction and normal liver function, despite the fact that the bleeding is often alarming, spontaneous cessation almost always occurs, and esophageal tamponade or vasopressin is rarely necessary. Hospitalization, bed rest, blood replacement, and sedation almost always suffice for patients with bleeding secondary to extrahepatic portal obstruction.

The majority of patients with acute bleeding varices is not in shock at the time of admission to the hospital, although the hematocrit is often reduced, and blood replacement may be necessary. Frozen red blood cells and fresh-frozen plasma should be employed for transfusing cirrhotic patients. This provides the clotting factors that are frequently diminished in the presence of hepatic disease and avoids the increased ammonia content and diminished platelet and prothrombin supply characteristic of old blood. There is a linearly progressive daily increment of 35 mg/dL ammonia nitrogen in banked blood, which can be responsible for exogenous hepatic coma. Therapy directed at preventing hyperammonemia and hepatic coma consists primarily of removing blood from the gastrointestinal tract. Catharsis, gastric lavage, and enemas are employed. If vasopressin has been administered, it will induce intestinal motility and effect a catharsis. A reduction in intestinal bacterial flora also contributes to the prevention of coma, and nonabsorbable antibiotic therapy is used to accomplish this.

Prevention of Recurrent Hemorrhage

Sclerotherapy has been used to eradicate varices subsequent to a major bleeding episode. This has been achieved in 80 to 90 percent of patients with an average of five injection sessions over a 9-month period. Endoscopic banding provides equivalent results. Somatostatin has been administered after bleeding was controlled initially by sclerotherapy to obviate future bleeding episodes.

The case for surgical intervention is based on the precept that a patient who has bled from esophageal varices is likely to rebleed and that subsequent bleeding episodes are associated with a higher mortality than an elective operative procedure.

Opinions differ about the role of decompressive procedures in children and in adults with portal hypertension due to extrahepatic portal venous thrombosis. Some children can be treated satisfactorily and safely without operation despite repeated episodes of variceal bleeding. The results of operation in terms of survival are significantly more encouraging in this population than in adults. Therefore, many series have suggested an aggressive approach in children with recurrent bleeding episodes. Central or distal splenorenal shunt or an anastomosis between the inferior vena cava and the superior mesenteric vein is applicable to this group of patients. The incidence of postoperative encephalopathy has been negligible. Devascularization operations such as the Sugiura procedure also have been used for these patients.

Presinusoidal obstruction (hepatic fibrosis, extrahepatic portal venous thrombosis, schistosomiasis) is characterized by portal hypertension and may be associated with normal hepatic function. In patients with hepatic fibrosis and extrahepatic portal venous obstruction, the results are gratifying; the surgical procedure generally will prevent subsequent bleeding and provide the patient with an essentially normal life expectancy. The patients with schistosomiasis are a unique group in that they are extremely liable to postshunt encephalopathy. A selective splenorenal shunt or devascularization procedure is the preferred operation.

Postsinusoidal portal hypertension is invariably complicated by impaired hepatic function. The role of decompressive procedures is least well defined for this group of cirrhotic patients. An elective procedure should be considered when the presence of an active intrahepatic process such as hyaline necrosis or acute fatty infiltration has been ruled out. Ascites that fails to respond to medical therapy, a prothrombin time that remains prolonged following parenteral administration of vitamin K, a serum bilirubin level above 3 mg/dL, a sulfobromophthalein retention greater than 20 percent, and a serum albumin level less than 2.5 g/dL are all associated with a poor postoperative prognosis. In these patients there

is immediate deterioration following portacaval shunting, but this is actually no greater than after other operations of comparable severity. Child's criteria and other assessments of hepatic function are not completely predictive and relate only to the immediate postoperative course.

In patients with end-stage liver disease and marked hepatocellular dysfunction, after the bleeding had been controlled with sclerotherapy or *TIPS*, orthotopic liver transplantation often is appropriate.

Ascites

Etiology

The mechanisms contributing to the formation of ascites are complex and incompletely understood. Portal hypertension is regarded as a contributory but minor factor, since there is no correlation between the degree of portal hypertension and the extent of ascites. Ascites is not a usual accompaniment of extrahepatic portal venous obstruction but has been noted occasionally. Impairment of hepatic venous outflow with subsequent congestion of the liver is accompanied by an increase in the size of lymphatic vessels and increased production of the hepatic lymph that extravasates through the capsule of the liver into the peritoneal cavity. In clinical cirrhosis, there is an increase in the size of hepatic channels and an augmented flow of thoracic duct lymph. Two distinct patterns of intrahepatic vasculature have been correlated with the presence or absence of ascites. With irreversible ascites, there is an absolute decrease in the hepatic venous bed and a concomitant increase in both the portal venous and hepatic arterial beds. By contrast, when cirrhosis is unaccompanied by ascites, there is a deficit in all vascular systems.

Reduced serum osmotic pressure related to hypoalbuminemia does exert some influence. The response of patients to albumin infusion is variable, and the reduced osmotic pressure may represent the result rather than the cause of fluid accumulation. The most profound biochemical change that accompanies the formation of ascites is the retention of sodium and water. There is evidence that adrenocortical hormone is a factor in the renal retention of sodium, and higher concentrations of antidiuretic substances have been noted in the urine of patients with cirrhosis and ascites.

Treatment

Bed rest reduces the functional demand on the liver. A diet high in calories with an excess of carbohydrates and proteins, supplemented by vitamins, is directed toward improving hepatic function, while low sodium (10 to 20 meq daily) intake is essential. Fluid is usually not restricted, and potassium supplements are provided routinely to treat the potassium depletion that accompanies the formation of ascites.

Chlorothiazide is usually used to initiate diuretic therapy, and approximately two-thirds of patients will respond to this medication. Potassium supplements are required. The aldosterone antagonists are employed for patients with incipient hepatic coma. Abdominal paracentesis as an initial procedure has diagnostic value, but repeated procedures are contraindicated because they deplete the body of protein and contribute to the development of systemic hyponatremia. Furosemide (Lasix) is the most frequently used drug. In some refractory cases, ethacrynic acid will help.

Emphasis on the importance of obstruction of hepatic venous outflow led to the proposal of side-to-side portacaval shunts as a method of therapy. These procedures were based on the hypothesis of providing a second outflow tract, with the portal vein acting as a hepatofugal conduit. At present, the operation is limited to patients with Budd-Chiari syndrome and to those patients who cannot be managed on a strict low-sodium diet and diuretic therapy, an unusual circumstance. Peritoneal venous shunts of the Leveen and Denver types have effectively controlled medically intractable ascites. The peritoneovenous shunt provides greater relief of ascites in a shorter period of time than does intensive medical therapy; also, ascites recurs more rapidly in medically treated patients. Improvement may be related to increased creatinine clearance and normalization of renin activity and aldosterone levels. Adverse consequence of the procedures include disseminated intravascular coagulopathy and initiation of variceal bleeding. Persistent patency of any peritoneal venous shunt is difficult to achieve.

Umbilical herniorrhaphy in a cirrhotic patient with marked ascites presents a significant risk, with hazards of leakage of ascitic fluid, infection, necrosis of the abdominal wall, and variceal bleeding due to interruption of collateral veins.

Hypersplenism

Splenomegaly, with engorgement of the vascular spaces, frequently accompanies portal hypertension. There is little correlation between the size of the spleen and the degree of hypertension. When hematologic abnormalities occur, they have been related to sequestration and destruction of the circulating cells by immune mechanisms mediated by the enlarged spleen or secretion by the hyperactive spleen of a substance that inhibits bone marrow activity. Patients may demonstrate reduction of any or all of the cellular elements of blood. The usual criteria for hypersplenism are a white blood cell count below 4000/mm³ and a platelet count below 100,000/mm³. Schistosomal fibrosis frequently induces hypersplenism.

No correlation exists between degree of anemia or leukopenia and the 5- year survival rate in patients. Splenectomy is rarely indicated and does not permanently reduce portal pressure. Removal of the spleen negates the possibility of performing a selective shunt. Decompression of the portal venous system is rarely indicated for treatment of hypersplenism alone. Significant hypersplenism in a patient undergoing elective surgical treatment for bleeding varices favors a splenorenal anastomosis, but both portacaval anastomosis and selective splenorenal shunt have been accompanied by reduction of the spleen and correction of the hypersplenism in about two-thirds of patients.

Encephalopathy and Coma

The development of neuropsychiatric symptoms and signs is related to natural and surgically created portal-systemic shunts and is identified by the term *portal-systemic encephalopathy*. This rarely occurs in patients with obstruction of the extrahepatic portal venous system without hepatocellular dysfunction. The neuropsychiatric syndrome usually is associated with cirrhosis and occurs in patients with marked hepatic dysfunction. Postshunt encephalopathy rarely occurs in patients with extrahepatic portal obstruction unaccompanied by hepatic dysfunction. Operative procedures that decompress the portal system also have been associated with varying incidences of encephalopathy in cirrhotic patients. With splenorenal anastomoses, the syndrome is demonstrated in 5 to 19 percent, whereas it has been reported in 11 to 38 percent following a portacaval anastomosis. The Warren distal splenorenal shunt has been associated with a reduced incidence of postoperative encephalopathy compared with portacaval and mesocaval shunts.

Hepatic coma has been related to hyperammonemia and ammonia intoxication. Both exogenous and endogenous sources contribute to the blood ammonia level. Dietary protein is the usual source of intestinal ammonia. In patients who bleed, blood within the intestinal tract is also converted into ammonia by bacteria. In the patient with hepatic disease, the ammonia formed within the intestine is carried to the liver but, because of hepatic dysfunction, cannot enter the Krebs-Henseleit (ornithine- citrulline-arginine) cycle. Endogenous urea produced within the gastrointestinal tract also represents an important source of ammonia, and gastric ammonia production from urea is a significant factor in patients with azotemia and cirrhosis. There was less deterioration of maximum urea synthesis following a selective splenorenal shunt than after total shunts.

In the cirrhotic patient with portal hypertension, the two factors implicated in the disturbed ammonia metabolism are impairment of hepatocellular function and portal-systemic collateralization. The blood ammonia level is also raised by increased ammonia production by the kidneys and by muscles that are actively contracting during delirium tremens.

The neuropsychiatric manifestations involve the state of consciousness, motor activity, and deep tendon reflexes. These have been divided into three stages: delirium, stupor, and coma. In the early stages there are mental confusion and exaggerated reflexes. The characteristic "liver flap" may be elicited. In the second stage there is an accentuation of muscular hypertonicity, to the extent of rigidity, and in the final stage there is complete flaccidity. The electroencephalogram is a sensitive indicator of portal-systemic encephalopathy, and the changes antedate clinical manifestations. Blood ammonia level does not define precisely the nature of material measured by standard tests. In patients with hepatic coma, the concentration of ammonia in the blood has correlated well with clinical progress in over 90 percent of patients. An elevated level, over 125 mg/dL, is usually associated with the clinical features of hepatic coma. Treatment with antibiotics negates the value of the test.

Treatment

Treatment is directed at (1) reducing nitrogenous material within the intestinal tract, (2) reducing the production of ammonia from the nitrogenous material, and (3) increasing ammonia metabolism. Since ammonia is an end product of protein metabolism, dietary protein must be reduced drastically to 50 g daily or less. Glucose is included in the diet, since it inhibits ammonia production by bacteria. Gastrointestinal hemorrhage frequently precipitates portal-systemic encephalopathy, with blood acting as a course of ammonia. A major factor in the prophylaxis of hepatic coma is prompt control of active bleeding. Potassium supplements are administered, particularly in patients who are receiving thiazide diuretics, because the rise in blood ammonia that accompanies diuresis has been related to hypokalemia.

The protein substrate on which bacteria can act may be reduced initially by using cathartics and enemas to purge the gastrointestinal tract. If active bleeding has occurred, infused vasopressin plays a dual role in temporarily stopping the bleeding as well as stimulating motility and evacuation of the intestine. Bacteria within the bowel are reduced by administering nonabsorbable antibiotics such as neomycin or kanamycin. In the presence of renal disease, kanamycin is preferred because there is less associated renal toxicity. For patients with severe renal impairment, chlortetracycline is more appropriate because it is not excreted primarily by the kidneys. Lactulose acts as a mild cathartic, and the products of its oxidation by bacteria include lactic and acetic acids, which lower the colonic pH and interfere with ammonia transfer across colonic mucosa. This drug has produced encouraging results in the treatment of hepatic encephalopathy.

Since the colon is the site of most ammonia absorption into the portal circulation, partial colectomy has been suggested as treatment of intractable encephalopathy but has not been effective.

Surgery of Portal Hypertension

The surgical therapy of portal hypertension may be divided into two major categories (1) procedures that directly attack a manifestation of portal hypertension, such as bleeding varices or ascites; and (2) procedures aimed at decreasing the portal hypertension and/or portal venous flow (Table [28-5](#)).

Transesophageal Ligation of Varices and Esophageal Transection

Using either a transthoracic or transabdominal approach, transesophageal ligation of varices and esophageal transection have been directed at controlling bleeding varices. The end-to-end stapler has facilitated the procedure. The procedures do provide temporary control, particularly in children with extrahepatic portal block who are too small to be considered for splenorenal anastomosis.

Technique

Transthoracic ligation (Fig. [28-30](#)) is performed through the eighth left intercostal space. The lower esophagus is freed, but the esophageal hiatus of the diaphragm is not disturbed. An umbilical tape is tightened around the esophagus just above the hiatus to minimize bleeding. A 7-cm longitudinal incision is made through all layers, and three tortuous columns

of veins, coursing longitudinally and communicating with one another, are obliterated with continuous locking sutures of 3-0 chromic catgut. The esophagotomy is closed in two layers with interrupted silk sutures, and the edges of the defect in the mediastinal pleura are reapproximated. Direct ligation of varices can be performed transabdominally.

For transection of the esophagus, the peritoneal cavity is entered, and the esophagogastric junction is exposed. The lower 3cm of the esophagus is mobilized, and care is taken to avoid the vagus nerves. The periesophageal veins are ligated. A high vertical incision is made in the anterior wall of the stomach, and the stapler is inserted, using the largest-size cartridge possible. The esophagus is tied over the center rod 2 cm above the gastric junction. The instrument is fired, resulting in simultaneous transection and reanastomosis of the esophagus (Fig. 28-31).

Esophageal Transection with Paraesophageal Devascularization

This procedure was introduced by Sugiura (Fig. 28-32), who found the perioperative mortality to be about 7 percent in elective cases and 25 percent in emergency cases. Hepatic function is not compromised, and postoperative encephalopathy does not occur. Unfortunately, American studies have not duplicated the Oriental results. The initial and late mortality rates were high, and over half the patients rebled. At least one Western report indicates that better results are achieved with a modified Sugiura procedure compared with portal-systemic results, citing better survival and a lower incidence of encephalopathy accompanied by only a slight increase in recurrent bleeding. However, the consensus is that best results are effected by a shunt. The procedure consists of esophageal transection and paraesophagogastric devascularization, splenectomy, and either selective vagotomy with pyloroplasty or highly selective vagotomy without pyloroplasty.

Procedures for Reduction of Portal Pressure

The operations directed at portal hypertension are based on the consideration that any reduction in portal pressure should decrease the potential for bleeding from varices. Splenectomy is effective only when portal hypertension is due to splenic vein thrombosis or increased flow, as in the massive splenomegaly of myeloproliferative disorders. Reduction of portal pressure by diffuse shunts between a high-pressure portal venous system and low-pressure systemic circulation was attempted. These procedures included omentopexy, posterior mediastinal packing, and transposition of the spleen into the thoracic cavity. Application of Poiseuille's law would suggest that any diffuse shunting would be less effective than a major portal-systemic shunt.

Functionally, portal-systemic shunts have been categorized as either totally or partially diverting portal venous flow away from the liver and also as decompressing or failing to decompress intrahepatic venous hypertension (Fig. 28-33). The end-to-side portacaval shunt prevents blood from reaching the liver by providing complete drainage of the splanchnic venous circulation to the vena cava. The associated alteration of hepatic blood flow is extremely variable, ranging from an increase of 34 percent to a decrease of 53 percent. This shunt also prevents the portal vein from serving as an efferent conduit from the liver. It has been shown that following the end- to-side shunt, the wedged hepatic vein pressure declines. Proponents of the end-to-side shunt have indicated this as the best method of preventing recurrent bleeding from varices, since it most completely decompresses the portal venous system's splanchnic circulation. In refutation, it has been shown that end-to-side and side-to-side portacaval shunts demonstrate equal flow and equivalent reductions of portal pressure.

The idea that the side-to-side portacaval shunt provides the liver with portal venous flow is erroneous in most circumstances. Injection of radioisotope into the distal portal vein results in minimal recovery in the portal vein cephalad to a side-to-side anastomosis. A variant of the side- to-side portacaval shunt uses a small-diameter (8 to 10 mm) portacaval H graft. Patients who maintained portal perfusion postoperatively had improved survival.

Accumulated evidence indicates that the side-to-side portacaval shunt converts the cephalad portion of the portal vein to an outflow from the liver. This is an important feature in patients with Budd-Chiari syndrome. Whether decompression of sinusoidal hypertension is greater with the side- to-side shunt than with the end-to-side shunt has not been defined. There is also concern as to whether the portal vein is beneficial or harmful as an efferent conduit because of a possible siphon effect that may reduce the blood available to the hepatic cell.

The classic end-to-side splenorenal shunt using the central end of the splenic vein also prevents the flow of portal venous blood to the liver if it adequately performs its prescribed function of decompressing esophagogastric varices. The other functional side-to-side shunts, including the end-to-side inferior vena cava–superior mesenteric artery shunt and the interposition of an H graft between these vessels, do not maintain hepatic perfusion with portal venous blood.

The distal splenorenal shunt proposed by Warren and associates is the one procedure that can be classified as truly selective because it decompresses esophagogastric varices while maintaining portal hypertension within the portal veins and hepatic sinusoids. Portal perfusion has been demonstrated in over 90 percent of patients with selective distal splenorenal shunts; the total hepatic blood flow has been shown to be unchanged, while the splenic venous circulation and esophagogastric veins have been reasonably decompressed. Angiographic and other studies have demonstrated that the distal splenorenal shunt loses its selectivity a few months after the operation. Prospective, randomized trials comparing the distal splenorenal shunt with nonselective (portacaval or mesocaval) shunts demonstrate no difference in survival, postoperative complications, or ascites. It is generally believed that the selective shunt is associated with a lower incidence of encephalopathy.

Selection of Procedure

Before a major shunt is performed, the presence of portal hypertension should be defined manometrically. Infrahepatic vena caval pressure should be measured to rule out caval hypertension that would interfere with splanchnic venous decompression. A splenoportogram or a selective superior mesenteric arteriogram with a venous phase should be obtained to determine the status of the major veins. Portal vein thrombosis occurs as a complication in approximately 2 percent of patients with portal cirrhosis. However, lack of visualization of the portal vein in and of itself does not establish this diagnosis, and collateral veins of Sappey must be visualized. If a selective or central splenorenal shunt is being considered, the anatomy of the left renal vein and its relation to the splenic vein should be identified.

Estimations of hepatic blood flow are fraught with errors in interpretation, particularly in cirrhotic patients. Pressure determinations and differentials within the portal venous system, measurements of the estimated hepatic blood flow, and splenoportographic findings did not approximate true flow, and these findings could not be related to the subsequent development of postshunt encephalopathy. The increase in hepatic arterial flow subsequent to the creation of a portacaval shunt has offered a hemodynamic correlate with the patient's prognosis.

The end-to-side portacaval shunt is the procedure most commonly performed, since it is technically easiest and has been associated with the lowest incidence of thrombosis. The presence of a large caudate lobe is less compromising to this procedure than to a side-to-side shunt. For some patients with extensive adhesions from previous operative procedures in the right upper quadrant, splenorenal and mesocaval shunts are preferred. Thrombosis with or without recanalization of the portal vein (cavernomatous transformation) generally precludes a portacaval anastomosis. The Budd- Chiari syndrome, related to endophlebitis of the hepatic veins, dictates a side-to-side shunt to decompress the liver.

In reference to the factor of ascites as a determinant of the decompressive procedure, in one series 39 percent of patients with end-to-side shunts who had preoperative ascites experienced postoperative relief; in 12 percent ascites appeared after the shunt, while all patients with side-to-side shunts and ascites had permanent relief. The splenorenal shunt, which is a functional side-to-side shunt, failed to relieve ascites in 12 percent of patients, and ascites appeared after the shunt in 16 percent. It is therefore thought that ascites per se cannot be considered a significant factor in determining the shunt to be performed. Some series have indicated that the selective splenorenal shunt is not ascitogenic.

Similarly, whether previous encephalopathy and the presence of asterixis are important determinants of the type of shunt has not been resolved. In general, the selective splenorenal shunt is associated with a lower incidence of encephalopathy than other decompressive procedures, but patients with minimal hepatocellular dysfunction rarely develop encephalopathy regardless of the shunt used; devascularization operations rarely cause encephalopathy. Another consideration in shunt selection is the anticipation of performing an orthotopic transplant on the patient in the future. In this circumstance, it is preferable to perform a distal splenorenal shunt.

Selection of Patients

Ascites that failed to respond to medical therapy, a prothrombin time that remained prolonged after parenteral administration of vitamin K, serum albumin level less than 3 g/dL, serum bilirubin level greater than 1 mg/dL, and sulfobromophthalein retention greater than 10 percent are all associated with poor postoperative prognosis. Child divided patients into three groups, including those with good hepatic function (A), those with moderate hepatic function (B), and those with advanced disease and poor reserve (C). In group A are patients with a serum bilirubin level below 2 mg/dL, albumin level above 3.5 g/dL, no ascites, no neurologic disorders, and excellent nutrition. Patients in group B have bilirubin levels between 2 and 3 mg/dL, albumin levels between 3 and 3.5 g/dL, easily controlled ascites, minimum neurologic disorder, and good nutrition. In the C group, the bilirubin level is above 3 mg/dL, the albumin level is below 3 g/dL, and the ascites is poorly controlled, with advanced coma and wasting. Operative mortality following portacaval shunts in group A was zero, in group B was 9 percent, and in group C was 53 percent. There is general agreement that hepatic function is more important than the type of shunt in determining prognosis.

Portacaval Shunt Technique

This shunt (Fig. 28-34) generally is performed through a subcostal incision, but in the presence of extreme hepatomegaly or obesity, a thoracoabdominal approach may be used. The liver is retracted craniad, and a Kocher maneuver is performed to permit mobilization of the duodenum. Dissection is begun in the hepatoduodenal ligament, and the portal vein, which resides posteriorly in relation to the common bile duct, is dissected free along the entire course. Attention is then directed to dissection of the inferior vena cava. The incision in the retroperitoneum is extended, and the anterior and lateral aspects of the inferior vena cava are exposed from the renal veins to the point where the vessel passes retrohepatically. Atraumatic clamps are applied to the portal vein just above its origin and just below its bifurcation, after which the vein is transected as far craniad as possible. The hepatic end is either ligated or oversewn. A sidearm, nonocclusive clamp is positioned along the anterior aspect of the inferior vena cava, and an incision is made in the inferior vena cava wall. This should be approximately one and one half times as long as the diameter of the portal vein. Employing vascular suture techniques, an anastomosis is made between the portal vein and the side of the inferior vena cava, utilizing a continuous suture that is interrupted at the two ends. Following mobilization of the portal vein and inferior vena cava, a side-to-side shunt may be performed. After the anastomosis is completed, the clamp is removed from the inferior vena cava, and then the clamp occluding the portal vein is removed. Pressure should be recorded directly from the portal vein.

Central Splenorenal Shunt

In the cirrhotic patient with postsinusoidal obstruction, the splenorenal anastomosis is not as hemodynamically efficient as the portacaval shunt and is associated with a higher incidence of recurrent bleeding. Proponents of this procedure indicate that in their experience the prevention of recurrent bleeding is similar to that resulting from a portacaval shunt, and the incidence of portal-systemic encephalopathy and persistent hypersplenism is reduced. The operation is generally employed in patients with obstruction of the extrahepatic portal venous system in which the portal vein is not available for shunting. In the pediatric age group, it is preferable to postpone this procedure until the child is 10 years old and the splenic vein is large enough to maintain its patency.

Technique

(Fig. 28-35).

An oblique subcostal incision or thoracoabdominal approach may be used. The transverse colon and splenic flexure are mobilized and retracted caudad, and the short gastric vessels are doubly ligated and transected. The splenophrenic splenorenal ligaments are then transected, and dissection is continued in the hilus until an ultimate pedicle of splenic artery and vein remains. The splenic vein is freed as it courses along the pancreas. The posterior peritoneum is incised just medial to the hilus of the kidney, and the left renal artery and vein are dissected free. Tapes are passed around the main renal vein and the major branches in the hilus of the kidney. Traction on these tapes establishes control of bleeding and minimizes the number of clamps interfering with the anastomosis. The splenic artery is double ligated and transected, and an atraumatic clamp is applied to the central end of the splenic vein. The splenic vein is then transected as close to the hilus of the spleen as possible and brought down to an appropriate site on the anterosuperior aspect of the renal vein. The renal artery is occluded temporarily with a bulldog clamp, and an incision is made in the renal vein. An end-to-side anastomosis is performed by initially securing two stay sutures and completing the posterior layer as a continuous suture. Anastomosis of the anterior layer is accomplished with either horizontal mattress sutures or a continuous suture.

Superior Mesenteric Vein-Inferior Vena Cava Shunt

This operation is generally used for patients with extrahepatic vein obstruction and is particularly applicable to the patient in whom a previous splenorenal shunt failed or to a small child in whom a splenorenal anastomosis is doomed to failure because of the size of the splenic vein. The operation is also advised for patients with cirrhosis if there is associated thrombosis of the portal vein or extensive scarring in the right upper quadrant that precludes safe dissection of the portal vein or marked enlargement of the caudate lobe of the liver.

Interruption of the inferior cava, which is required for the end-to-side shunt, results in venous stasis in the lower extremity, and in the immediate postoperative period the foot of the bed must be elevated to reduce potential edema. The procedure is well tolerated by young patients, in whom postoperative chronic dependent edema of the legs is uncommon. In older patients, any edema may be readily controlled with elastic stockings.

Technique

(Fig. 28-36).

The peritoneal cavity is entered through a midline or right paramedian incision extending from the xiphoid process to well below the umbilicus. Upward traction on the transverse colon exposes the superior mesenteric vessels. The peritoneum is incised in the region of the superior mesenteric arterial pulse, and the superior mesenteric vein is identified and dissected free. The lateral reflection of the ascending colon is then incised along its entire length to permit medial displacement of the transverse and ascending colons and medial reflection of the ascending mesocolon. This exposes the inferior vena cava and the third portion of the duodenum. The inferior vena cava is mobilized from its origin up to the entrance of the right renal vein. The paired lumbar veins are ligated and transected. After the entire inferior vena cava has been freed, vascular clamps are applied immediately below the renal veins and at the junction of the iliac veins. The inferior vena cava is transected as far distal as possible, and the caudal stump is ligated. The right iliac vein may be left attached to the vena cava to achieve greater length. A “window” is created in the mesentery of the small intestine between the ileocolic vessels and the origin of the main ideal trunk to permit approximation of the end of the inferior vena cava to the right posterolateral aspect of the superior mesenteric vein. The anastomosis between the inferior vena cava and the superior mesenteric vein is usually performed proximal to the right colic vein, utilizing a continuous arterial suture interrupted at both ends.

Decompression is more commonly accomplished by the construction of an H graft using a 19- to 22-mm prosthesis interposed between the superior mesenteric vein and the inferior vena cava (Fig. 28-37). A high incidence of immediate and long-term thrombosis of the graft has been reported from some institutions, but this is not uniform.

Selective Splenorenal Shunt

The indications for the selective distal splenorenal shunt include a substantial portal venous flow to the liver, favorable anatomic features related to the site and patency of the splenic vein and the site and size of the left renal vein, and satisfactory liver function in the absence of marked ascites.

Technique

The operative procedure is shown in schematic fashion in Fig. 28-38. Decompression is effected through the short gastric vessels in the spleen. The spleen is not removed, and the distal or splenic side of the splenic vein is used for an anastomosis to the left renal vein. This technique can be modified by transecting the left renal vein close to the hilus of the kidney, turning and anastomosing the caval side of the renal vein to the side of the splenic vein, and ligating the splenic vein close to the confluence with the superior mesenteric vein. Included in the procedure is ligation of the coronary vein and devascularization of the stomach by ligating all vessels with the exception of the right gastric artery and the short gastric veins. Both Inokuchi and Warren have devised modifications of splenopancreatic disconnection to prevent future portal malcirculation.

Complications of Portal-Systemic Shunts

The complications uniquely associated with portal-systemic shunting procedures occur intraoperatively or during the postoperative period. The intraoperative complications include bleeding and a nonshuntable situation, whereas the postoperative complications include rebleeding, hepatic failure, changes in cardiorespiratory dynamics, the hepatorenal syndrome, and delayed complications of hemosiderosis, peptic ulcer, and portal-systemic encephalopathy.

The complication of intraoperative bleeding can be reduced by correction of coagulation defects and by continuing the infusion of vasopressin during the operative procedure. A nonshuntable situation may be related to extension of cavernomatous transformation of the portal vein to involve the superior mesenteric vein and the splenic vein. In these patients, the so-called makeshift shunt, using large collaterals, is generally doomed to failure. The circumstance of caval hypertension caused by hypertrophy and nodularity of the caudate lobe encroaching on the infrahepatic vena cava has been referred to previously. Attempts have been made to shunt between the superior mesenteric vein and the atrium.

Early postoperative bleeding is usually related to thrombosis of a reconstructed shunt. This can be defined by splenoportography in the case of a portacaval shunt. The rapid onset of ascites during the early postoperative period may be managed by restriction of sodium intake and the administration of diuretic agents. In this circumstance, the early institution of a peritoneal-venous shunt has provided dramatic relief. Renal failure following a portal-systemic shunt is not predictable, and treatment consists of supportive measures. The early postoperative development of hepatic coma is an ominous sign.

Portacaval Shunt for Glycogen Storage Disease and Hypercholesterolemia

In 1963, Starzl and associates performed a portacaval transposition on an 8-year-old child with type III glycogen storage disease, which resulted in resumption of weight gain and growth rate. Since then, several investigators have reported clinical improvement following an end-to-side portacaval shunt in children with type I glycogen storage disease in which the enzyme glucose 6-phosphatase is deficient or absent. Intravenous hyperalimentation for 2 to 3 weeks prior to surgery is advised to reduce liver size, restore bleeding time to normal, correct acidosis and hypoglycemia, and promote a favorable outcome of the shunting procedure. Children who have undergone diversion of portal flow have shown no evidence of encephalopathy. Portacaval shunts also have been applied with success to a few patients with homozygous hypercholesterolemia and to an occasional patient with heterozygous type 2 hypercholesterolemia.

FULMINANT HEPATIC FAILURE

Fulminant hepatic failure refers to the clinical syndrome characterized by sudden, severe impairment of hepatic function generally as a consequence of massive necrosis of liver cells. In most instances, the cause is acute hepatitis of viral origin. Massive necrosis and dysfunction have been reported with Reye's syndrome, as a rarity in pregnancy, and following exposure to a variety of drugs. Circulatory necrosis of hepatic cells has been reported following vasopressor therapy and following the inadvertent intraarterial infusion of vasopressin into the hepatic artery.

General supportive therapy is directed at reducing hyperammonemia. A number of intercessions are all aimed at removing so-called noxious elements from the affected patient and preserving life long enough for the liver to regenerate sufficiently. Included in this category are hemodialysis and peritoneal dialysis, exchange transfusion, plasmapheresis, asanguineous hypothermic total-body perfusion, ex vivo perfusion of a liver of the same or different species, and cross-circulation with a human being or a subhuman primate. Orthotopic liver transplantation has led to the survival of several moribund patients and is now the treatment of choice.

Patients in fulminant hepatic failure who are categorized at the comatose level with spasticity have a mortality of between 85 and 90 percent; this approaches 100 percent when the flaccidity stage is reached. The effect of any form of therapy is difficult to evaluate because spontaneous recovery occurs in 10 to 20 percent of patients.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 29 - Gallbladder and Extrahepatic Biliary System

Seymour I. Schwartz

ANATOMY

Duct System

The extrahepatic biliary system begins with the hepatic ducts and ends at the stoma of the common bile duct in the duodenum. The right hepatic duct is formed by the intrahepatic confluence of dorsocaudal and ventrocranial branches. The former enters with a sharp curve, which accounts for the fact that calculi are less common in this segment. The left hepatic duct is longer than the right and has a greater propensity for dilatation as a consequence of distal obstruction. The two ducts join to form a common hepatic duct that is 3 to 4 cm in length. It is then joined at an acute angle by the cystic duct to form the common bile duct (Fig. 29-1).

The common bile duct is approximately 8 to 11.5 cm in length and 6 to 10 mm in diameter. The upper portion is situated in the free edge of the lesser omentum, to the right of the hepatic artery and anterior to the portal vein. The middle third of the common duct curves to the right behind the first portion of the duodenum, where it diverges from the portal vein and hepatic arteries. The lower third of the common bile duct curves more to the right behind the head of the pancreas, which it grooves, and enters the duodenum at the hepatopancreatic ampulla (of Vater), where it is frequently joined by the pancreatic duct. The portions of the duct are referred to according to their relationship to intestinal viscera—*suprapancreatic*, *intrapancreatic*, and *intraduodenal*.

The union of the bile duct and the main pancreatic duct follows one of three patterns. The structures may: (1) unite outside the duodenum and traverse the duodenal wall and papilla as a single duct; (2) join within the duodenal wall and have a short, common, terminal portion; or (3) exit independently into the duodenum. Separate orifices have been demonstrated in 29 percent of autopsy specimens, and injection into cadavers reveals reflux from the common bile duct into the pancreatic duct in 54 percent. Radiographically, reflux from the common bile duct into the pancreatic duct is present in about 16 percent of cases. The sphincter of Oddi surrounds the common bile duct at the ampulla of Vater. This provides control of the flow of bile and, in some cases, pancreatic juice. An ampullary sphincter that is present in one-third of adults may produce a common channel for the terminal common and pancreatic ducts.

Gallbladder

The gallbladder is located in the bed of the liver in line with that organ's anatomic division into right and left lobes. It is a pear-shaped organ with an average capacity of 50 mL and is divided into four anatomic portions: the fundus, the corpus or body, the infundibulum, and the neck. The fundus is the rounded, blind end that normally extends beyond the liver's margin. It may be unusually kinked and present the appearance of a “phrygian cap.” It contains most of the smooth muscle of the organ, in contrast to the corpus or body, which is the major storage area and contains most of the elastic tissue. The body tapers into the neck, which is funnel-shaped and connects with the cystic duct. The neck usually follows a gentle curve, the convexity of which may be distended into a dilatation known as the infundibulum, or Hartmann's pouch.

The wall of the gallbladder is made up of smooth muscle and fibrous tissue, and the lumen is lined with a high columnar epithelium that contains cholesterol and fat globules. The mucus secreted into the gallbladder originates in the tubular alveolar glands in the globular cells of the mucosa lining the infundibulum and neck.

The gallbladder is supplied by the cystic artery, which normally originates from the right hepatic artery behind the cystic duct. It is approximately 2 mm in diameter and courses above the cystic duct for a variable distance, until it passes down the peritoneal surface of the gallbladder and branches. Venous return is carried through small veins, which enter directly into the liver from the gallbladder, and a large cystic vein, which carries blood back to the right portal vein. Lymph flows directly from the gallbladder to the liver and drains into several nodes along the surface of the portal vein. The nerves of the gallbladder arise from the celiac plexus and lie along the hepatic artery. Motor nerves are made up of vagus fibers mixed with postganglionic fibers from the celiac ganglion. The preganglionic sympathetic level is at T8 and T9. Sensory supply is provided by fibers in the sympathetic nerves coursing to the celiac plexus through the posterior root ganglion at T8 and T9 on the right side.

The gallbladder is connected with the common duct system via the cystic duct, which joins the common hepatic duct at an acute angle. The segment of the cystic duct adjacent to the gallbladder bears a variable number of mucosal folds that have been referred to as the “valves of Heister” but do not have any valvular function. Immediately behind the cystic duct resides the right branch of the hepatic artery. The length of the cystic duct is highly variable, though the average is around 4 cm. Variations of the cystic duct and its point of union with the common hepatic duct are surgically important (Fig. 29-2). The cystic duct may run parallel to the common hepatic duct and actually be adherent to it. It may be extremely long and unite with the hepatic duct at the duodenum. It may be absent or very short and have a high (cephalad) union with the hepatic duct, in some cases joining the right hepatic duct instead. The cystic duct may spiral anteriorly or posteriorly in relation to the common hepatic duct and join it on the left side. Congenital biliary atresia is discussed in Chap. 37.

Anomalies

The classic description of the extrahepatic biliary passages and their arteries applies in only about one-third of patients. There are surgically important anomalies in the gallbladder's position and form, and even its number (Fig. 29-3). Isolated congenital absence of the gallbladder is extremely rare; autopsy incidences of 0.03 percent have been reported. Before the diagnosis is made, the presence of an intrahepatic vesicle or left-sided organ must be ruled out. Duplication of the gallbladder with two separate cavities and two separate cystic ducts has an incidence of approximately 1 in 4000. The accessory gallbladder may be situated on the left side, and its cystic duct may empty into the left hepatic duct rather than the common duct. Pathologic processes such as cholelithiasis and cholecystitis may involve one organ while the other is spared.

The gallbladder may be found in a variety of anomalous positions. The so-called “floating gallbladder” occurs when there is an increase in the peritoneal investment. The organ may be completely invested by peritoneum with no mesentery. In other instances, the gallbladder may be suspended from the liver by a complete mesentery, or the neck may have a mesentery in which the cystic artery lies, while the fundus and body are free. This condition occurs in about 5 percent of patients and predisposes to torsion and resulting gangrene or perforation of the viscus. A left-sided gallbladder with the cystic duct entering directly into the left hepatic duct or common duct is extremely rare, as is the situation known as *retrodisplacement*, in which the fundus extends backward in the free margin of the gastrohepatic omentum. The gallbladder may also be totally intrahepatic, a situation that occurs in many animals. In human beings, the partial or complete intrahepatic gallbladder is associated with an increased incidence of cholelithiasis.

Anomalies of the cystic duct were described earlier (see Fig. 29-2). Accessory hepatic ducts are present in approximately 15 percent of cases. Large ducts are usually single and drain a portion of the right lobe of the liver joining the right hepatic duct, common hepatic duct, or infundibulum of the gallbladder. Small ducts (of Luschka) may drain directly from the liver into the body of the gallbladder. When these ducts go unrecognized and are not ligated or clipped at cholecystectomy, an accumulation of bile (biloma) may occur in the subhepatic area.

Anomalies of the hepatic artery and the cystic artery are present in about 50 percent of cases (Fig. 29-4). A large accessory left hepatic artery, originating from the left gastric artery, occurs in about 5 percent of cases. In about 20 percent of cases, the right hepatic artery originates from the superior mesenteric artery, and in about 5 percent of cases there are two hepatic arteries: one originating from the common hepatic and other from the superior mesenteric artery. The right hepatic artery is vulnerable during surgical procedures, particularly when it parallels the cystic duct and is adherent to it or when it resides in the mesentery of the gallbladder. A “caterpillar hump” right hepatic artery may be mistaken for the cystic artery. The right hepatic artery may course anteriorly to the common duct. In 10 percent of cases, the cystic artery originates from the left hepatic artery or from the junction of the left or right hepatic arteries with the common hepatic artery. In about 15 percent of cases, the cystic artery passes in front of the common hepatic duct, rather than to the right of or posterior to this duct. Double cystic arteries occur in about 25 percent of cases, and they may both arise from the right hepatic artery, or one may have another origin.

Cystic Disease of the Extrahepatic Biliary Tract (Choledochal Cyst)

Congenital cystic abnormalities may occur throughout the entire biliary system, i.e., from intrahepatic biliary radicles to the terminal common duct. Intrahepatic cystic dilatation is discussed in Chap. 28. Choledochal cysts are discussed in Chap. 37. There are three major varieties (Fig. 29-5): cystic dilatation involving the entire common bile duct and common hepatic duct with the cystic duct entering the choledochal cyst; a small cyst usually localized to the distal common bile duct; and diffuse fusiform dilatation of the common bile duct.

Congenital biliary atresia is discussed in Chap. 37.

PHYSIOLOGY

Bile Formation

The normal adult with an intact hepatic circulation and consuming an average diet produces within the liver 250 to 1000 mL bile per day. This is, in large part, an active process that takes place within the hepatocytes and is dependent on a supply of oxygen. The secretion of bile is responsive to neurogenic, humoral, and chemical control. Vagal stimulation increases secretion, whereas splanchnic nerve stimulation results in decreased bile flow, probably related to vasoconstriction. The release of secretin from the duodenum after the stimulus of hydrochloric acid, breakdown products of proteins, and fatty acids increases bile flow and the production of an alkaline solution by the canaliculi. Bile salts are also choleretic and augment bile secretion by the liver.

The active transport of bile acids from the hepatocytes into the canaliculi creates an osmotic gradient that causes water to diffuse into those canaliculi. In addition, there is a “bile acid-independent” active transport of electrolytes and other solutes into the canaliculi with consequent passive diffusion of water and solutes.

Composition of Bile

The main constituents of bile are water, electrolytes, bile salts, proteins, lipids, and bile pigments. Sodium, potassium, calcium, and chloride have the same concentration in bile as in extracellular fluid or plasma. As secretion increases, there is an increase in the concentration of bicarbonate and in pH, and a slight increase in chloride. The pH of hepatic bile is usually neutral or slightly alkaline and varies with diet; an increase in protein shifts the pH to the acidic side.

Cholesterol and phospholipids are synthesized in the liver. The rate of cholesterol synthesis, subject to a negative feedback mechanism, is inhibited by high cholesterol intake. Bile acids, produced endogenously or taken orally, reduce cholesterol synthesis and increase cholesterol absorption from the intestine. The synthesis of phospholipids is also regulated by bile acids. The concentrations of cholesterol and phospholipids are both lower in hepatic bile than in plasma.

The principal bile acids, cholic and deoxycholic acids, are synthesized from cholesterol within the liver; they are conjugated there with taurine and glycine and act within the bile as anions that are balanced by sodium. The concentration of these salts within liver bile is 10 to 20 mEq/L. Proteins are present in bile in lesser concentrations than in plasma, with the exception of mucoproteins and lipoproteins that are not present in plasma. Liver bile also contains unesterified cholesterol, lecithin, and neutral fats.

The color of the bile secreted by the liver is related to the presence of the pigment bilirubin diglucuronide, which is the metabolic product of the breakdown of hemoglobin and is present in bile in concentrations 100 times greater than in plasma. After this pigment has been acted upon by bacteria within the intestine and converted into urobilinogen, a small fraction of the urobilinogen is absorbed and secreted into the bile.

Gallbladder Function

The gallbladder provides storage and concentration of bile. The selective absorption of sodium, chloride, and water results in a concentration of bile salts, bile pigments, and cholesterol ten times higher than in liver bile. The gallbladder mucosa has the greatest absorptive power per unit area of any structure in the body. This rapid absorption prevents a rise in pressure within the biliary system under normal circumstances. The absorption of fluid by the gallbladder is driven by an energy-dependent active transport of sodium and a consequent passive transport of water.

Secretion of mucus, approximately at the rate of 20 mL/h, protects the mucosa from the lytic action of bile and facilitates the passage of bile through the cystic duct. This mucus makes up the colorless, “white bile” present in hydrops of the gallbladder resulting from obstruction of the cystic duct. The gallbladder also secretes calcium in the presence of inflammation or obstruction of the cystic duct.

Motor Activity

The passage of bile into the duodenum involves the coordinated contraction of the gallbladder and relaxation of the sphincter of Oddi. Some bile flows from the gallbladder continuously and there are rhythmic contractions occurring two to six times per minute and mediating pressures less than 30 mmH₂O. The gallbladder's emptying, however, is mainly a response to the ingestion of food and the release of *cholecystokinin (CCK)* by the duodenum. CCK also relaxes the terminal bile duct, the sphincter of Oddi, and the duodenum. After the intravenous injection of CCK, the gallbladder is two-thirds evacuated within 30 min. CCK exerts its contractile effects mainly through action directly on the gallbladder smooth muscle cells, but also via interaction with cholinergic nerves. There is a feedback inhibition of CCK secretion by bile acids and proteases. Somatostatin has a direct inhibitory action against CCK-induced gallbladder contraction.

The vagus nerve stimulates contraction of the gallbladder, and splanchnic sympathetic stimulation is inhibitory to its motor activity. Although vagotomy for duodenal ulcer increases the size and volume of the gallbladder, the rate of emptying is unchanged. Parasympathomimetic drugs contract the gallbladder, whereas atropine leads to relaxation. Magnesium sulfate is a potent evacuator of the gallbladder. Emptying of the gallbladder takes place 30 min after ingestion of a fatty meal. There is an increased risk of gallbladder disease in patients on prolonged total parenteral nutrition (TPN) because of the lack of intestinal stimulus and consequent stasis of bile within the organ.

Gallbladder filling occurs when the pressure in the bile duct is greater than that within the gallbladder. This is correlated with reduced CCK levels but is also affected by vasoactive intestinal polypeptide (VIP), pancreatic polypeptide (PP), and peptide YY (PYY).

The common bile duct can be shown to have waves of peristalsis. During starvation, the sphincter of Oddi maintains an intraductal pressure that approximates the maximal expulsive pressure of the gallbladder, i.e., 30 cmH₂O, thereby preventing emptying. During the interdigestive periods, the hormone motilin regulates sphincteric pressure to allow continuous flow of small amounts of bile into the duodenum. After the ingestion of food, the sphincteric pressure is reduced to 10 cmH₂O. When pressure within the extrahepatic bile ducts is greater than 36 cmH₂O, secretion of bile is suppressed.

Biliary dyskinesia lacks objective findings. The term has been used to describe disturbances of biliary tract motility that occur in the absence of anatomic changes. It has been applied as a primary condition and as a complication of biliary tract surgery. Pain has been noted to occur after the ingestion of fatty foods and the injection of CCK at the time that contraction of the gallbladder is induced. Biliary tract pain has also been ascribed to spasm of the sphincter of Oddi. The concept of hyperplastic cholecystosis, characterized by hyperconcentration and excessive emptying of the gallbladder manifest on a cholecystogram, is questionable, but cholecystectomy has been reported to be curative in symptomatic patients.

Enterohepatic Circulation

After the bile enters the duodenum, over 80 percent of the conjugated bile acids are absorbed in the terminal ileum, and the remainder is deconjugated by bacterial activity and absorbed in the colon. Eventually, almost 95 percent of the bile acid pool is absorbed and returns via the portal venous system to the liver. Only 5 percent is excreted in the stool, thereby permitting a relatively small pool of bile acids to have maximal effectiveness. A negative feedback mechanism regulates the hepatic synthesis of bile acids. When the distal ileum has been resected, there is usually adaptation, but occasionally the lack of a feedback mechanism persists and causes significant diarrhea.

DIAGNOSIS OF BILIARY TRACT DISEASE

(See also Jaundice in Chap. [22](#).)

Radiologic Studies

Abdominal Ultrasonography

Ultrasound imaging is the most widely applied diagnostic technique for biliary tract disease in elective and emergent situations. It provides anatomic and pathologic information with great flexibility and portability, and at low cost. The technique may be limited by obesity and large amounts of intestinal gas. Ultrasonography employs a high-frequency vibration in which alternate compression and rarefaction waves travel through the tissue and are reflected off of tissues or items that differ in acoustic impedance. The reflected portion of the sound beam returns to a transducer to create an image. There is variability in the quality of images, and the technique is operator dependent. The gallbladder is readily imaged because echo-free bile contrasts with the organ's wall and the liver parenchyma. The intrahepatic and extrahepatic major bile ducts are also defined. Calculi can be demonstrated in more than 95 percent of cases in which they are present. The discrimination of ductal dilatation has an accuracy of 90 percent.

Ultrasonography is the most cost effective and reliable method for demonstrating gallstones. They appear as reflective foci within the gallbladder or ducts and cast acoustic shadows (Fig. [29-6](#)). A stone impacted in the gallbladder neck or cystic duct may be difficult to detect because the walls themselves return strong echoes. Ultrasonography has been used to guide lithotripsy.

Ultrasound imaging also provides diagnostic information for acute and chronic cholecystitis. The characteristic signs include edema and thickening of the gallbladder wall, occasionally gas in the wall, and absence of visualization of the organ. Thickening and edema of the wall is particularly useful in establishing the diagnosis of acalculous cholecystitis when they are coupled with tenderness over the organ evoked by pressure of the ultrasound probe. Ultrasonography also can establish the diagnoses of hydrops, porcelain gallbladder, adenomas, and carcinomas.

Ultrasound is the first radiologic step in the evaluation of jaundice because it provides a sensitive method for detecting intrahepatic and extrahepatic ductal dilatation. The level of obstruction can be defined by tracing the dilatation down to a point or termination. It can distinguish between intraductal calculi and tumors as the causative agent. Postoperatively it readily defines bilomas and subhepatic abscesses.

Abdominal Radiography

Plain x-ray films of the abdomen are of limited value in assessing patients with gallstones or with jaundice. Supine and upright films of the abdomen may be useful in excluding other causes of abdominal pain, such as a perforated viscus or a bowel obstruction. The presence of significant amounts of calcium within gallstones, which occurs in 15 to 20 percent of patients, causes stones to appear as opacified objects located in the right upper quadrant on plain x-ray films (Fig. [29-7](#)). There are a number of unusual circumstances in which complications from gallstones may be suggested by specific radiographic findings. The presence of gas within the biliary tree outlining its anatomy occurs in patients with a cholecystenteric fistula (abnormal communication between the gallbladder and duodenum, which typically occurs as a consequence of chronic cholecystitis). Opacification of the gallbladder, or of parts of it, occurs in patients with a “porcelain” gallbladder. Gas bubbles may be present in the wall of the gallbladder in patients with emphysematous cholecystitis (infection secondary to anaerobic, gas-producing organisms).

Oral Cholecystography

Oral cholecystography, a relatively simple and effective test for diagnosing gallstones, was introduced by Graham and Cole in 1924. While this test may permit visualization of gallstones within the gallbladder, the critical function that is assessed is the absorptive ability of the gallbladder. A radiopaque iodine containing halogenated dye is orally ingested by the patient. The dye is first absorbed by the gastrointestinal tract and extracted in the liver. The liver excretes the dye into the biliary ductular system, and the dye then passes through the cystic duct into the gallbladder. Ultimately, if the gallbladder has normal mucosal function, the dye becomes concentrated through the physiologic absorption of water and solutes. A “positive” study—one suggestive of gallstones or gallbladder pathology—occurs when stones are noted as filling defects in a visualized, opacified gallbladder (Fig. [29-8](#)) or when the dye is not adequately concentrated and the gallbladder cannot be visualized. When nonvisualization occurs, a second, double dose of contrast medium is frequently administered. Although the accuracy of this modality has been reported to be as high as 95 percent, a number of important limitations have reduced its use. False positives may occur when patients have been noncompliant or have been unable to ingest the tablets because of nausea and emesis or general medical conditions; when the tablets have not been absorbed through the gastrointestinal tract or have not been excreted into the biliary tract as a result of hepatic dysfunction; or when there is some technical problem with the equipment. Oral cholecystography has been largely replaced by the development and refinement of abdominal ultrasonography.

Computed Tomography and Magnetic Resonance Imaging

Computed tomography (CT) is used to assess biliary dilatation and calculi (Fig. [29-9](#)), but because both can be studied more readily by ultrasound imaging, CT is not routinely performed. CT is inferior to ultrasonography for the detection of stones. The major application of CT is to define the course and status of the extrahepatic biliary tree and adjacent structures. Intravenous contrast enhancement of the biliary tract is mandatory in this study. Use of CT is an integral part of the differential diagnosis of obstructive jaundice. Magnetic resonance imaging currently has little application in biliary disease. With the introduction of new contrast agents, its applicability may be expanded.

Biliary Scintigraphy

After technetium 99m-labeled derivatives of iminodiacetic acid (HIDA) are injected intravenously, they are cleared by the Kupffer's cells in the liver and excreted in the bile. Normally, after injection, peak activity is detected in the liver in 10 min, and the biliary ducts can be identified shortly thereafter. The gallbladder is visualized within 60 min in fasting subjects (Fig. [29-10](#)). The test is particularly applicable when the diagnosis of acute cholecystitis is being considered. Evidence of cystic duct obstruction, as indicated by nonvisualization of the gallbladder, is highly diagnostic. The isotopic visualization of the gallbladder essentially precludes the diagnosis. The accuracy of the test in diagnosing acute cholecystitis is about 97 percent.

Percutaneous Transhepatic Cholangiography (PTC)

With fluoroscopic guidance, a small needle is introduced under local anesthesia through the abdominal wall and into the substance of the liver. After position in a bile duct has been confirmed, a guide wire is introduced and a catheter can be placed. PTC facilitates diagnosis by providing a cholangiogram and permits therapeutic intervention, as necessary, based on the clinical situation. The technique has little role in the management of patients with uncomplicated gallstone disease, but it has been particularly useful for patients with more complex biliary problems, including strictures and tumors (Fig. [29-11](#)). PTC is the preferred approach for patients in whom ultrasonography demonstrates intrahepatic ductal dilatation and no extra hepatic ductal dilatation, e.g., Klatskin's tumor at the confluence of the right and left hepatic ducts. As with any invasive procedure, there are potential risks: bleeding, cholangitis, bile leak, and other catheter-related problems. Hematobilia occurs often but is usually self-limiting and of little consequence.

Endoscopic Retrograde Cholangiopancreatography (ERCP)

Using a side-viewing endoscope, the biliary tract and pancreatic duct can be intubated and visualized. This procedure is generally performed with the patient under light intravenous sedation. Advantages of ERCP include direct visualization of the ampullary region and direct access to the distal bile duct, which facilitates diagnosis and therapeutic intervention. This test is generally not required when dealing with patients who have benign gallbladder disease, though it has been of enormous benefit for patients with common bile duct disease (benign and malignant). This is particularly true for the evaluation and treatment of patients with obstructive jaundice when there is dilatation of the common duct or gallbladder (Fig. [29-12](#)).

Choledochoscopy

Rigid and flexible choledochoscopes inserted into the supraduodenal common duct to visualize the lumen of the extrahepatic ducts have been used to determine the presence or absence of calculi; an accuracy of over 90 percent has been reported. The technique is used as an adjunct to operative cholangiography when the common duct is explored. Choledochoscopy can also aid in the removal of stones and bile duct tumors, and in inspecting and obtaining biopsy samples from stenoses.

TRAUMA

Penetrating and Nonpenetrating Injuries of the Gallbladder

Injuries of the gallbladder are uncommon, occurring in 2 to 8 percent of patients with major abdominal trauma. Penetrating injuries are usually caused by gunshot wounds or stab wounds; they also occur rarely during a needle biopsy procedure of the liver. Nonpenetrating injuries are extremely rare. Fewer than 100 cases have been reported, and in only 20 percent was the trauma isolated to the gallbladder.

The types of traumatic injuries to the gallbladder include contusion, avulsion, laceration, rupture, and traumatic cholecystitis. Contusion is difficult to verify but may be associated with vague or temporary symptoms that require no specific therapy. The contused area may undergo necrosis and perforate. Avulsion of the gallbladder from its liver bed occurs as a result of nonpenetrating injury. When the gallbladder's attachments are torn the organ usually hangs by its neck but may be attached only by the cystic duct and artery. Volvulus of the gallbladder may result. Traumatic cholecystectomy, in which the cystic duct, cystic artery, and gallbladder attachments are transected, has been reported. Laceration is the most common type of injury following penetrating wounds but also may result from blunt trauma. Delayed rupture of the gallbladder can occur days to weeks following injury. Traumatic cholecystitis is an unusual condition that occurs as a result of blunt trauma. Bleeding into the gallbladder, from injury of the gallbladder or of the liver, precipitates cholecystitis and sometimes gangrene of the gallbladder. The retained blood may clot and block the cystic duct, in which case the patient presents with the manifestations of hematobilia, including intermittent jaundice, colicky pain, hematemesis, and melena.

Effects of Intraperitoneal Bile

The effects of extravasation of bile into the peritoneal cavity depend on whether or not the bile is infected. When infected bile escapes into the peritoneal cavity, a fulminating and often fatal peritonitis results. When bile is sterile, however, it is well tolerated and results in a chemical peritonitis that may be relatively mild. In the majority of gallbladder injuries, the organ is normal and the bile is sterile. The fact that sterile bile is relatively innocuous is borne out by the very low mortality rate associated with nonpenetrating wounds of the gallbladder. Continuous leakage of noninfected bile, however, is not innocuous. The extravasated bile may produce ascites or become encysted, and extensive chemical peritonitis causes an outpouring of fluid into the peritoneal cavity from the general circulation that may result in shock. There is also some evidence that large amounts of bile salts may be toxic.

Clinical Manifestations

Bile leakage through the penetrating wound suggests the possibility of damage to the biliary system, but duodenal laceration may have a similar manifestation. With blunt trauma, manifestations may be delayed for 36 h or more, in part because typically there are other serious injuries that mask injury of the biliary tract, and sterile bile itself causes only minimal symptoms. The presence of severe shock and pain in the right upper quadrant or lower part of the right side of the chest should raise clinical suspicion of gallbladder injury. The manifestations of bacterial peritonitis may ensue, or if the bile leakage is minimal, the patient may appear to recover but subsequently develop ascites or an intraperitoneal cyst. The finding of bile-stained fluid during diagnostic paracentesis is suggestive, but a negative tap does not exclude gallbladder injury. In most instances the diagnosis is made at celiotomy, emphasizing the need for careful examination of the biliary system after abdominal trauma.

Treatment

The injured gallbladder has been successfully treated by simple suture of the laceration, cholecystostomy, and cholecystectomy. In general, it is preferable to remove the traumatized gallbladder. Cholecystectomy is usually quite easy to perform, since the gallbladder is rarely diseased, and it must be performed if the gallbladder has been avulsed or the cystic artery torn. In the severely ill patient, cholecystostomy may be used for treatment of the extensive laceration or traumatic cholecystitis in order to reduce the time of operative procedure and avoid injury to the common duct. Prognosis is directly related to the incidence of associated injuries.

Injury of the Extrahepatic Bile Ducts

Rare cases of solitary penetrating wounds involving the bile duct have been reported, but there is usually associated trauma to other viscera. Approximately 120 cases of traumatic rupture of the extrahepatic bile duct have been reported, and in 20 cases complete transection occurred. The clinical manifestations are similar to those described for gallbladder injury.

Treatment consists initially of meticulous exploration, particularly if injury to the gallbladder has been excluded and bile has been demonstrated retroperitoneally or within the peritoneal cavity. A Kocher maneuver should be performed to rule out perforation of the common duct behind the duodenum. The presence of hematoma in this region should raise the surgeon's suspicions. Tangential injuries may be treated by primary repair. Complete transection of the common hepatic duct or the common bile duct (e.g., by a penetrating knife wound) may be treated by debridement and an end-to-end anastomosis over a T tube, which should be left in place for several weeks. In most cases of complete transection and injuries caused by blunt trauma, however, the proximal end of the duct should be anastomosed to a Roux-en-Y limb of jejunum. The patient should be placed on an appropriate regimen of antibiotics.

Operative Injury of the Bile Ducts

The great majority of injuries of the extrahepatic biliary duct system are iatrogenic, occurring in the course of laparoscopic or open cholecystectomy. In over 70 percent of cases, the cholecystectomy had apparently been carried out without incident.

Diagnosis

In approximately 15 percent of the cases, ductal injuries are recognized and treated at the time of operation. The remaining 85 percent become manifest by either increasing obstructive jaundice or profuse and persistent drainage of bile through a fistula. Jaundice usually becomes manifest in 2 to 3 days, but in some instances it does not develop for weeks. It may be continuous or intermittent; if intermittent, it is frequently accompanied by attacks of chills and fever, suggesting ascending cholangitis. Hepatomegaly almost always accompanies jaundice if it has been persistent for several weeks, and splenomegaly also may occur if secondary biliary cirrhosis has evolved. Some patients do not display the signs or symptoms of partial or complete blockage until months or years after surgical treatment. Blockage in such cases is the result of increasing fibrosis and narrowing of the channel or of repeated episodes of cholangitis, which in turn leads to fibrosis. *ERCP* or *PTC* most clearly defines the site of obstruction or leak.

Treatment

Patients with jaundice or persistent fistula require a vigorous preoperative regimen that includes a high-protein, low-fat diet and intravenous administration of fat-soluble vitamins, particularly vitamin K. Concomitant portal hypertension with bleeding varices may preclude repair of the common duct; the portal hypertension is usually best treated by a splenorenal shunt because of extensive scarring in the right upper quadrant.

Operative Approach

Injury of the bile duct recognized during surgical operation should be corrected with an immediate reconstructive procedure. Restoration of the continuity of the duct with an end-to-end anastomosis over a T tube may be feasible after a sharp transection, but stricture develops in about half the cases. Direct anastomosis is usually impractical for acute injuries and chronic strictures where the proximal end of the duct should be anastomosed to a Roux-en-Y of jejunum. A mucosa-to-mucosa approximation provides the best long-term results. If this is not feasible, a lateral-lateral anastomosis between the left hepatic duct and a Roux-en-Y limb of jejunum (Hepp- Soupault) is preferable to the Smith transhepatic mucosal pull-through technique. The Longmire operation, with transection of the left lobe of the liver and anastomosis of the jejunum to a large intrahepatic bile duct, has been associated with discouraging results.

The operative mortality of patients with chronic stricture is reported to be 3 to 5 percent. A satisfactory result is obtained in about 70 percent of patients after one or more operative procedures. If the patient is symptom-free 4 years after reconstruction, the cure is almost always permanent.

GALLSTONES

Composition

The major elements involved in the formation of gallstones are cholesterol, bile pigment, and calcium. Other constituents include iron, phosphorus, carbonates, proteins, carbohydrates, mucus, and cellular debris. In Western cultures, most stones are made up of the three major elements and have a particularly high content of cholesterol, averaging 71 percent. Pure cholesterol stones are uncommon, usually large with smooth surfaces, and solitary. Bilirubin pigment stones are also uncommon, with a characteristic smooth, glistening, green or black surface. The pigment stones may be "pure" or consist of calcium bilirubinates. The "pure" pigment stones are usually associated with hemolytic jaundice or situations in which the bile is abnormally concentrated. Increased red blood cell destruction after cardiac valve replacement has resulted in production of gallstones. Calcium

bilirubinate stones are prevalent in Asia, where they constitute 30 to 40 percent of all gallstones.

Formation

Gallstones form as a result of solids settling out of solution. The solubility of cholesterol depends on the concentrations of conjugated bile salts, phospholipids, and cholesterol in bile. Lecithin is the predominant phospholipid in bile, and, although insoluble in aqueous solutions, it is dissolved by bile salts in micelles. Cholesterol is also insoluble in aqueous solution but becomes soluble when incorporated into the lecithin-bile salt micelle. By plotting the percentages of cholesterol, lecithin, and bile salts on triangular coordinates (Fig. 29-13), the limits of micellar liquid in which bile is less than saturated with cholesterol may be defined. Above these limits, the bile is either a supersaturated liquid or a two-phase system of liquid bile and solid crystalline cholesterol.

Perhaps no more than 30 percent of biliary cholesterol is transported in micelles, and of that the majority is carried in a vesicular form. These vesicles are made up of lipid bilayers similar to those found in cell membranes. The vesicles are able to solubilize more cholesterol than are micelles, and the stability of these structures is believed to be the key determinant of cholesterol saturation and precipitation. Current theory suggests that there is an equilibrium between the physicochemical phases of these vesicles such that the formation of liquid crystals may or may not result in actual gallstones. When crystals achieve macroscopic size during a period of entrapment in the gallbladder, gallstones form. The basic secretory defect in nonobese patients is decreased bile salt and phospholipid secretion. Conversely, in obese subjects cholesterol secretion is greatly increased without any reduction in bile salt or phospholipid secretion.

Nucleation is the process by which cholesterol monohydrate crystals form and aggregate. The time required for nucleation is shorter in patients with gallstones than in those without stones. Specific heat-labile glycoproteins within cholesterol-saturated bile induce vesicular aggregation and consequent stone growth. Factors that have been implicated in the formation and precipitation of cholesterol include constitutional elements, bacteria, fungi, reflux of intestinal and pancreatic fluid, hormones, and bile stasis. Constitutional elements are best exemplified in the Pima Indians, of whom 70 percent of females by age thirty and 70 percent of males by age sixty have gallstones. The Masai of Kenya, in contrast, do not have gallstones. Evidence in favor of infection as a cause includes the isolation of such organisms as *Escherichia coli*, *Salmonella typhi*, and *Streptococcus* species from gallbladder walls and from the center of stones in a high percentage of cases, and the demonstration of slow-growing actinomycetes recovered from over half the stones examined in one series. Given that gallstones develop in the absence of infection or inflammation, infection appears not to be a universal factor. In Asians, concretions are known to form about liver flukes and other parasites within the bile ducts.

The reflux factor receives support from the findings of pancreatic enzymes in the gallbladders of patients with cholelithiasis. Trypsin disturbs colloidal balance, and pancreatic phospholipase A can convert lecithin into toxic lysolecithin. Hormones have been implicated in an unproved correlation between calculi and parity, diabetes, hyperthyroidism, and the predominance in females.

Stasis, which includes temporary cessation of bile flow into the intestine and stagnation in the gallbladder, has also been assigned a major role in stone formation. Temporary bile stasis may be due to functional disorders or to a mechanical blockage in the region of the choledochoduodenal junction or the gallbladder. The interruption of bile flow to the intestine is associated with an interruption in enterohepatic circulation, which in turn is accompanied by a decrease in the output of bile salts and phospholipids, reducing the solubility of cholesterol. When more than 20 percent of bile is diverted, the bile salt pool cannot be maintained. Bile salt secretion is also diminished by reduction of the distal third of the intestine, explaining the development of stones in patients with ileal resection or disease. Cholecystectomy causes a greater fraction of the bile salt pool to cycle around the enterohepatic circulation, thereby increasing bile salt and phospholipid secretion.

Solubility has been investigated as a possible regimen to prevent the development of stones in patients at risk as well as to dissolve stones already formed. Chenodeoxycholic acid and ursodeoxycholic acid, which replenish the bile acid pool and reduce cholesterol synthesis and secretion, administered to potential stone formers may return supersaturated bile to its normal composition, preventing stone formation. In one series the drug was administered for 2 years; complete dissolution of radiolucent stones occurred in 13.5 percent of patients. Partial dissolution occurred in 41 percent. The effects were more frequent in women, in thin patients, and in patients with serum cholesterol levels greater than 227 mg/dL. Clinically significant hepatotoxicity was rare.

The direct instillation into the gallbladder of agents that are capable of dissolving cholesterol gallstones has become a reality largely as a result of advances in interventional radiologic technology. Although experience is limited, infusion of a potent cholesterol solvent, methyl- *tert*-butyl ether (MTBE), into the gallbladder via a percutaneously placed catheter has been shown to be effective in selected patients in achieving gallstone dissolution. This procedure is invasive and is therefore associated with some specific risks, including hemorrhage and catheter-related and drug-regulated problems. The major disadvantage of this technology is the high recurrence rate, which approaches 50 percent at 5 years. Lithotripsy has successfully fragmented biliary calculi but generally is not regarded as appropriate therapy because a diseased organ remains to form new stones and the flushing effected by normal bile flow is not equivalent to that of urinary flow.

Pigment stones can be further classified as either “brown” or “black” stones. Brown stones have a characteristic appearance and consistency and are typically found in Asia. These stones presumably occur as a result of infection and are quite similar to primary bile duct stones. Black stones, by contrast, typically are not associated with infected bile. These stones are found in patients with hemolytic disorders or cirrhosis. Altered solubilization of unconjugated bilirubin with precipitation of calcium bilirubinate and insoluble salts is presumed to be the common final pathway for the formation of all pigment stones, regardless of the clinical setting.

Asymptomatic Gallstones

The liberal use of cholecystography and ultrasonography has resulted in the diagnosis of calculi in patients without symptoms referable to the biliary tract. In several large series of asymptomatic patients with gallstones who were followed without surgical treatment, symptoms developed in 50 percent, and serious complications occurred in 20 percent. By contrast, McSherry and associates reported that only 10 percent of patients developed symptoms during a mean 5-year follow-up. Similarly, Gracie and Ransohoff reported a 15-year cumulative probability of developing symptoms of 18 percent for 123 patients with asymptomatic gallstones, and no deaths from gallbladder disease.

The relationship of cholelithiasis and carcinoma of the gallbladder is also of some significance. A review of several series showed that the incidence of calculi in cancer of the gallbladder ranged from 65 to 100 percent, with a mean of 90 percent. Conversely, the incidence of cancer of the gallbladder in patients with symptomatic gallstones ranged from 1 to 15 percent, with a mean of 4.5 percent. Comfort and associates reported no carcinoma among 112 patients with asymptomatic cholelithiasis.

In general, patients with asymptomatic gallstones should not be treated. Dyspepsia, eructations, and flatulence are not regarded as specific symptoms. With the advent of laparoscopic cholecystectomy, the number of cholecystectomies performed has increased. Cholecystectomy for asymptomatic stones may be appropriate for elderly patients with diabetes and for individuals who will be isolated from medical care for an extended period.

Cystic Duct Obstruction

Temporary obstruction to the outflow of bile from the gallbladder is responsible for the most common manifestation of calculous disease, which is biliary colic. This consists of the intermittent spasmodic pain in the right upper quadrant, often radiating to the shoulder or scapula, and precipitated by a fatty or fried meal. The attacks are self-limiting but have a tendency to recur in an unpredictable manner. Significant temperature elevation or leukocytosis are uncommon. The bilirubin and alkaline phosphatase levels are normal or slightly elevated because of an inflammatory process, and hyperamylasemia may be present. The treatment is cholecystectomy, preferably by the laparoscopic approach, and is best performed during that hospitalization but not as an emergent procedure.

Calculi, usually of the cholesterol type, may become impacted in the cystic duct or the neck of the gallbladder, resulting in what is called *hydrops of the gallbladder*. The bile is absorbed, and the gallbladder becomes filled and distended with mucinous material. The gallbladder is generally palpable and tender, and the impacted stone with the resulting edema may encroach on the common duct and cause mild jaundice. Although hydrops may persist with few consequences, early cholecystectomy is generally indicated to avoid the complications of biliary tract infection, empyema, or perforation of the gallbladder. In questionable cases, isotopic scanning of the gallbladder following intravenous CCK can define cystic obstruction or patency.

Choledocholithiasis

Common duct stones may be single or multiple and are found in 4 to 12 percent of cases subjected to cholecystectomy. Most common duct calculi are formed within the gallbladder and migrate down the cystic duct into the common bile duct. Less commonly, stones are thought to form within the ducts. These are classified as *primary stones*, in contradistinction to the *secondary stones* formed in the gallbladder. Primary stones are usually soft, nonfaceted, yellowish brown, and friable. In patients infected with tropical parasites such as *Clonorchis sinensis* and in the Asian population of the Far East, stones may form within the hepatic ducts or the common bile duct itself. Although small stones may pass via the common duct into the duodenum, the distal duct with its narrow lumen (2 to 3 mm) and thick wall frequently obstructs their passage. Edema, spasm, or fibrosis of the distal duct secondary to irritation by the calculi contribute to biliary obstruction. Both extrahepatic and intrahepatic bile ducts become dilated, and there is evidence of laking in the biliary radicles of the liver. There is also thickening of the duct walls and inflammatory cell infiltration. Chronic biliary obstruction may cause secondary biliary cirrhosis with bile thrombi, bile duct proliferation, and fibrosis of the portal tracts. Also associated with chronic obstruction is the development of infection within the bile duct, giving rise to ascending cholangitis and occasionally extending up to the liver, resulting in hepatic abscesses. The offending organism is almost always *E. coli*.

Gallstone pancreatitis is generally associated with the presence or passage of common bile duct stones. The best evidence for this is the frequency with which stones can be found if the stool is filtered at the time of an attack. The frequency with which stones are found in the common duct varies from 1 percent to 70 percent, depending on the time of the operation. At the time of exploration, the pancreas may appear entirely normal, or it may demonstrate edema and rarely necrosis (necrotizing pancreatitis).

Clinical Manifestations

The manifestations of calculi within the common duct are variable. Stones may be present within the extrahepatic duct system for many years without causing symptoms. Characteristically, the symptom complex consists of colicky pain in the right upper quadrant radiating to the right shoulder with intermittent jaundice accompanied by pale stools and dark urine. Biliary obstruction is usually chronic and incomplete but may be acute or complete. If obstruction is complete, jaundice progresses but is rarely intense. In contrast to patients with neoplastic obstruction of the common bile duct or the ampulla of Vater, the gallbladder is usually not distended because of associated inflammation (Courvoisier's law). Liver function tests demonstrate the pattern of obstructive jaundice, and the alkaline phosphatase level usually becomes elevated earlier and remains abnormal for longer periods than the serum bilirubin level. The prothrombin time is frequently prolonged because the absorption of vitamin K is dependent on bile entering the intestine, but a normal level can usually be achieved with parenteral vitamin K. Tests of hepatocellular function generally have normal results. In patients with ascending cholangitis, Charcot's intermittent fever accompanied by abdominal pain and jaundice is characteristic. The diagnosis may be established by *ERCP* or *PTC*.

Treatment

The indications for the removal of common duct stones are: (1) their presence as defined preoperatively in a symptomatic patient or by palpation or cholangiographically at the time of operation; (2) a dilated extrahepatic duct; (3) jaundice; (4) recurrent chills and fevers suggestive of cholangitis; and (5) gallstone pancreatitis.

Common duct stones can be removed by *ERCP*, and the performance of an adequate destruction of the sphincter of Oddi will permit stones that were not extracted or form at a later date to pass into the duodenum without obstruction in the extrahepatic ducts. In a patient undergoing an elective cholecystectomy in whom common duct stones are thought to be present, a preoperative ERCP and sphincterotomy can be followed by laparoscopic cholecystectomy. In some elderly patients ERCP and sphincterotomy have constituted definitive treatment and the gallbladder was not removed.

If common duct stones are detected during laparoscopic cholangiogram, they can be removed by subsequent *ERCP* or during the procedure by trans-cystic duct retrieval or pushing them into the duodenum. Alternatively, the common duct can be opened, the stones extracted, and a T tube inserted. If common duct stones are suspected or detected during open cholecystectomy, the same alternatives apply. The use of the choledochoscope and ureteral baskets facilitates the procedure. In the patient population as a whole, concomitant choledochostomy at the time of cholecystectomy increases the operative mortality by less than 1 percent. In addition, in the face of dilated common duct and multiple stones, a choledochoduodenostomy can provide definitive treatment.

Retained Common Duct Stones

If stones are noted to be present when a T-tube cholangiogram is performed postoperatively (Fig. 29-14), several approaches can be entertained. Small stones, particularly those located in the branches of the hepatic duct, may be disregarded; the majority will remain asymptomatic, and for those that do generate symptoms operative extraction is not associated with significantly increased morbidity. Another approach employs either flushing or chemical dissolution. Capmul 8210, a mono- octanoin, is the agent of choice. The use of heparin, 250,000 units in a 250-mL solution infused every 8 h for 5 days, has been successful.

The mechanical extraction of the retained stone can be performed under radiographic control. Mazzariello reported a 96 percent success rate for 1086 cases, and Burhenne and associates reported a 91 percent success rate for 612 patients managed at 38 hospitals, with no deaths and no significant complications. The T tube is generally left in place for at least 4 weeks after the operation; it is then extracted and a polyethylene catheter is used to instill radiopaque material into the common duct. A Dormia basket is then advanced through the catheter to entrap the stone (Fig. 29-15).

The most commonly used approach is transduodenal papillotomy with extraction of the stone under endoscopic visualization (Fig. 29-16). The success rate for extraction or spontaneous passage after this procedure was 86 percent for 731 collected cases. A complication rate of 7 percent was noted, but two-thirds of complications were treated conservatively. The mortality rate related to the technique was 1.25 percent. Operative intervention is indicated in some cases if there is evidence of obstruction or cholangitis, or if nonoperative methods fail.

Some calculi remain within the liver and may cause irreversible damage. The most common location is a left main hepatic duct that forms a cisterna, and successful treatment is best achieved in this circumstance by resection of the left lobe of the liver. In occasional patients with recurrent hepatic duct stones, a Roux-en-Y limb can be anastomosed to the hepatic duct (usually the left main duct) and positioned so that it can be entered under radiographic guidance to permit stone extraction.

Biliary Enteric Fistula and Gallstone Ileus

A stone in the ampulla of the gallbladder (Hartmann's pouch) can encroach upon and erode the common bile duct. This is known as Mirizzi's syndrome. Operative management depends on the extent to which the common duct has been compromised. If there is only a pressure effect, cholecystectomy is sufficient. If the common duct segment is partially or completely destroyed, a reconstructive procedure is mandated and may require a Roux-en-Y limb anastomosis to the proximal normal duct. When biliary enteric fistulas develop, they usually run between the gallbladder and the duodenum, but 15 percent are cholecystocolic fistulas. Mechanical obstruction of the gastrointestinal tract caused by gallstones is a relatively infrequent occurrence. Gallstone ileus causes 1 to 2 percent of mechanical small-intestine obstructions; the mortality rate is less than 10 percent.

Since cholelithiasis occurs three to six times more commonly in the female than in the male, a higher incidence of gallstone ileus in the female is to be anticipated. Preponderance in the female is actually higher than one would expect, and in several series all patients were female. It is characteristically a disease of the aged, with an average age of sixty-four, and is unusual under the age of fifty.

The process usually begins with formation of the stone within the gallbladder, but cases have been reported in which the gallbladder was not present, having been removed several years prior to the intestinal obstruction. After the gallstone has left the gallbladder, it may obstruct the alimentary tract in one of two ways. Typically, *intraluminal obstruction* is produced by the entrance of the stone into the gastrointestinal tract. Rarely, the stone enters the peritoneal cavity, causing kinking or inflammation and *extrinsic obstruction* of the intestine. The stone may enter the duodenum via the common duct, but this is unusual, and almost always the offending calculus enters through a cholecystenteric fistula. The fistulous tract may connect the gallbladder with the stomach, duodenum, jejunum, ileum, or colon. In addition, internal biliary fistulas may communicate with the pleural or pericardial cavities, tracheobronchial tree, pregnant uterus, ovarian cyst, renal pelvis, and urinary bladder. In a series of 176 fistulas caused by gallstones, the duodenum was involved in 101, the colon in 33, the stomach in 7, and multiple sites in 11.

The fistula probably originates with a stone obstructing the cystic duct, acute cholecystitis, empyema, and the formation of adhesions between the gallbladder and adjacent viscera. Perforation then occurs between the intimately adherent organs, and the stone traverses the fistula. The cholecystenteric fistula then frequently closes, and only a fibrous remnant remains. Having entered the alimentary tract, the gallstone, which is usually single, may be vomited or passed spontaneously via the rectum. The size of the stone is important, since stones smaller than 2 to 3 cm usually pass. When obstruction occurs, the site is usually at the terminal ileum, which is the narrowest portion of the small intestine. Of 154 cases, the duodenum was obstructed in 6, the jejunum in 14, the proximal ileum in 6, the middle ileum in 31, the terminal ileum in 88, the colon in 3, and the rectum in 2. When a gallstone blocks the small intestine, the morbid anatomic and physiologic effects of a mechanical obstruction obtain. There are very large losses of fluid into the intestine. Edema, ulceration, or necrosis of the bowel may occur, and perforation may result.

Clinical Manifestations

A past history suggestive of cholelithiasis is present in 50 to 75 percent of patients. Symptoms of acute cholecystitis immediately preceding the onset of gallstone ileus occur in one-quarter to one-third of the cases. A history of jaundice is present in about 10 percent of the cases. Occasionally, there may be an initial episode of pain suggestive of biliary colic, but major pain is usually not experienced until the intestinal colic results. There is associated cramping, nausea, and vomiting, which may be intermittent. When complete small intestinal obstruction occurs, the vomiting increases and obstipation results. Vomiting is present in almost 100 percent, cramps in 90 percent, distention in 90 percent, obstipation in 78 percent, and feculent vomiting in 67 percent. Serum electrolyte levels reveal the pattern of lower intestinal obstruction with marked hypochloremia, hyponatremia, hypokalemia, and an elevated carbonate level.

The correct preoperative diagnosis is infrequently made, ranging between 13 and 30 percent in several series. The usual diagnosis is that of intestinal obstruction of unknown cause. Radiologic examination may be diagnostic if gas is demonstrated within the biliary tract (Fig. 29-17). Flat, upright, and lateral films plus spot films over the liver are indicated if the diagnosis is considered. The plain x-ray film reveals the pattern of small-intestine obstruction, and a stone is visualized in less than 20 percent of the cases. The diagnosis has also been based on the migration of a previously observed radiopaque gallstone.

Treatment

Biliary enteric fistulas are managed by cholecystectomy and closure by primary repair of the intestinal opening. The patient with gallstone ileus often requires fluid and electrolyte replacement in order to correct deficiency, and a nasogastric tube is used to decompress the stomach. Definitive therapy consists of locating the stone or stones, enterotomy proximal to the stone, and removal of the offending calculi with closure of the intestine. The recurrence rate of gallstone ileus is 5 to 9 percent, and it is important to palpate the entire small intestine, gallbladder, and common duct for retained stones, particularly if the obstructing stone is faceted. Either concomitant or planned interval cholecystectomy and closure of the fistula, if patent, is indicated, since recurrent symptoms or complications develop in one-third of the patients. Carcinoma of the gallbladder has also been present or developed 5 to 16 years after removal of the obstructing gallstone. Performance of concomitant cholecystectomy is determined by the patient's general condition. Many of these patients are extremely ill and depleted, and prolongation of the operative procedure may be contraindicated.

INFLAMMATORY AND OTHER BENIGN LESIONS

Acute Cholecystitis

Acute cholecystitis is usually associated with an obstruction of the neck of the gallbladder or cystic duct caused by stones impacted in Hartmann's pouch. Direct pressure of the calculus on the mucosa results in ischemia, necrosis, and ulceration with swelling, edema, and impairment of venous return. These processes in turn increase and extend the intensity of the inflammation. The ulceration may be so extensive that the mucosa is frequently hard to define on microscopic examination, and segmented leukocytes are found infiltrating all layers. The results of necrosis are perforation with pericholecystic abscess formation, fistulization, or bile peritonitis. In the past, acute cholecystitis secondary to systemic infection occurred most commonly with typhoid fever, but this is now rare. A bacterial cause has been proposed and positive bile cultures have been noted in 60 percent of patients. *E. coli*, *Klebsiella* species, streptococci, *Enterobacter aerogenes*, salmonellae, and clostridia have all been implicated.

Acute cholecystitis, caused by generalized sepsis or by stasis or impaction of a calculus, may occur while the patient is recovering from trauma or an operation. Among other causes of acute cholecystitis are the vascular effects of collagen disease, terminal states of hypertensive vascular disease, and thrombosis of the main cystic artery. Acute cholecystitis in which the gallbladder is devoid of stones is known as acalculous cholecystitis. Less than 1 percent of acutely inflamed gallbladders contain a malignant tumor that may play a role in causing obstruction. The incidences of common duct calculi are similar in acute and in chronic cholecystitis, averaging 7 to 15 percent.

Clinical Manifestations

Most attacks of acute cholecystitis occur in patients who give a history compatible with chronic cholecystitis and cholelithiasis. Acute cholecystitis can occur at any age, but the greatest incidence is between the fourth and eighth decades, and patients over the age of sixty comprise between one-quarter and one-third of the group. Caucasians are afflicted more frequently than blacks, and women more than men.

The onset of acute symptoms is frequently related to a vigorous attempt of the gallbladder to empty its contents, usually after a heavy, fatty, or fried meal. Moderate to severe pain is experienced in the right upper quadrant and epigastrium and may radiate to the back in the region of the angle of the scapula or in the interscapular area. The patient is often febrile, and vomiting may be severe. Tenderness, usually along the right costal margin, often associated with rebound tenderness and spasm, is characteristic. The gallbladder may be palpable, or a palpable mass in the region may be the result of omentum wrapped around the gallbladder. Mild icterus may be present and may be caused by calculi within the ampulla and edema encroaching on the common duct. Moderate to marked jaundice, particularly with a serum bilirubin level greater than 6 mg/dL, suggests the presence of

associated choledocholithiasis but can occur with isolated cholecystitis.

The differential diagnosis includes perforation or penetration of peptic ulcer, appendicitis, pancreatitis, hepatitis, myocardial ischemia or infarction, pneumonia, pleurisy, and herpes zoster involving an intercostal nerve.

The hemogram usually demonstrates leukocytosis with a shift to the left. Radiographs of the chest and abdomen are indicated to rule out pneumonia. A radiopaque calculus is noted in less than 20 percent of cases. The serum bilirubin level may determine the presence of common duct obstruction. Although an elevated amylase level is generally regarded as evidence of acute pancreatitis, levels as high as 1000 Somogyi units have been associated with acute cholecystitis uncomplicated by pancreatitis. To rule out myocardial ischemia, an electrocardiogram should be performed on any patient over the age of forty-five being considered for surgical treatment. Acute cholecystitis may be responsible for some electrocardiographic changes. Oral cholecystography is of limited value because of impaired absorption of dye. An ultrasonogram may demonstrate calculi and/or a thickened wall of the gallbladder and is the diagnostic procedure of choice. Radionuclide scanning with DISIDA (diisopropyl iminodiacetic acid) or PIPIDA (N-para-isopropyl-acetanilide-iminodiacetic acid) is the most effective diagnostic study in this situation.

Treatment

There have been conflicting opinions on the management of acute cholecystitis, particularly on the optimal time for surgical intervention. For the purposes of discussion, early operation is defined as one performed within 72 h after the onset of symptoms; intermediate operation is one carried out between 72 h and the cessation of clinical manifestations; delayed operative management permits the acute inflammatory process to subside; and scheduled elective surgery is performed after an interval of 6 weeks to 3 months. Most surgeons now favor early operation, i.e., with 24 to 48 h. The mortality rate for emergent cholecystectomy ranges from 0 to 5 percent. In the majority of cases, laparoscopic cholecystectomy is successful, but the incidence of conversion to open cholecystectomy is greater in this group of patients when compared to those without acute inflammation. In rare instances of extremely ill patients, cholecystostomy under local anesthesia is applicable.

Emphysematous Cholecystitis

Emphysematous cholecystitis is a rare form of acute, usually gangrenous, cholecystitis, associated with the presence of gas in the gallbladder (Fig. 29-18). Unlike ordinary acute cholecystitis, which is more prevalent among women, emphysematous cholecystitis is more often found in men, with incidences of 75 percent for males and 25 percent for females. Pathogenesis is related to acute inflammation of the gallbladder, which often begins aseptically, complicated by a secondary infection with gas-forming bacilli. These may reach the gallbladder by bile ducts, bloodstream, or lymphatic channels and grow in an anaerobic environment. The clinical manifestations are similar to those of acute cholecystitis. In approximately half the patients, a history of previous gallbladder attacks can be elicited. Cholelithiasis is also present in half the patients, who are frequently diabetic.

The diagnosis is usually made on the basis of radiographs that show a globular, gas-filled shadow in the region of the gallbladder. Later, intramural or submucosal gas may appear, and gas may also appear in the pericholecystic area, denoting extension of the pathologic process outside the confines of the gallbladder. The treatment of choice is early operation, since the incidence of free perforation is reported to be 40 to 60 percent. Cholecystectomy is indicated, but if it is not feasible, cholecystostomy should be performed. In 9 percent of cases, choledocholithiasis is present, and exploration of the common duct may be required. Although positive bile cultures are found in only half the cases, antibiotics directed toward the clostridial and coliform organisms are indicated. The mortality rate is significantly greater than that for nonemphysematous cholecystitis.

Chronic Cholecystitis

Chronic inflammation of the gallbladder is generally associated with cholelithiasis and consists of round cell infiltration and fibrosis of the wall. Buried crypts of mucosa (Rokitansky-Aschoff sinuses) may be seen dipping into the mucosa (Fig. 29-19). Obstruction by gallstones of the neck of the cystic duct may produce a mucocele of the gallbladder (hydrops). The bile is initially sterile but may be secondarily infected with coliform bacilli, *Klebsiella* species, streptococci, and occasionally clostridia or *Salmonella typhi*. Secondary effects of cholecystitis include obstruction of the common duct, cholangitis, perforation of the gallbladder with formation of a pericholecystic abscess or a cholecystenteric fistula, bile peritonitis, and pancreatitis. There may be associated carcinoma of the gallbladder.

Clinical Manifestations

The patients generally present with moderate intermittent abdominal pain in the right upper quadrant and epigastrium, occasionally radiating to the scapula and interscapular region. There is usually a history of intolerance of fatty or fried foods, and the patient may have noted intermittent nausea and anorexia. If the patient is not experiencing acute pain, there may be no diagnostic findings on physical examination. Occasionally tenderness is elicited over the gallbladder. Diagnosis is usually established by ultrasound scanning or an oral cholecystogram, which demonstrates either the absence of filling of the gallbladder or the presence of stones.

Hyperplastic Cholecystoses

Hyperplastic cholecystoses are characterized by the proliferation of normal tissue elements. The two most common of these lesions are cholesterolosis and adenomyomatosis. In patients with cholesterolosis there is evidence of cholesterol deposition within the epithelial cells of the lamina propria. The bile of these patients contains significantly more cholesterol than that of normal adults, and the abnormality presumably arises from some aberration in cholesterol transport and absorption by the gallbladder epithelium. The deposition of cholesterol within the wall gives rise to the gross description of the "strawberry gallbladder." Adenomyomatosis is characterized by hyperplasia of the muscle and mucosa of the gallbladder. Malignant degeneration is unusual, and it is unclear whether these disorders can truly be the source of symptoms. Cholecystectomy should be offered to these patients only if warranted by symptoms.

Treatment

The treatment of chronic cholecystitis and cholelithiasis is cholecystectomy, and the results are usually excellent. Laparoscopic cholecystectomy is the procedure of choice. Early cholecystectomy is particularly important for the diabetic patient. Operative mortality of less than 1 percent has been reported for large series. Seventy-five percent of patients undergoing cholecystectomy for cholelithiasis are completely relieved of all preoperative symptoms, and the remaining 25 percent have only mild symptoms that are apparently unrelated to the biliary system.

Acalculous Cholecystitis

Acute and chronic inflammatory disease of the gallbladder can occur without stones. Acute acalculous cholecystitis frequently is a complication of burns, sepsis, multiple system failure, cardiovascular disease, diabetes, prolonged illness, or a major operation.

The incidence of chronic acalculous cholecystitis is difficult to establish. It is present in over 50 percent of children and 35 percent of Nigerians with gallbladder disease, and the accepted incidence of adults in the United States is less than 5 percent of cases of cholecystitis. Possible causes include (1) anatomic conditions such as kinking, fibrosis, and obstruction of the cystic duct by tumor or anomalous vessels; (2) thrombosis of major blood vessels, producing ischemia and gangrene; (3) spasm or fibrosis of the sphincter of Oddi in patients with a "common channel" with or without associated pancreatitis; (4) systemic diseases such as diabetes mellitus and collagen diseases; (5) specific infections such as typhoid fever, actinomycosis, and parasitic infestation; and (6) scarlet fever and a wide variety of febrile illnesses in young children. The DISIDA or PIPIDA scan and the ultrasound scan are occasionally normal in these patients, but characteristically the ultrasound demonstrates thickening of the wall. Percutaneous cholecystostomy has been used successfully (80 percent) for diagnosis and treatment of acalculous cholecystitis.

Treatment

Cholecystectomy is preferable, but in one series the patient's condition mandated cholecystostomy in 14 of 16 cases. In children with acute febrile illness, cholecystostomy has been particularly effective, and subsequent cholecystectomy has not been required in many of these patients.

Cholangitis

Infection within the biliary duct system is most frequently associated with choledocholithiasis but also has accompanied choledochal cysts and carcinoma of the bile duct, and has followed sphincteroplasty. Infection and inflammatory changes may extend up the duct system into the liver and give rise to multiple hepatic abscesses. Clinically, the condition is characterized by intermittent fever, upper abdominal pain, exacerbation of jaundice, pruritus, and at times rigor.

In patients with common duct stones in whom there is ascending cholangitis, a broad-spectrum antibiotic directed particularly at *E. coli*, which is the most common offending organism, should be given for several days before surgical treatment. Antibiotics usually control the infection, but if the patient's temperature does not fall, surgical drainage should not be delayed. This can be accomplished percutaneously by the transduodenal or transhepatic routes or operatively.

Acute Suppurative Cholangitis

Suppurative cholangitis, in which there is gross pus within the biliary tract, constitutes one of the most urgent causes for laparotomy in patients with obstructive jaundice. The condition was first described in 1877 by Charcot, who suggested a diagnostic triad of jaundice, chills and fever, and pain in the right upper quadrant. To these, Reynolds and Dargan added shock and central nervous system depression as specific identifying features of the condition.

The disease occurs almost exclusively in patients over 70 years of age. All patients are febrile, and a majority are jaundiced. Hypotension, confusion, or lethargy occurs in about 20 percent of cases. A white blood cell count of less than 12,000/mm³ has been reported in over half the patients, probably related to the age and lack of marrow response. Bilirubin, SGOT, and alkaline phosphatase levels are characteristically elevated, but the serum amylase level is usually normal. The correct diagnosis has been made in less than one-third of the patients. Patients have been managed emergently by establishing initial drainage via *ERCP* or *PTC* followed by definitive operation.

At operation, all patients demonstrate gross distention of the common bile duct, with frank pus, frequently under considerable pressure, and choledocholithiasis or a tumor obstructing the distal bile duct. If the gallbladder is present, it is invariably distended and inflamed. Spontaneous perforation of the bile ducts has been reported. Surgical treatment is directed at rapid decompression of the duct system and is combined with large doses of antibiotics, particularly those that achieve high levels in the bile. In a review of the literature, it was reported that all patients who were not operated on died, and mortality following drainage or surgical procedures ranged from zero to 88 percent, averaging 33 percent.

Cholangiohepatitis

Cholangiohepatitis, which is also known as recurrent pyogenic cholangitis, is found almost exclusively among the Chinese, with the largest number of cases seen among Cantonese living in the Pearl River delta in China. In Hong Kong it is the most commonly encountered disease of the biliary passages and is the third most common abdominal surgical emergency after appendicitis and perforated ulcer. It has also been encountered in Great Britain, in Australia, and in the Chinese population in the United States. Cholangiohepatitis

occurs most frequently in the third and fourth decades but has been reported at all ages and occurs with equal frequency among men and women.

The etiology of cholangiohepatitis is summarized in Fig. 29-20. The pyogenic element probably originates from the bowel and is caused by *E. coli*, *Klebsiella* species, *Bacteroides* species, or *Enterococcus faecalis*. In most instances, positive cultures are obtainable from the bile and the portal venous blood. The Chinese liver fluke, *C. sinensis*, was thought to be an important contributing factor. Other factors that have been implicated as contributing causes of cholangiohepatitis include ascariasis and hemolysis associated with malaria.

Pathology

The gallbladder wall is thickened but not grossly inflamed. The common bile duct is also usually grossly distended and contains large stones. The stones are produced by precipitation of bile pigments, desquamation of epithelium, and products of inflammation; the nucleus of the stone may contain an adult *Clonorchis* worm, an ovum, or an ascarid. Acute or hemorrhagic pancreatitis occurs in less than 1 percent of cases. The most marked changes occur in the liver, where the intrahepatic bile ducts are both dilated and constricted. Inflammatory changes are present in the periductal tissue and may progress to frank abscess formation.

Clinical Manifestations

In highly endemic areas, cholangiohepatitis is the first consideration in patients with jaundice, pain, and pyrexia. Pain is usually located in the right upper quadrant and epigastrium and may be colicky or constant. In most acute attacks there is fever accompanied by chills and rigors, and 50 percent of the patients are jaundiced, while the remainder have an elevation of the serum bilirubin level. Recurrence of symptoms is one of the most characteristic features of the disease.

Most patients appear to be in a toxic condition, with temperatures up to 40°C. There is tenderness and guarding in the right upper quadrant. The white blood cell count is usually about 15,000/mm³, and the serum bilirubin level is generally above 2mg/dL, with accompanying bilirubinuria. There may be evidence of impairment of hepatocellular function. In the majority of cases calculi are not demonstrable on routine x-ray films. An occasional finding of significance is the presence of gas in the biliary tree, which may be due to a secondary gas-forming organism or a fistula between the duct and duodenum. *ERCP* or *PTC* may establish the diagnosis. Manifestations of portal hypertension may be present.

Treatment

Patients are generally prepared with antibiotics. Surgical therapy, however, should not be delayed for the patient who is jaundiced and has pain and pyrexia. The operation consists of removal of the stones and debris from the extrahepatic bile ducts followed by establishment of open drainage between the involved ducts and intestine, usually with a Roux-en-Y limb. Anchoring the Roux-en-Y limb to the anterior abdominal wall facilitates subsequent repeated dilatation and stone extractions.

If large hepatic abscesses are noted, drainage should be performed. Left hepatic lobectomy has been carried out on occasion, when there has been gross dilatation of the ducts and abscess formation in the left lobe while the right was apparently normal.

The prognosis is generally guarded, since recurrence is not uncommon. In one study, common duct exploration, transhepatic intubation, and hepatotomy were associated with recurrence rates of 24, 37, and 75 percent, respectively. Hepatic resection had a failure rate of only 4 percent, and none of the patients had recurrent stones. In advanced cases, particularly with multiple abscesses, the prognosis is poor, and the patient eventually succumbs to liver failure, septicemia, or cholangiocarcinoma.

Sclerosing Cholangitis

Sclerosing cholangitis is an uncommon disease that involves all or part of the extrahepatic biliary duct system and often affects the intrahepatic biliary radicals as well. The disease has also been called *obliterative cholangitis* and *stenosing cholangitis*, in reference to a progressive thickening of the bile duct walls encroaching upon the lumen. It may be associated with gallstones, but several series have been presented in which there were no stones in the gallbladder or the common duct. A significant number of cases have been associated with ulcerative colitis, Crohn's disease, Riedel's struma, retroperitoneal fibrosis, and porphyria cutanea tarda.

The cause of sclerosing cholangitis is unknown. Histologic sections in several cases failed to reveal any granulomatous lesion, metaplasia, or neoplasia. In several series, none of the patients had previous surgical treatment, and therefore local trauma was excluded as an etiologic agent; irritation of the common duct by passage of calculi is unlikely given that there are usually no stones present in either the common duct or the gallbladder. It has been suggested that the disease may be caused by local response to viral infection, since a relative lymphocytosis with atypical lymphocytes has been noted. Immune response and collagen disease have also been considered as possible causes. A positive cellular immune response to biliary antigens has been demonstrated. The disease has been noted in patients with HIV infection.

Pathology

Grossly, there is diffuse thickening of the wall of the extrahepatic biliary tract and sometimes of the intrahepatic ducts, with a concomitant encroachment on the lumen, resulting in marked luminal narrowing. The duct system may be completely involved, or the hepatic ducts may be spared and the disease restricted to the entire length of the common duct. The gallbladder is usually not involved, but the lymph nodes in the region of the common duct and foramen of Winslow are usually markedly enlarged and succulent. Microscopic analyses of the affected duct show that the walls are as much as eight times thicker than normal. The areas of inflammation and fibrosis are in the submucosal and subserosal portions, with an edematous field between them. The mucosa is intact throughout. Biopsy examination of the liver may reveal bile stasis or, in long-standing cases, biliary cirrhosis. The histologic evaluation is critical, since it is difficult to differentiate this disease from sclerosing carcinoma of the bile ducts.

Clinical Manifestations

The diagnosis is to be considered in patients (particularly middle-aged men) with a clinical and laboratory picture of extrahepatic jaundice. Jaundice is usually associated with intermittent pain in the right upper quadrant, nausea, vomiting, and occasionally chills and fever. In long-standing cases with biliary cirrhosis, the manifestations of portal hypertension, such as bleeding varices and ascites, may be apparent. The diagnosis has been established by *ERCP*. At operation a dense inflammatory reaction in the region of the gallbladder and gastrohepatic ligament is noted. Palpation of the duct reveals a cordlike structure that may feel like a thrombosed blood vessel, but the wall of the common duct is obviously thickened and cuts with difficulty. The edges of the incision characteristically pout out. Usually only a fine probe or small Bakes dilator can be inserted into the lumen. Cholangiography may vividly demonstrate the extensive narrowing of the lumen (Fig. 29-21).

Treatment

The appropriate management of sclerosing cholangitis remains unclear. No drug therapy has achieved consistent, or even usual, success.

The asymptomatic anicteric patient is not treated and is not studied with repeated cholangiograms if jaundice or cholangitis does not develop. The pruritic and icteric patient is treated for 4 to 6 weeks with prednisone; if there is no improvement, or if cholangitis is present or develops, an operation is performed with a preoperative cholangiogram as a guide. If there is minimal intrahepatic involvement and dilatation of a segment of the common duct or common hepatic duct proximal to marked stenosis, the stenotic segment is excised as a biopsy section to rule out cholangiocarcinoma, and a direct mucosa-to-mucosa anastomosis is effected between the dilated segment of duct and a Roux-en-Y limb of jejunum, preferably without a stent. Stricture of the confluence of the hepatic ducts is managed by excision of the distal ducts for pathologic evaluation and anastomosis of the hepatic ducts to the Roux-en-Y limb of jejunum by the mucosa-to-mucosa technique. If the hepatic ducts are sufficiently dilated, no stent is used. If these ducts are small, transhepatic stents are used, but no attempt is made to dilate intrahepatic ducts.

Data from several large centers suggest that selected patients with primarily extrahepatic disease can be successfully managed with hepaticojejunostomy and long-term stenting. In patients with more diffuse or advanced parenchymal disease, hepatic transplantation has become the procedure of choice. The role of transplantation might be extended as we begin to understand more fully the risk of cholangiocarcinoma developing in patients with sclerosing cholangitis.

Fibrosis or Stenosis of the Sphincter of Oddi

In 1884 Langenbuch, only 2 years after reporting the first successful removal of a gallbladder, suggested transduodenal division of the "diverticulum" of Vater in cases of cicatricial stenosis for chronic inflammation. In 1901 Opie called attention to the "common channel" theory as the cause of pancreatitis, and in 1913 Archibald suggested sphincteroplasty as the treatment for pancreatitis.

The pathogenesis of fibrosis or stenosis of the sphincter of Oddi and the papilla of Vater is not fully understood. Long-standing spasm may play an important role, and infection of the biliary tract or pancreas has also been implicated. Irritation from stones within the common duct may also lead to fibrosis. In a series of 50 patients in whom sphincteroplasty was performed because a small Bakes dilator could not be passed through the sphincter of Oddi, biopsy analysis revealed no abnormalities in 18, while 18 showed inflammatory infiltration, 17 had minimal fibrosis, and 2 had diffuse fibrosis. No definite correlation could be found between the various manifestations of biliary tract disease and the histologic changes.

Clinical Manifestations

The main symptom of fibrosis or stenosis of the sphincter of Oddi is abdominal pain, usually colicky and frequently associated with nausea and vomiting. The pain begins in the right upper quadrant and radiates to the shoulder, and it may be intermittent. Over half the patients give a history of intermittent jaundice, and many indicate that they have had previous cholecystectomy without relief of symptoms.

Treatment

The diagnosis is generally made when there is difficulty in passing a No. 3 Bakes dilator through the ampulla of Vater. Cholangiography and pressure studies on the common bile duct have theoretical application. If a 3-mm dilator cannot be easily passed through the ampulla, a transduodenal exploration should be carried out. Thomas and associates compared the results of transduodenal sphincteroplasty and choledochoduodenostomy in 30 patients with stenosis or stricture of the sphincter. The procedures were equally and highly effective, and neither was associated with a significant incidence of subsequent cholangitis. Sphincteroplasty is preferable if the common duct is small, and a transduodenal approach is indicated if an ampullary tumor is suspected. Endoscopic papillotomy has been used successfully, particularly in Europe, where the disorder is more frequently diagnosed.

Papillitis

In 1926 DelValle first described a benign inflammatory and fibrous process of the ampulla of Vater and indicated that it was a factor in producing stenosis. It was postulated that acute and subacute inflammatory changes occur and that stenosis is the final and irreversible result of these changes. Acosta and Nardi have presented 61 cases of papillitis, 21 of which were chronic ulcerative papillitis, 20 chronic sclerosing papillitis, 15 chronic granulomatous papillitis, and 5 chronic adenomatous papillitis. The acute stage, which is characterized by edema, papillary dilatation, hemorrhage, and infiltration, may be reversible, whereas sclerosing papillitis and chronic granulomatous papillitis are considered

irreversible in view of their inevitable evolution into scar tissue.

The clinical and pathologic features associated with papillitis include the postcholecystectomy syndrome in 30 percent, dilatation of the common duct in 50 percent, biliary disease without stones in 25 percent, obstructive jaundice in 60 percent, pancreatitis in 70 percent, and liver damage in 25 percent. There has been no correlation between the specific clinical syndromes and the pathologic changes. A pancreatic evocative test, using morphine-neostigmine or secretin-CCK, has been used. Elevation of at least one serum pancreatic enzyme level by a factor of four over the normal level, coupled with reproduction of the patient's pain, is considered a positive test result. The efficacy of this test has been disputed. Since the majority of patients with papillitis have irreversible lesions, sphincteroplasty is generally employed.

TUMORS

Carcinoma of the Gallbladder

Carcinoma of the gallbladder accounts for 2 to 4 percent of gastrointestinal malignancies. Its occurrence in random autopsy series is about 0.4 percent, and approximately 1 percent of patients undergoing biliary tract operations have carcinoma either as an anticipated diagnosis or found incidentally.

Etiology

Approximately 90 percent of patients with carcinoma of the gallbladder have cholelithiasis, but the pathogenesis has not been defined. There is also an association with polypoid lesions of the gallbladder. Areas of dysplasia have been noted in juxtaposition to larger (greater than 2.5 cm) stones and in adenomas. Malignant changes have been noted more frequently in polypoid lesions greater than 10 mm. The calcified "porcelain" gallbladder is associated with a 20 percent incidence of gallbladder carcinoma.

Pathology

Approximately 80 percent of the tumors are adenocarcinomas (75 percent of these are scirrhous, 15 percent polypoid, and 10 percent mucoid). Squamous carcinomas, adenoacanthomas, and melanomas occur rarely. The routes of metastasis include spread along the lymphatics to the choledochal, peripancreatic, and periduodenal nodes. There is often localized invasion of vessels within the wall of the gallbladder, and the tumor frequently extends transmurally into the parenchyma of the liver. When metastases are present, the liver is involved in two-thirds of patients, the regional lymph nodes in about one-half, and the omentum, duodenum, colon, or porta hepatis in about one-fourth. Early reports suggest more frequent and rapid recurrences of carcinomas of the gallbladder after laparoscopic cholecystectomy. By 1995, 15 cases of trocar-site metastases from unsuspected gallbladder carcinoma following laparoscopic cholecystectomy were reported.

Clinical Manifestations

Signs and symptoms of carcinoma of the gallbladder are generally indistinguishable from those associated with cholecystitis and cholelithiasis. These include abdominal discomfort, right upper quadrant pain, nausea, vomiting, and weight loss. Half the patients are jaundiced, and two-thirds of those with clinical manifestations have a palpable mass. Laboratory findings are not diagnostic. Ultrasound or CT scan may suggest the diagnosis.

Treatment

Most long-term survivors are patients who underwent cholecystectomy for cholelithiasis and in whom the malignancy was an incidental finding. The management of these patients is controversial. There is some suggestion that T2 or more advanced tumors have a better prognosis if a radical second procedure, which includes lymphadenectomy and partial hepatic resection for lesions located adjacent to the liver, is performed. For lesions that are apparent at operation, removal of the hepatoduodenal nodes, resection of segment IV/V or extended right hepatectomy and, in some cases, pancreaticoduodenectomy has been reported to improve survival. By contrast, other reports have indicated that there has been no improvement associated with these procedures. Some groups have reported that adjuvant radiation therapy or chemotherapy improves survival.

Prognosis

Large cumulative series report 5-year survival rates of 5 percent. The overwhelming majority of survivors are in the group that had incidentally diagnosed tumors. T1 lesions have a 5-year survival rate of approximately 100 percent. T2 lesions have a survival rate of 40 percent when subjected to the more radical operations. In some series, no difference was noted when node-negative and node-positive patients were compared.

Bile Duct Carcinoma

Pathology

The autopsy incidence of bile duct carcinoma is about 0.3 percent. Unlike gallbladder carcinomas, bile duct tumors occur more frequently in men. There is no evidence that bile duct stones have a role, and the relationship between the tumors and sclerosing cholangitis remains ill-defined. Approximately two-thirds of the lesions are located in the proximal ducts, often at the confluence of the right and left main hepatic ducts (Klatskin tumors). The tumors are generally small but involve the whole thickness of the duct, growing in a scirrhous concentric manner and resulting in ductal obstruction. The proximal lesions often extend into the hepatic parenchyma. Rarely, the ductal tumors are polypoid tumors, which are associated with a more favorable prognosis, as is the case with carcinomas of the ampulla of Vater.

Ductal lesions are cholangiocarcinomas of the adenocarcinoma type. Most tumors are well differentiated and associated with a marked fibrous reaction. Perineural involvement is common. Intraoperative frozen-section diagnosis is often difficult. The liver and regional lymph nodes are the most frequent sites of metastasis. The incidence of metastasis at operation is 50 percent. In some cases multicentric ductal tumors have been reported.

Clinical Manifestations

Characteristically, patients present with the recent onset of jaundice, acholic stools, and dark urine. The jaundice is often preceded by pruritus. Almost all patients have had significant weight loss associated with loss of appetite. Half the patients have abdominal pain; cholangitis may result from the obstruction. The gallbladder is palpable in one-third of the patients who have distal lesions and is not felt with proximal tumors.

The laboratory findings are compatible with the diagnosis of obstructive jaundice with elevation of the bilirubin and alkaline phosphatase and mild increase of the transaminases. Increased level of carcinoembryonic antigen (CEA) can be detected in the bile. Ultrasound scanning demonstrates intrahepatic ductal dilatation and distention of the extrahepatic ducts proximal to the point of obstruction. CT scanning will also define the extent and location of ductal dilatation; it rarely demonstrates the tumor itself. Precise demonstration of the site of obstruction is achieved by *PTC* or *ERCP*. In the face of isolated intrahepatic ductal dilatation, the former is preferable, and when there is evidence of distention of the gallbladder or extrahepatic ducts, ERCP is more rewarding (Fig. [29-22](#)).

Treatment

Treatment is directed at resecting the tumor, if possible, or palliation by relieving the obstruction. Cure can be achieved only by surgical removal of the lesion, while palliation can be effected by operation, radiologic intervention, or endoscopic decompression. Curative resection generally entails removal of the common duct and the common hepatic duct up to and sometimes including the confluence of the right and left hepatic ducts, followed by anastomosis of the proximal dilated system to a Roux-en-Y limb of small intestine (Fig. [29-23](#)). The resectability of proximal tumors is about 20 percent, and the cure rate remains under 15 percent. If a proximal lesion extends into the liver parenchyma, varying amounts of the liver are removed en bloc, and proximal intrahepatic ducts are anastomosed to the intestine (Fig. [29-24](#)). Orthotopic liver transplants have been performed for intrahepatic tumors with some long-term survivors. The cure of distal bile duct tumors has been improved by radical lymphadenectomy and pancreaticoduodenectomy (Whipple procedure).

Surgical palliation is performed by anastomosing the dilated ductal system proximal to the point of obstruction to the limb of intestine. In more proximal lesions, this can be achieved by using the extrahepatic portion of the left main hepatic duct (Fig. [29-25](#)). Transection of the liver to the left of the falciform ligament and anastomosing the intestine to the dilated hepatic duct (Longmire procedure) rarely provides long-term relief of jaundice. Using *PTC* or *ERCP*, a stent can be passed through the tumor to provide drainage of the dilated ducts. These require frequent changes. Effort should be directed to providing drainage into the intestine rather than externally because external drainage of bile prevents an enterohepatic circulation and is accompanied by loss of appetite. There have been conflicting reports regarding the efficacy of adjuvant radiotherapy and chemotherapy for either cure or palliation.

OPERATIONS OF THE BILIARY TRACT

Perioperative Considerations

Prophylactic antibiotics are not indicated for patients undergoing elective cholecystectomy unless there are specific risk factors. These include jaundice, common duct stones, diabetes, and age greater than 65 years. Interventional procedures such as *PTC* or *ERCP*, especially in patients with biliary tract obstruction, should be covered with prophylactic antibiotic therapy. The antibiotic usually is selected with the assumption that the most likely involved organisms are *E. coli*, *Klebsiella* species, and enterococci. A second-generation cephalosporin is appropriate. The drug is administered before the operation or intervention, and two subsequent doses are given at 6-h intervals.

Preoperative decompression of the proximal distended ducts in patients with obstructive jaundice does not significantly improve the outcome and has been associated with an increased incidence of complications. Short-term drainage may be indicated as a bridge to an operation in a patient with cholangitis and sepsis. The preoperative placement of catheters transhepatically may facilitate dissection of proximal tumors, and they can be left in place as stents that traverse anastomoses.

Cholecystostomy

Cholecystostomy accomplishes decompression and drainage of the distended, hydropic, or purulent gallbladder. It is particularly applicable if the patient's general condition is such that it precludes prolonged anesthesia, since the operation may be performed under local anesthesia. It is also performed in cases in which marked inflammatory reaction obscures the anatomic relation of critical structures. Cholecystostomy may be a definitive procedure, particularly if a postoperative tube cholangiogram is normal.

Technique

(Fig. [29-26](#)). A circumferential purse-string suture is placed in the fundus of the gallbladder, and a small incision is made through the serosa within the suture. A trocar is inserted into the lumen of the gallbladder, which is then decompressed. After the gallbladder has been emptied, a stone forceps may be introduced to the junction of the ampulla and cystic duct, and obstructing calculi may be removed. A mushroom or Foley catheter is inserted into the lumen of the gallbladder, and a second purse-string suture is placed concentrically to the first one. The sutures are tied, inverting the serosa. Unless a small, oblique incision was used initially, the drainage tube should be brought out through a stab wound.

If the fundus of the gallbladder is necrotic, the gangrenous portion should be excised and the remainder of the gallbladder closed around the catheter, using purse-string sutures.

Open Cholecystectomy

A principal aim of open cholecystectomy is to avoid injury to the common duct close to its junction with the common bile duct to obviate a long cystic duct remnant. A more conservative approach toward elective cholecystectomy is indicated for cirrhotic patients. If an operation is performed, increased bleeding should be anticipated; extensive intrahepatic dissection should be avoided. Intraoperative infusion of vasopressin and an antifibrinolytic agent should be considered.

Technique

(Fig. 29-27). The gallbladder may be approached through an oblique right upper quadrant incision (Kocher or Courvoisier), through a vertical right paramedian incision, or through the upper midline. There are frequently adhesions between the gallbladder, particularly the ampulla, and the duodenum and colon. These should be lysed by sharp dissection. By applying traction laterally to the ampulla and retracting the duodenum medially, the veil of peritoneum running from ampulla to hepatoduodenal ligament may be accentuated and incised. The cystic duct is identified and a silk ligature passed around it. Traction is applied to the ligature to prevent passage of a stone down the cystic duct during dissection of the gallbladder. Dissection is continued craniad in this peritoneal fold, and the cystic artery is identified. The course of this artery to the gallbladder should be demonstrated to avoid ligating the right hepatic artery. The cystic artery should be doubly ligated and transected. If bleeding occurs from the cystic artery, it is best controlled by applying pressure on the hepatic artery within the hepatoduodenal ligament. The artery is compressed between the index finger, which is inserted into the foramen of Winslow, and the thumb anteriorly. The peritoneum overlying the gallbladder is then incised close to the liver, and dissection is begun from the fundus of the gallbladder down to an ultimate pedicle of cystic duct. During this dissection, blood vessels coursing from the liver may require ligation, and the gallbladder bed should be inspected for large draining ducts, which should also be ligated. Attention is then directed toward visualization of the junction of the cystic duct and the common duct. The cystic duct is transected and ligated 3 to 5 mm from the common bile duct. It is not necessary to close the bed of the gallbladder. A drain may be brought out from the hepatorenal pouch, which is the most dependent portion of the upper abdomen with the patient in the supine position, via a separate stab wound if there is any concern that blood will accumulate or if there is marked pericholecystic inflammation and edema. Several series have shown that in the absence of specific indications, drainage is not required.

This method is directed at facilitating demonstration of the junction between the cystic duct and the common bile duct. The gallbladder may also be removed in the so-called retrograde fashion, in which the cystic duct is ligated close to the junction with the common duct as the initial part of the procedure. Then, after the cystic duct and artery have been transected, dissection is begun from the cystic duct and continued outward toward the fundus (Fig. 29-28).

Laparoscopic Cholecystectomy

The application of minimally invasive surgical techniques to removal of the gallbladder has emerged as the preferred way of treating symptomatic gallstone disease. Although a subcostal incision is avoided, these operations should be viewed with the same respect for surgical principles as are the open procedures. Trocars are introduced after the instillation of a pneumoperitoneum, and the gallbladder and liver can be retracted so as to provide optimal visualization. It is essential that the ductal structures be carefully identified before division. Because of the nature of this procedure, bleeding is a particularly worrisome problem and should be guarded against. This procedure begins by retracting the gallbladder up over the edge of the liver so as to facilitate exposure of the triangle of Calot (Fig. 29-29). The cystic duct and artery are then identified, and ductal anatomy is confirmed. Operative cholangiography can be used to selectively define anatomy as well as to search for common bile duct stones. The cystic duct is then divided. The gallbladder is dissected from this area up toward the fundus. Before complete separation of the gallbladder from the liver bed, the gallbladder fossa should be carefully inspected for bleeding. After this has been managed, the gallbladder is carefully withdrawn through one of the ports and the pneumoperitoneum released. As with open cholecystectomy, bile duct injury and hemorrhage can and should be avoided.

Operations of the Extrahepatic Bile Duct

Exploration for Choledocholithiasis

(Fig. 29-30). Exploration for choledocholithiasis is indicated when ductal stones have been identified by palpation or cholangiography or when the cause of obstruction has not been defined. The procedure is often facilitated by performing a Kocher maneuver (freeing the lateral and posterior attachments of the second portion of the duodenum). Common duct dissection can be facilitated by traction on the cystic duct. After the anterior aspect of the duct has been visualized, aspiration of bile with a fine needle provides confirmation. After a fixation suture has been placed laterally and medially and traction applied, a vertical incision is made between these through the anterior wall. A choledochoscope can be introduced at this time to visualize the lumen and determine whether any stones are present.

Ductal stones can be removed by irrigation, balloon-tipped catheters, scoops, or forceps. These procedures should be applied initially to the distal common duct and subsequently proximally to each of the main hepatic ducts. After the stones have been removed, a No. 3 Bakes dilator is passed into the duodenum and the tip is visualized through the anterior wall. When the duct is clear of stones, a T tube is inserted into the duct. The limbs of the T tube should be short so that the distal limb does not pass through the ampulla, and the proximal limb does not obstruct either of the hepatic ducts. The incision in the duct is closed around the long limb of the tube, and saline is injected to demonstrate the absence of leaks. A completion cholangiogram confirms the absence of stones and the passage of dye into the duodenum. The latter can be facilitated by the injection of glucagon. The T tube should be brought out through a stab wound. A postoperative cholangiogram is performed about 1 week postoperatively, and if absence of stones and clear passage of opaque medium into the duodenum are demonstrated, the tube is removed.

The common duct can be explored and cleared of stones laparoscopically (see Chap. 44).

Transduodenal Sphincteroplasty

(Fig. 29-31). Division of the sphincter of Oddi is occasionally indicated for a stone impacted at the ampulla, a stricture, or a functional disorder. The procedure is also applicable for multiple or recurrent ductal stones. A generous Kocher maneuver should be performed initially, followed by a longitudinal anterior duodenotomy. The passage of a Bakes dilator or catheter down the duct facilitates identification of the sphincter, which should be incised at the 11-o'clock position to avoid damaging the pancreatic duct. A pie-shaped segment is removed from the sphincter and the duodenal and ductal mucosa are coapted with fine absorbable sutures. A T tube is inserted into the common duct, and the duodenotomy is closed either longitudinally or horizontally, with care being taken not to compromise the lumen.

Choledochoduodenostomy

(Fig. 29-32). This procedure is applicable to patients with multiple common duct stones to obviate the use of a T tube. The sine qua non for the performance of this procedure is a dilated common duct. A Kocher maneuver is performed to relieve any tension on the anastomosis. The distal common duct is incised longitudinally, as is the anterior portion of the duodenum, and a one-layer, large-diameter anastomosis is made.

Choledochojejunostomy (Roux-en-Y)

(see Fig. 29-23). Although an occasional transection of the common duct with a sharp instrument can be repaired by end-to-end anastomosis over a T tube, most ductal injuries and strictures are preferably repaired with choledochojejunostomy or, more often, a hepaticojejunostomy, using a 45-cm defunctionalized Roux-en-Y limb of jejunum to obviate reflux of intestinal contents into the biliary tree. A precise mucosa-to-mucosa anastomosis should be performed using the antimesenteric side of the jejunum with interrupted absorbable sutures. In most instances, a stent is not necessary.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 30 - Pancreas

Howard A. Reber

ANATOMY

The pancreas lies almost transversely in the retroperitoneum, behind the stomach, between the duodenum on the right and the spleen on the left. It is arbitrarily divided into a head with its uncinat process, a neck, a body, and a tail. The *head* of the pancreas lies at the level of the second lumbar vertebra near the midline. The posterior surface of the head rests next to the medial border of the right kidney, on the right renal vessels, and the inferior vena cava (Fig. 30-1). The *uncinate process* extends posterior and to the left of the head of the gland, behind the portal vein and the superior mesenteric vessels. The *neck* of the pancreas, which joins the head with the body of the gland, is anterior to the superior mesenteric vessels and portal vein.

The portal vein is formed by the confluence of the superior mesenteric and splenic veins. There are usually no tributaries between the anterior surface of the superior mesenteric and portal veins and the posterior surface of the neck. Thus, during resection of the head of the pancreas, the two can be separated safely by blunt dissection, and the neck may be transected.

The *body* of the pancreas lies at the level of the first lumbar vertebra and begins at the left border of the superior mesenteric vein. The posterior surface is in contact with the aorta, the left adrenal gland and kidney, the left renal vessels, and the splenic artery and vein, which run along its superior border. The *tail* of the pancreas lies at the level of the twelfth thoracic vertebra, and its tip usually reaches the hilus of the spleen.

Bile and Pancreatic Ducts

The common bile duct passes posterior to the head of the pancreas on its way to the duodenum and is partially or completely covered by the pancreas in over 70 percent of cases. The duct enters the duodenal wall and runs within it for about 1.5 cm. The main pancreatic duct (*Wirsung*) arises in the tail of the pancreas and enters the duodenal wall caudal to the bile duct. The two ducts usually lie side by side for several millimeters before they join as a common channel. The single pancreaticobiliary duct enters the duodenal lumen on the papilla of Vater, located on the posteromedial wall of the second portion of the duodenum, at the level of the second or third lumbar vertebra. The *ampulla of Vater* is a dilatation of the common pancreaticobiliary channel within the papilla and is distal to the junction of the two ducts. An ampulla is present in about 90 percent of cases, and it usually is quite short (5 mm or less). In 10 percent of people, the individual ducts empty separately into the duodenum, and there is no ampulla. When a common channel exists, it may be important in the pathogenesis of gallstone pancreatitis because a calculus could lodge in the ampulla and obstruct the ducts.

The accessory pancreatic duct (*Santorini*) usually drains the anterior and superior portions of the head of the pancreas. In 60 percent of cases, it enters the duodenum about 2 cm cranial and slightly anterior to the papilla of Vater, through the minor papilla. Because the accessory duct often communicates with the duct of Wirsung, drainage can occur through either the minor or major papilla. The minor papilla is always found distal to the point where the gastroduodenal artery crosses posterior to the duodenum. To avoid injury to the accessory pancreatic duct during gastrectomy, dissection of the duodenum should not progress past the artery.

Numerous variations of ductal anatomy occur, but only *pancreas divisum*, which occurs in up to 10 percent of normal people, occasionally may cause disease. Pancreas divisum occurs when the embryologic dorsal and ventral pancreatic ducts do not fuse, and the main pancreatic duct empties through the minor papilla (Fig. 30-2). In this case, only the duct draining the uncinat process empties through the papilla of Vater. In some people with pancreas divisum, pancreatitis may occur as a result of relative obstruction to the outflow of pancreatic juice through a minor papilla not normally required to handle such a large volume.

In adults, the main pancreatic duct is about 3 to 4 mm in diameter in the head of the pancreas, 2 to 3 mm in the body, and 1 to 2 mm in the tail. Some dilatation occurs normally with aging, and a diameter of 5 to 6 mm may be normal in patients age 70. Nevertheless, ductal dilatation is more commonly an indication of disease and implies ductal obstruction.

Blood Supply (Fig. 30-3)

The blood supply to the head of the gland comes from the superior pancreaticoduodenal *artery*, which arises from the gastroduodenal artery and divides into anterior and posterior branches. These form collaterals with branches of the inferior pancreaticoduodenal artery, which arise from the superior mesenteric artery. The dorsal pancreatic artery usually arises from the proximal 2 cm of the splenic artery and, after supplying some branches to the head, passes to the left to supply the body and tail of the gland. There it is called the transverse pancreatic artery. Numerous branches from the splenic artery anastomose with the transverse artery and also supply the body and tail. Variations of the arterial supply are common, although most are not important to the surgeon. In the most common significant variation (25 percent of cases), the right hepatic artery originates from the superior mesenteric instead of from the common hepatic artery. The right hepatic artery may then pass through the substance of the pancreas, usually the uncinat process, where it can be injured during pancreatic resection. It continues posterior to the portal vein and often along the right side of the common bile duct. Some surgeons perform preoperative arteriography before pancreatic resection to forewarn themselves of such anomalies.

The head of the pancreas is drained by *veins* that parallel the arteries. The superior pancreaticoduodenal, right gastroepiploic, and a colic vein join to form a major gastrocolic trunk on the anterior surface of the head. This trunk empties into the superior mesenteric vein just before it passes under the neck of the pancreas and may be a useful anatomic landmark to identify the vessel during pancreatic surgery. The trunk may be ligated without consequence when a pancreatectomy is performed. Another fairly constant unnamed vein drains the posterior portion of the head of the pancreas and empties into the anterior aspect or right side of the portal vein at the superior margin of the gland. Additional delicate veins also empty into the right side of the superior mesenteric and portal veins directly from the pancreas. These can only be seen during resection of the head of the gland and must be ligated carefully. Venous drainage of the body and tail of the pancreas is directly to the splenic vein and through the inferior pancreatic vein to the inferior or superior mesenteric veins.

Lymphatics

Lymphatic drainage of the pancreas is rich and, in general, follows venous drainage in all directions. The *superior* nodes, located along the superior border of the pancreas, collect lymph from the anterior and superior upper half of the gland. The *inferior* nodes, along the inferior margin of the head and body, drain the anterior and posterior lower half. *Anterior* nodes drain the anterior surface of the head of the pancreas. They are located beneath the pylorus, anteriorly in the groove between the pancreas and duodenum, and the root of the mesentery of the transverse colon. *Posterior* nodes drain the posterior surface of the head. They are found posteriorly in the groove between the pancreas and duodenum, along the common bile duct, the aorta as high as the origin of the celiac axis artery, and at the origin of the superior mesenteric artery. *Splenic* nodes drain the tail of the pancreas. Lymphatic drainage is important in regard to the spread of pancreatic cancer, which arises most commonly in the head of the gland. Most patients with cancer already have nodal metastases at the time of diagnosis, usually involving several of the groups mentioned, except for the splenic. Any curative procedure must eradicate the involved nodes as well as the primary tumor.

Nerves

The pancreas receives sympathetic fibers by way of the splanchnic nerves and parasympathetic innervation by way of the vagus nerves (celiac division of the posterior vagus trunk). In general, the nerves follow the blood vessels and pancreatic ducts as they travel to the pancreatic acini. The splanchnic nerves also carry visceral afferent pain fibers that pass through the celiac plexus and ganglia. It is not known whether afferent fibers of the vagus are involved in pancreatic pain. Because pancreatic cancer and chronic pancreatitis are often accompanied by significant pain, efforts to relieve it sometimes include destruction of the celiac ganglia, with variable success.

PHYSIOLOGY

Exocrine Function

The pancreas secretes 1 to 2 L/day of a clear, watery, alkaline (pH 8.0-8.3) liquid that contains more than 20 different digestive enzymes. This electrolyte-rich juice is isosmotic with plasma and is secreted chiefly by the centroacinar cells and those cells which make up the walls of the smaller pancreatic ducts in response to secretin. The principal cations of this juice are sodium and potassium, which are always present in concentrations similar to those found in plasma (the sum of the two is approximately 165 mmol/L). The concentrations of the principal anions (bicarbonate and chloride) vary. When the stimulus to secrete is minimal, the chloride concentration is high (e.g., 110 mmol/L) and that of bicarbonate low (e.g., 50 mmol/L). When the secretory stimulus is maximal, the bicarbonate concentration rises (to approximately 140 mmol/L), and the chloride concentration falls (to ~20 mmol/L). The secretion of bicarbonate at such high concentrations requires active transport. In part, the reciprocal relationship of the two anions is achieved by the passive exchange of intraductal bicarbonate for interstitial chloride. This exchange occurs as the juice flows through the larger pancreatic ducts on its way to the duodenum. At slower flow rates, there is more opportunity for exchange to take place, and more bicarbonate is lost. The alkaline pancreatic juice helps to neutralize gastric acid in the duodenum and provides the optimal pH for the activity of the pancreatic digestive enzymes.

The digestive enzymes are synthesized and stored in the pancreatic acinar cells and are released in response to *cholecystokinin (CCK)* and vagal cholinergic stimulation. Pancreatic enzymes are proteolytic (e.g., trypsin, chymotrypsin, carboxypeptidase, ribonuclease, deoxyribonuclease, and elastase), lipolytic (lipase, colipase, and phospholipase A₂), and amylolytic (amylase). Although lipase and amylase are secreted in their active forms, the proteolytic enzymes and phospholipase A₂ are secreted as inactive “zymogens.” Activation of trypsinogen to trypsin occurs when the zymogen is exposed to the duodenal enzyme enterokinase. Trypsin then converts the other zymogens to their active forms. In the intestine, the proteolytic enzymes digest proteins into peptides, lipase breaks fats into glycerol and fatty acids, phospholipase A₂ catalyzes the conversion of biliary lecithin to lysolecithin, and amylase converts starch to disaccharides and dextrins.

Enzyme synthesis occurs on polysomes attached to the rough endoplasmic reticulum, following which the newly formed enzymes migrate through various subcellular compartments and end up as zymogen granules, which are stored. Lysosomal enzymes, which degrade unneeded intracellular constituents, are also synthesized by a similar process. They are segregated from the digestive enzymes and stored separately in other subcellular structures called *lysosomes*. Secretory stimulation results in fusion of the zymogen granule membrane with the apical (luminal) membrane of the acinar cell, and the enzymes are discharged into the duct. There is experimental evidence that the lysosomal enzymes (e.g., cathepsin B) could activate trypsinogen inside the cell if the two ever came into contact. Thus segregation of the lysosomal enzymes from the inactive digestive enzymes may be important to prevent the cellular injury that might follow enzyme activation. Such abnormal intracellular enzyme activation seems to be the cause of pancreatitis in several experimental animal models, although it is unknown whether this can occur in human beings.

Regulation of Secretion

Pancreatic secretion is under complex neurohormonal control. Ductal fluid and bicarbonate are secreted chiefly in response to the hormone secretin, which is released from the mucosa of the duodenum and proximal small bowel in response to acid in the bowel lumen. Bicarbonate secretion is triggered when the pH of the duodenal lumen falls below 4.5.

Acinar enzyme secretion is stimulated both by vagal cholinergic discharge and by the hormone CCK. CCK is released from the proximal small bowel by fatty acids, oligopeptides, and some amino acids.

Pancreatic secretion has been divided artificially into cephalic, gastric, and intestinal phases. The *cephalic* phase begins with the sight, smell, and anticipation of food. Although mostly enzymes are secreted, some bicarbonate is also produced. Enzymes are released by vagal cholinergic pathways; *vasoactive intestinal polypeptide (VIP)* , which is structurally related to secretin, may cause the bicarbonate response. The *gastric* phase begins when food enters the stomach. Gastric distention, as well as the exposure of the stomach mucosa to nutrients (e.g., peptides), causes primarily an enzyme-rich pancreatic secretion. This effect is blocked by atropine or truncal vagotomy and probably involves a gastropancreatic vagovagal reflex. The *intestinal* phase begins when the chyme empties into the duodenum. This phase is important for the maintenance of enzyme output as gastric distention diminishes. Gastric acid is also responsible for pancreatic bicarbonate secretion. When the buffering effect of the food is lost, the duodenal pH begins to fall, and secretin is released. The mechanisms responsible for this phase are both neural and hormonal, involving both stimulatory (e.g., secretin, CCK, gastrin-releasing peptide [GRP], bombesin, neurotensin, VIP) and inhibitory (e.g., *pancreatic polypeptide [PP]* , calcitonin-gene-related peptide [CGRP], peptide YY [PYY], neuropeptide Y [NPY], somatostatin) effects on the pancreas. In addition to these inhibitory effects, there is also evidence for a negative feedback mechanism for pancreatic secretion. The presence of pancreatic proteases and bile within the duodenum appears to inhibit the release of CCK. Thus pancreaticobiliary diversion from the intestine is characterized by elevated plasma CCK levels and stimulation of pancreatic secretion. This may have some clinical relevance in patients with chronic pancreatitis who are sometimes treated with large amounts of oral pancreatic enzymes in an effort to inhibit pancreatic secretion.

The quantity of pancreatic enzymes that is secreted in response to a meal is almost 90 percent in excess of what is actually needed for normal digestion. Thus patients develop symptoms of malabsorption only when secretion falls to about 10 percent or less of normal. Pancreatic insufficiency can result from blockage of the main pancreatic duct, which prevents the enzymes from entering the intestine (e.g., pancreatic cancer), from destruction of the pancreatic secretory parenchyma (e.g., chronic pancreatitis), and from surgical procedures that remove the pancreas or result in poor mixing of the gastric chyme with the pancreatic secretions (e.g., partial gastrectomy with gastrojejunostomy). Pancreatic insufficiency affects fat absorption much more than protein or carbohydrate absorption because protein digestion is aided by gastric pepsin and small bowel brush border enzymes and carbohydrate digestion by salivary amylase. Moreover, fat malabsorption causes troublesome diarrhea, while malabsorption of protein and carbohydrate is generally asymptomatic. On a diet containing 100 g of fat a day, normal subjects excrete 5 to 7 g/day of fat in the feces (i.e., 5 to 7 percent fat malabsorption is normal). A total pancreatectomy causes 70 percent fat malabsorption; partial pancreatic resections may not cause any malabsorption, providing that the pancreatic remnant is normal. Vitamin malabsorption is rarely a problem, since water-soluble B vitamins are absorbed throughout the small intestine, and fat-soluble vitamins do not require pancreatic enzymes for absorption. Vitamin B₁₂malabsorption occurs in some patients but is rarely a significant problem. Therefore, vitamin B₁₂ replacement is unnecessary.

Because complete correction of malabsorption is not possible, the goal of treatment of pancreatic exocrine insufficiency is to eliminate the diarrhea that the steatorrhea provokes. At least 5 to 10 percent of the normal amount of enzymes must be delivered to the duodenum at the same time as the gastric chyme in order for this goal to be achieved. This represents about 30,000 units of lipase for each meal, which can be given either throughout the meal or as a single dose at the beginning. A number of different enzyme preparations are available; the least expensive and most effective one should be chosen. Because lipase is inactivated below pH 4.0 and gastroduodenal pH may fall below this level, enzyme therapy may be ineffective if gastric acidity is high.

Patients should be instructed to begin a diet consisting of 4 meals per day and a minimum of 3000 kcal/day. The diet should contain at least 400 g of carbohydrate, 100 to 150 g of protein, and moderate amounts of fat (25 g/meal). If the enzymes eliminate the diarrhea, the fat in the diet can be liberalized until symptoms recur and then maintained just below that level. If the enzymes do not eliminate the symptoms, fat should be restricted further to 50 to 75 g/day. If the diarrhea still persists, H₂ blockers should be added to decrease gastric acidity. If this is still not effective, the H₂ blockers should be stopped, and an enteric-coated enzyme preparation may be tried. In this form, these enzymes are coated with a protective polymer that resists acid and dissolves only above pH 5.0. Thus lipase is more likely to be delivered to the duodenum in its active form.

Endocrine Function

The islets of Langerhans constitute 1 to 2 percent of the pancreatic mass, but they receive about 20 percent of the total pancreatic blood flow. They are dispersed throughout the pancreatic parenchyma, but some of the hormone-secreting cells are concentrated in specific areas. The insulin- secreting cells are evenly distributed and are usually localized in the core of each islet, where they form 60 to 80 percent of the cells. They are surrounded peripherally by a mantle of other cells that are specialized to secrete glucagon (15 to 20 percent of the islet cells), somatostatin (5 to 10 percent), or pancreatic polypeptide (15 to 20 percent). These last cells are found chiefly in the islets of the head of the pancreas. The glucagon cells occur chiefly in the body and tail of the gland.

Insulin

Insulin is a 51-amino-acid peptide (MW 5734 Da) that consists of two polypeptide chains. It is synthesized by the beta cells by means of a precursor, proinsulin. Insulin lowers blood glucose concentration by enhancing glucose uptake by all the body cells, promoting glycogenesis and inhibiting gluconeogenesis. It also stimulates lipogenesis, inhibits lipolysis, and enhances protein synthesis. Although the factors that govern insulin secretion are multiple and complex, the most important is an elevation of the extracellular blood glucose concentration. Other substrates, including amino acids (e.g., arginine) and fatty acids, also stimulate insulin release. The primary sites of action of the hormone are the liver, muscle, and fat cells.

Glucagon

Glucagon is a single-chain 29-amino-acid peptide (MW 3483 Da) formed in the alpha cells of the islets. It increases blood glucose by causing hepatic glycogenolysis and gluconeogenesis. Thus the effects of glucagon counterbalance those of insulin. Glucagon also relaxes and dilates smooth muscle such as the stomach, duodenum, and sphincter of Oddi. The hormone is released by a low blood glucose concentration, amino acids, catecholamines (e.g., during stress), sympathetic nervous discharge, and CCK. Insulin and hyperglycemia suppress glucagon release.

Somatostatin

Somatostatin is a tetradecapeptide (MW 1637 Da) arising in the delta cells of the islets. Both oral and intravenous nutrients stimulate somatostatin secretion, and both neural and hormonal mechanisms may be involved. The hormone inhibits the release of insulin, and in turn, its own release is inhibited by insulin. It may act as an important local paracrine regulator of glucose homeostasis. Somatostatin has a wide spectrum of activities, most of which are inhibitory. These include suppression of the release of gastrin, secretin, VIP, PP, gastric acid, pepsin, pancreatic enzymes, and glucagon. Somatostatin also inhibits intestinal, biliary, and gastric motility.

Pancreatic Polypeptide

Pancreatic polypeptide (PP) is a 36-amino-acid peptide (MW 4200 Da) arising in the PP cells. Orally ingested protein, vagal cholinergic stimulation, and hypoglycemia are all potent stimulants for the release of PP. PP itself inhibits pancreatic exocrine secretion.

The endocrine pancreas influences the synthesis and secretion of pancreatic enzymes from the exocrine parenchyma. At least in part, this is the result of an *islet-acinar portal system* in which the blood that first perfuses the islets collects into a vascular network that surrounds the acini before draining out of the gland. Thus the acini are exposed to higher concentrations of the islet hormones than are the peripheral tissues of the body. This may be important clinically. For example, insulin activates amylase gene expression and amylase synthesis in acinar cells and regulates their sensitivity to CCK. This could explain why about 40 percent of patients with type I diabetes have some impairment in exocrine pancreatic function. Up to 70 percent of patients with chronic pancreatitis also develop diabetes.

CONGENITAL ANOMALIES

Annular Pancreas

Annular pancreas is formed by a thin band of normal pancreatic tissue that completely encircles the second portion of the duodenum and is continuous with the head of the pancreas anteriorly and posteriorly. It is probably the result of abnormal rotation and fusion of the ventral pancreatic primordium during the second month of fetal life. Forty percent of infants with annular pancreas have associated duodenal stenosis or atresia. Symptoms of duodenal obstruction (gastric distention and vomiting) occur in about one- third of cases in the first week of life and about one-half of cases in the first year. The rest are asymptomatic until adulthood, when abdominal pain, nausea, and vomiting may occur. Some cases, however, probably never cause problems and remain undiagnosed. There is no gender predilection in infancy, although adult patients are more often males. The diagnosis may be suspected when an upper gastrointestinal series shows obstruction of the second portion of the duodenum, but it can only be confirmed by direct inspection at surgery. In adults, duodenal narrowing from chronic ulcer disease also must be considered in the differential diagnosis. The treatment is duodenojejunostomy or gastrojejunostomy, either of which bypasses the obstruction. The obstructing pancreatic ring must not be simply divided because this may not relieve the obstruction, and pancreatic fistula can result.

Pancreatic Cysts

Congenital cysts of the pancreas are rare but are increasingly likely to be diagnosed at any age on abdominal CT scans for unrelated disease. If they are single, they may require surgery to distinguish them from a pseudocyst or a cystic neoplasm. Congenital cysts have an epithelial lining; pseudocysts do not. The diagnosis of benign polycystic disease of the pancreas may be possible from the CT image of multiple cysts of various sizes involving the entire pancreas, especially when cysts are also present in other organs (e.g., kidneys, liver, cerebellum, as in von Hippel-Lindau disease). In the absence of symptoms or serious concern that the cysts are neoplastic, surgery may be unnecessary.

Ectopic Pancreas

Ectopic pancreas usually occurs in the stomach or duodenum but has been found in Meckel's diverticulum, the colon, the appendix, and elsewhere. The tissue usually functions and contains islets in about one-third of cases. The condition is usually asymptomatic. However, if symptoms occur, they may simulate duodenal ulcer, appendicitis, intestinal obstruction, etc., depending on the site. Treatment is surgical, and the diagnosis may be suspected from the characteristic gross appearance. The lesions are submucosal, firm, yellowish nodules with a visible central umbilication.

PANCREATITIS

Acute pancreatitis is a nonbacterial inflammation of the pancreas caused by the activation, interstitial liberation, and the digestion of the gland by its own enzymes. It is characterized clinically by acute abdominal pain, elevated concentrations of pancreatic enzymes in blood, and an increase in the amount of pancreatic enzymes excreted in urine. There may be just a single episode of pancreatitis, or several episodes may recur. In the United States, the most common cause of acute pancreatitis is cholelithiasis.

During a mild attack, the morphologic changes are characterized by pancreatic and peripancreatic edema and fat necrosis, but pancreatic necrosis is absent. This is often referred to as *edematous pancreatitis*. The mild form may turn into a severe form (or the episode may be severe from the beginning). In its severe form, extensive pancreatic and

peripancreatic fat necrosis, pancreatic parenchymal necrosis, and hemorrhage into and around the pancreas are evident. This form of the disease is often referred to as *necrotizing pancreatitis*. During such an episode of acute inflammation, both the exocrine and endocrine functions of the gland are impaired for some weeks or even months. If the cause (e.g., cholelithiasis) and any complications (e.g., a pseudocyst) of the pancreatitis are eliminated, the pancreas usually returns to normal. Some scarring may persist after a severe attack, but both exocrine and endocrine functions are normal, and there is little, if any, histologic abnormality evident. Even multiple attacks of acute pancreatitis rarely lead to chronic pancreatitis.

Chronic pancreatitis is characterized clinically by recurrent, acute episodes of abdominal pain indistinguishable from those of acute pancreatitis. For several years or more, patients may be asymptomatic in between episodes that typically occur two or three times a year. In most cases the pain-free intervals shorten, and pain eventually may become constant. As the disease progresses, pancreatic function becomes impaired. The exocrine insufficiency is manifested as steatorrhea and malabsorption. The endocrine insufficiency causes diabetes, often requiring insulin for its management. An unusual form of chronic pancreatitis occurring in about 5 percent of cases is not associated with pain. The principal clinical manifestation of the disease in these patients is pancreatic functional insufficiency. In the United States, the most common cause of chronic pancreatitis is alcoholism. Morphologically, chronic pancreatitis is characterized by a permanent and usually progressive destruction of pancreatic parenchyma. The acinar cells are destroyed first, replaced by dense fibrous scar tissue. Eventually, the islet cells are also damaged by this sclerotic process. Often the main pancreatic duct is dilated, sometimes with focal areas of narrowing along its length. Intraductal protein plugs and/or calculi are common. Both acute and chronic inflammatory cells are present, as well as edema and focal necrosis. Cysts and pseudocysts occur frequently.

A distinct morphologic form of chronic pancreatitis is *obstructive chronic pancreatitis*. Here the ductal system is moderately dilated proximal to the obstruction, which is most often caused by a tumor or scar from a previous injury. In the obstructed part of the pancreas, there is uniform, diffuse atrophy and fibrosis of the acinar parenchyma. Calculi and intraductal protein plugs are unusual.

In chronic pancreatitis (unlike acute pancreatitis), the morphologic changes in the pancreas are irreversible and often progressive, even if the cause (e.g., alcoholism) is removed. This usually leads to permanent functional impairment as more and more acinar and islet cell mass is destroyed. The only exception is obstructive chronic pancreatitis. Here there does seem to be some improvement both morphologically and functionally when the obstruction is relieved.

The terms *recurrent acute* and *recurrent or relapsing chronic pancreatitis* are obsolescent. They referred to repetitive episodes of pancreatic inflammation in patients with either of the two forms of the disease.

Etiology

Gallstone pancreatitis accounts for about 40 percent of all cases of pancreatitis and about 90 percent of cases of acute pancreatitis in the United States. Eradication of the gallstone disease prevents further attacks of pancreatitis. The etiologic mechanism probably involves the transient obstruction of the pancreatic duct by a gallstone in the common bile duct at the ampulla of Vater. Gallstones are recoverable in the feces of over 90 percent of patients within 10 days of an attack, suggesting that the obstruction is brief and that most such stones pass into the duodenum. The usual age at onset is the middle to late forties, with women more commonly affected than men.

Alcoholic pancreatitis accounts for 40 percent of all cases of pancreatitis and about 75 percent of cases of chronic pancreatitis in the United States. In some cases, an episode of binge drinking may precipitate an attack typical of acute pancreatitis. This usually occurs in a patient who is a chronic alcoholic, in whom the pancreas has probably already been permanently damaged by alcohol even though other evidence of pancreatic disease is absent (e.g., pancreatic exocrine insufficiency, diabetes). In other cases, the presentation is less acute. Patients may present with chronic abdominal and/or back pain, often with steatorrhea and diabetes from pancreatic exocrine and endocrine insufficiency. Pancreatic calcifications are found in at least one-third of patients. Weight loss is common, since food often aggravates the pain and intake is voluntarily restricted. The patients may be addicted to the narcotics necessary for pain relief. The usual age at onset is the midthirties, with men more commonly affected than women.

Alcoholic pancreatitis generally is diagnosed after patients have consumed alcohol for at least 2 years and usually between 6 to 10 years. It probably makes no difference whether the ethanol is consumed as wine, beer, or spirits, but the daily consumption averages 100 to 150 g/day. There is evidence that the diet of alcoholic patients who develop pancreatitis is richer in fat and protein than the diets of other alcoholics who do not. It is estimated that about 10 to 15 percent of individuals who consume such large amounts of alcohol eventually develop pancreatitis; a similar number develop cirrhosis of the liver. The factors that govern individual susceptibility to each are unknown.

Early studies suggested that ethanol caused pancreatitis by inducing spasm of the sphincter of Oddi, thereby creating an obstruction to the outflow of pancreatic juice. Experimentally, secretion in the face of similar obstruction has caused acute pancreatitis, but this cannot be the full explanation. Ethanol is also a cellular metabolic poison, and it has deleterious effects on the synthesis and secretion of digestive enzymes by the pancreatic acinar cells. This causes an increase in the concentration of enzyme protein in pancreatic juice and the eventual precipitation of this protein in the pancreatic ducts. Calcium, normally soluble in pancreatic juice, may then precipitate within the matrix of these protein plugs. This creates multiple points of ductal obstruction randomly scattered throughout the pancreas. Further damage follows continued secretion in the face of this widespread obstruction. Ethanol also increases the permeability of the pancreatic ducts. This allows pancreatic enzymes, normally contained within the ducts, to leak out into the surrounding tissue. If the enzymes are active, damage to the pancreas results. There is recent evidence that acute ethanol ingestion markedly but transiently depresses pancreatic blood flow. Thus cellular damage could be provoked by ischemia. Table [30-1](#) lists the common causes of acute pancreatitis.

Hypercalcemic states, most commonly hyperparathyroidism, may cause acute and chronic pancreatitis. The mechanism is unclear, but hypercalcemia may favor the intraductal precipitation of calcium stones, which are seen in up to half these patients. Calcium also influences the activation of certain pancreatic enzymes and a variety of acinar cell synthetic and secretory events. Acute hypercalcemia increases the permeability of pancreatic ducts, perhaps allowing enzymes to leak from them and damage the tissue.

Hyperlipidemia, especially that associated with elevated chylomicrons and very low density lipoproteins, may be a cause of acute pancreatitis. The mechanism may involve the release by pancreatic lipase of large amounts of toxic fatty acids in the pancreatic capillary circulation. Vascular endothelial damage might cause sludging of red blood cells, stasis, and pancreatic ischemic injury and inflammation. Hyperlipidemia also can occur transiently during an attack of pancreatitis, usually in alcoholics. In such patients, it is not thought to play an etiologic role.

Hyperlipidemia as a cause of pancreatitis should be recognized for two reasons. First, because the elevated serum lipid level interferes with the chemical determination of amylase, the serum amylase level in these patients with acute pancreatitis is often reported as "normal." The appearance of a milky white (lactescent) serum in a patient with acute abdominal pain, however, should alert the clinician to the correct diagnosis. Because the lipid is not excreted in the urine, the finding of a high urinary amylase excretion may still be diagnostic in this setting. Second, dietary or pharmacologic control of the hyperlipidemia minimizes the chances of recurrent episodes of pancreatitis.

Hereditary pancreatitis is transmitted as a mendelian dominant trait. At least one form is due to a mutation in the gene for trypsinogen, which may interfere with the normal protective mechanism by which active trypsin is degraded. Trypsin could then digest the pancreas, producing pancreatitis. The condition is rare, although over 250 patients have been reported. Symptoms typical of acute pancreatitis appear in most patients between the ages of 12 and 14 years. Relentless progression of the process is usual, with recurrent attacks of acute inflammation. Eventually, many patients develop the typical manifestations of chronic pancreatitis, including calcifications (32 percent), diabetes (19 percent), and steatorrhea from exocrine insufficiency (15 percent). As in patients with chronic pancreatitis from other causes, an increased incidence of pancreatic cancer has been noted. In most of the reported cases where *endoscopic retrograde cholangiopancreatography (ERCP)* has been performed, the pancreatic ducts have been dilated.

Protein deficiency in children with severe protein-calorie malnutrition may cause kwashiorkor and a specific type of chronic pancreatitis. Treatment with adequate diet rapidly restores pancreatic function to normal, provided that parenchymal fibrosis is not too extensive. Pancreatic calcification is absent, and episodes of abdominal pain are infrequent. Another variety of chronic pancreatitis, possibly related to malnutrition, occurs only in certain tropical and subtropical countries (e.g., Nigeria). It begins in childhood and in most cases is associated with abdominal pain, pancreatic calcification, steatorrhea, and diabetes. The pancreatic ducts are usually dilated and obstructed. Adequate nutrition does not reverse the process.

Postoperative (iatrogenic) pancreatitis occurs after a variety of procedures in proximity to the pancreas, and the causes are usually obvious. They include direct injury to the gland (e.g., pancreatic biopsy, pancreatic resection) or obstruction of the pancreatic duct (e.g., the placement of a long-arm T-tube through the sphincter of Oddi into the duodenum or forceful dilation of the sphincter of Oddi to a diameter of more than 5 mm during common duct exploration). For this reason, these latter two procedures are no longer done. Pancreatitis also occasionally follows endoscopic sphincterotomy or surgical sphincteroplasty.

Sometimes the operation is distant from the pancreas, and there is no obvious explanation. For example, pancreatitis complicates up to 5 percent of cases of heart surgery that require cardiopulmonary bypass. It has been suggested that impaired pancreatic perfusion may damage the pancreas in this setting.

Pancreatitis can follow Billroth II gastrectomy if the afferent limb of the jejunum becomes obstructed. In this case, pancreatic juice and bile accumulate in the limb and may reflux into the pancreatic duct under pressure.

Drugs are an uncommon but important cause of acute pancreatitis. Those most commonly incriminated include steroids, azathioprine, 6- mercaptopurine, thiazide diuretics, furosemide, sulfonamides, tetracycline, and estrogens (which produce pancreatitis by inducing hypertriglyceridemia).

Duct obstruction of brief duration may cause acute pancreatitis, but when it is persistent over months or years, it is a recognized cause of chronic pancreatitis. Unlike other forms of chronic pancreatitis, some morphologic and functional recovery of the pancreas can occur if the obstruction is removed. Of course, when the process is advanced, significant recovery is not possible. Obstruction occurs in a variety of clinical settings.

Chronic pancreatitis has been reported in patients with severe *strictures* at the point where fusion of the ventral and dorsal pancreatic ducts occurs during fetal life. Strictures also occur in other areas of the main duct and are probably due to scarring after trauma or acute pancreatitis from any cause.

Pancreas divisum has been most commonly associated with recurrent episodes of acute pancreatitis in young adulthood, but chronic pancreatitis may occur when the outflow obstruction is severe and persistent over a period of years. In patients with acute pancreatitis and pancreas divisum, therapy directed toward relief of obstruction at the minor papilla may be effective. However, once chronic pancreatitis is established, surgical treatment should consist of pancreatic resection or drainage.

Chronic *inflammation of the papilla of Vater* with obstruction of the duct of Wirsung is rare. In most cases the inflammation is probably caused by the passage of common duct

stones through the papilla. In others there is no apparent explanation. There are probably some cases of chronic pancreatitis that are secondary to the outflow obstruction caused by this process. Nevertheless, the clinician must be extremely cautious about attributing chronic pancreatitis to this entity. In any case, as with pancreas divisum, treatment of chronic pancreatitis directed toward relief of the papillary obstruction is ineffective.

Miscellaneous additional causes of pancreatitis exist. They include acute pancreatitis that follows scorpion (*Tityus trinitatis*) venom poisoning, various infectious agents (mumps, group B Coxsackie viruses, herpes simplex, mononucleosis), and exposure to anticholinesterase insecticides.

Idiopathic pancreatitis still accounts for 15 to 20 percent of the total, suggesting that a number of other causes are still obscure. About one- third of patients with no obvious cause for their episodes of acute pancreatitis eventually will be found to have cholelithiasis. Thus, to avoid overlooking a treatable condition, an extensive search for the cause always must be made before the designation *idiopathic* is applied.

Pathogenesis

Acute pancreatitis is thought to result from enzymatic digestion of the gland by its own enzymes, although the mechanism(s) by which enzyme activation occurs in humans remains unknown. The pathogenesis of pancreatitis is best discussed within the framework of various theories that attempt to relate experimental observations to known clinical characteristics of the disease.

Obstruction-Secretion

Ligation of the pancreatic duct in animals results in edema of the pancreas that may resemble mild acute edematous pancreatitis in humans. It is self- limited, however, and generally resolves within a week; over several months the acinar tissue is replaced by fibrous scar. Partial duct obstruction, along with stimulation of pancreatic secretion, produces a more severe pancreatic inflammation. These observations may be relevant to biliary pancreatitis, when a gallstone obstructs the pancreatic duct during pancreatic secretion. Acute alcoholic pancreatitis also may have a similar mechanism. Alcohol in the stomach stimulates the secretion of gastric acid, which enters the duodenum and releases secretin. This, in turn, stimulates the pancreas to secrete. Alcohol also causes spasm of the sphincter of Oddi, which would obstruct the flow of pancreatic juice. This theory does not explain the mechanism by which the pancreatic enzymes become activated to begin the autodigestive process.

Common Channel Theory

In two patients who died from acute pancreatitis, Opie (1901) observed at autopsy that a gallstone was lodged in the ampulla of Vater, obstructing the bile and pancreatic ducts and creating a common channel between them. He suggested that bile reflux through the channel into the pancreatic duct was the cause of the pancreatitis. This hypothesis continues to be debated. It is known that infected bile that contains deconjugated bile salts and bile incubated with pancreatic juice that contains lysolecithin injure the lining of the pancreatic ducts. Their permeability is increased so that pancreatic enzymes can leak from them into the surrounding pancreatic parenchyma. This could initiate pancreatitis. On the other hand, fresh bile is apparently innocuous to the ducts. Moreover, in 90 percent of people the common channel may be too short to allow for bile reflux in the presence of an obstructing gallstone.

Duodenal Reflux

Reflux of duodenal contents through the papilla of Vater into the pancreatic duct causes pancreatitis in experimental animals. The rare case of pancreatitis that occurs after a Billroth II gastrectomy with obstruction of the afferent loop probably has a similar causation. This theory is attractive because it explains pancreatic enzyme activation, since duodenal enterokinase would gain access to the pancreatic duct. Nevertheless, normal mechanisms inherent in the duodenal wall and sphincter of Oddi efficiently prevent reflux, and there is no direct evidence that this mechanism accounts for pancreatitis in humans.

Increased Pancreatic Duct Permeability

Animal experiments have shown that a variety of conditions can increase the permeability of the pancreatic ducts so that pancreatic enzymes, normally contained within the ducts, can leak out. These include the acute ingestion of ethanol, direct exposure of the duct to deconjugated bile salts, pancreatic secretion against an obstruction, and acute hypercalcemia. If those enzymes have been activated, they cause acute pancreatitis. Nevertheless, there is no direct evidence that this occurs in human pancreatitis, and it is unknown how enzyme activation might occur.

Enzyme Autoactivation

In experimental animals, intrapancreatic autoactivation of trypsinogen occurs, and pancreatitis develops when pancreatic secretion is stimulated by supramaximal doses of cerulein or when animals are fed a choline- deficient diet supplemented with ethionine. These observations may be important because they offer a novel explanation for enzyme activation that does not require exposure to duodenal enterokinase. However, there is no evidence that enzyme autoactivation occurs in human pancreatitis, and there is no apparent connection between these experimental models and the human disease.

Presently we have only an incomplete understanding of the pathogenesis of pancreatitis. The diverse causes of the disease suggest that any of a number of inciting factors may result in injury to the gland and that the pancreas responds in a limited fashion. Biliary pancreatitis is associated with *acute* inflammation and almost certainly follows transient obstruction of the pancreatic duct by a gallstone. It is unknown whether bile reflux, elevated ductal pressures, or other factors are important. Alcoholic pancreatitis is associated with *chronic* inflammation, although acute inflammatory episodes do occur. The chronic changes are due to the toxic effects of ethanol on the acinar secretory cells and probably to widespread ductal obstruction from protein plugs.

Acute Pancreatitis

Clinical Manifestations

The typical attack of acute pancreatitis begins with severe and persistent epigastric or upper abdominal pain that often radiates through to the back. Frequently it follows the ingestion of a large meal and is associated with nausea and persistent vomiting and retching. The findings are the same regardless of the causation, even if the event represents an episode of acute pancreatic inflammation in a patient with chronic pancreatitis.

The pain is of variable intensity and may be less severe with edematous pancreatitis compared with the necrotizing form of the disease. Examination of the abdomen reveals tenderness most marked in the epigastrium but sometimes present throughout. The bowel sounds are decreased or absent. Usually there are no masses palpable; when one is present, it most often represents a swollen pancreas (phlegmon), pseudocyst, or abscess. With necrotizing pancreatitis, the abdomen may be distended with intraperitoneal fluid. The temperature is only mildly elevated (100 to 101°F) in uncomplicated cases. There may be evidence of a pleural effusion, especially on the left side.

When the disease is more severe, the patient also may exhibit signs of profound fluid losses from sequestration of edema fluid and/or blood in the peripancreatic retroperitoneal spaces or in the peritoneal cavity (ascites). Severe dehydration, tachycardia, and hypotension may be present. In about 1 percent of patients, a bluish color is evident around the umbilicus (Cullen's sign) or in the flanks (Grey Turner's sign). This represents blood that has dissected to those areas from the retroperitoneum near the pancreas in patients with necrotizing pancreatitis.

Although the clinical presentation often suggests the correct diagnosis, and the laboratory findings usually confirm it, it is important to stress that *acute pancreatitis is a diagnosis of exclusion*. Other acute upper abdominal conditions (e.g., perforated peptic ulcer, acute cholecystitis, gangrenous small bowel obstruction) must be considered in every patient. If doubt remains, a laparotomy may be indicated for diagnosis (<5 percent of patients), since the diseases with which acute pancreatitis is most likely to be confused are often lethal if not treated surgically.

Diagnostic Studies

The hematocrit may be elevated because of dehydration or low as a result of pancreatic or retroperitoneal blood loss in necrotizing pancreatitis. In the absence of suppurative complications, white blood cell counts over 12,000/mm³ are unusual. Liver function tests are usually normal, but mild elevations of the serum bilirubin concentration (<2 mg/dL) may be seen. This is probably due to partial obstruction of the intrapancreatic portion of the common bile duct by the swollen head of the pancreas. Ultrasound examination may reveal evidence of pancreatic and peripancreatic edema or fluid collections. CT scans may indicate similar changes. Nevertheless, these latter two studies are not commonly needed to diagnose pancreatitis. Their main value is in the diagnosis and management of complications of the disease.

The most widely used tests to diagnose acute pancreatitis depend on the release of various pancreatic enzymes from the inflamed gland. The serum *amylase* concentration rises to more than 2.5 times normal within 6 h of the onset of an acute episode and usually remains elevated for several days. Values above 1000 I.U./dL are characteristic (but not diagnostic) of biliary pancreatitis; lower values are typical for acute alcoholic pancreatitis. The level of the serum amylase concentration does not correlate with the severity of the episode of pancreatitis.

There are several reasons why a patient with acute pancreatitis may have a normal serum amylase concentration: (1) The urinary clearance of amylase increases shortly after the onset of pancreatic inflammation. Thus the serum concentration may fall to normal because more is excreted in the urine. For this reason, the urinary excretion of amylase is often measured as well. Excretion of more than 5000 I.U./24 h is abnormal. (2) In patients with hyperlipidemia and pancreatitis, the chemical determination of amylase is interfered with by the lipid. Serum amylase values are usually falsely normal in this setting. Urinary amylase measurement is still accurate and is usually elevated, however. (3) In patients with chronic pancreatitis who have suffered significant destruction of pancreatic parenchyma, there may be insufficient amylase released to raise the serum concentration above normal, even with an episode of acute inflammation. Pancreatic fibrosis also may limit the permeability of the gland and its capillaries, so absorption of amylase into the blood is restricted.

It is important to recognize that the serum amylase concentration and the urinary amylase excretion both may be increased in patients without pancreatitis. Examples are patients with perforated duodenal ulcer, gangrenous cholecystitis, small bowel obstruction, and other acute intraabdominal inflammatory conditions. This probably occurs because the pancreatic amylase leaks directly into the peritoneal cavity (e.g., perforated viscus) or translocates through the inflamed and permeable bowel wall and is absorbed into the blood.

The usual laboratory assay measures amylase isoenzymes from both pancreatic (p-type amylase) and salivary or other (s-type amylase) sources. Thus serum amylase elevations in some patients without pancreatic disease may be due to excessive s-type amylase alone. Examples are salivary or ovarian tumors, chronic sialadenitis, and liver disease. Macroamylasemia is another condition in which the serum amylase concentration is elevated but the pancreas is normal. In this abnormality, the amylase molecules (usually s-type) exist in the serum as large macromolecular aggregates whose effective size and molecular mass are much greater than the normal amylase molecule's (~55,000 Da). These macromolecules are too large for glomerular filtration and excretion by the kidney, so their serum concentration increases. Urinary excretion of amylase is low. The identification of both macroamylasemia and s-type amylase can be made by electrophoretic analysis of the serum. This should be done when the serum amylase concentration is high and the clinical picture is inconsistent with the diagnosis of pancreatitis.

Other pancreatic enzymes also have been measured in an effort to improve the diagnostic accuracy of serum amylase determinations. The greatest experience has been with serum *lipase*, which appears to be more reliable than amylase to diagnose acute pancreatitis. This is probably because the pancreas is the main source of lipase in the blood. Nevertheless, the test is still not used widely because the older method for lipase determination was cumbersome and time-consuming. A new technique that is rapid and simple may increase its utility. Radioimmunoassay kits are now available for the rapid determination of serum *trypsin* and *elastase* concentrations, and these also appear to be sensitive markers of pancreatic inflammation. To date, they have been used mostly in research protocols.

Radiologic studies are not used commonly to diagnose acute pancreatitis, but certain nonspecific abnormalities are often present. The most frequent finding on a plain abdominal film is the dilatation of an isolated loop of intestine (duodenum, jejunum, or transverse colon) adjacent to the pancreas (*sentinel loop*). A gas-filled ascending and right transverse colon that stops abruptly in the middle or left transverse colon (*colon cutoff sign*) is due to colonic spasm near the inflamed pancreas. An upper gastrointestinal series may show a widened duodenal loop or a swollen ampulla of Vater because of edema in the head of the pancreas. Calcification in the pancreas itself signifies the presence of chronic pancreatitis but does not prove the presence of acute inflammation. Chest films may reveal pleural effusions, more common on the left side.

In patients with biliary pancreatitis, a plain abdominal film may show radiopaque gallstones in the region of the gallbladder. However, the best way to confirm the presence of gallstones when biliary pancreatitis is suspected is with *ultrasound* (US) examination. Ultrasound also may reveal an edematous, swollen pancreas, peripancreatic fluid collections, or pseudocysts. Serial US examinations may be useful to follow the clinical course of patients with protracted or complicated pancreatitis or pseudocysts. In about 20 percent of cases of acute pancreatitis, US examination of the pancreas is technically unsatisfactory because of the presence of overlying bowel gas.

A *CT scan* of the pancreas should be obtained in all patients with acute pancreatitis whose illness has not begun to resolve within 2 or 3 days or whenever complications are suspected. Possible findings include a normal pancreas, pancreatic edema, a mass of inflamed peripancreatic tissue and pancreas (phlegmon), pancreatic and/or peripancreatic fluid collections (acute pseudocysts), or abscess. Extension of the inflammation from the area of the pancreas to other tissue planes (retrocolic, perinephric) may be evident and implies a more serious prognosis. The adequacy of pancreatic perfusion also can be estimated if the CT is done first without and then with intravenous contrast material, which must be injected rapidly as a bolus. A viable pancreas "enhances" as the contrast material flows through it; lack of enhancement suggests necrosis of that part of the pancreas (Fig. [30-4](#)). This technique represents a major advance in the diagnosis and management of complicated acute pancreatitis, since patients with significant pancreatic necrosis (>50 percent of the pancreas) are more likely to develop pancreatic infection and to require surgical intervention. If infection is suspected, a percutaneous ("skinny") needle can be introduced into collections of fluid, especially near poorly perfused pancreatic tissue, using CT or US as a guide. The aspirated material should be sent for Gram stain and bacterial and fungal culture. Proof of infection is an indication for laparotomy and surgical drainage.

Prognostic Signs

Early identification of patients who are at increased risk for the development of serious complications of pancreatitis allows them to be managed more aggressively, which may decrease the mortality rate. Three different approaches have been used: (1) Specific biochemical markers have been sought that discriminate between edematous and necrotizing pancreatitis, since the latter is known to have a higher rate of complications. For example, in necrotizing pancreatitis, the serum level of a γ -macroglobulin (which binds trypsin) falls, and the levels of a α -protease inhibitor and complement (C3 and C4) increase. So far the best discrimination between edematous and necrotizing pancreatitis has been achieved with serum C-reactive protein, a nonspecific acute-phase reactant that is elevated in the latter. C-reactive protein is still not widely used outside the research setting, however. (2) Multiple biochemical and/or physiologic measurements have been made during the first several days after hospital admission, which, taken together, have prognostic value. Ranson developed a series of different criteria for the severity of acute pancreatitis (Table [30-2](#)). Modifications of these criteria exist, but Ranson's original ones are still used most widely. The APACHE II scoring system also has been applied recently to patients with acute pancreatitis and appears to have prognostic significance as well. (3) Contrast-enhanced CT scan provides the most objective information about the extent of the inflammatory process, the degree of pancreatic necrosis, and the presence of infection, all of which affect the treatment and the prognosis. Because it is an invasive and expensive test, however, it is not done on all patients who are admitted with the disease. It should be done on patients who appear to be at higher risk based on any of the preceding prognostic criteria (e.g., >3 Ranson signs), those whose illness is not resolving within 2 or 3 days of admission, or for any other reason if a complication is suspected.

Treatment

Medical Treatment

The treatment of uncomplicated acute pancreatitis is medical and is directed primarily toward the restoration of fluid and electrolyte balance and the avoidance of secretory stimulation of the pancreas. *Nasogastric suction* should be instituted in most patients, except those with mild disease unassociated with significant vomiting or pain. All oral intake should be withheld until the ileus has resolved and pain is absent. The enzyme abnormalities usually have returned to normal by that time. For unclear reasons, the serum enzymes sometimes remain elevated even though the patient is asymptomatic. The resumption of oral intake is usually tolerated in these circumstances. Occasionally, symptoms will recur when oral intake is resumed. When this occurs, another period of fasting is indicated.

Fluid requirements may be considerable, since large volumes can be sequestered in the retroperitoneum adjacent to the pancreas. The amount of crystalloid and colloid (albumin, blood) given should be sufficient to maintain an adequate hematocrit, circulating blood volume, and urine output. It should be stressed that the most important aspect of early resuscitation is adequate and aggressive fluid replacement. Renal failure from inadequate replacement is a frequent finding in patients who die from this disease. Moreover, the pancreatitis itself may progress from the impaired pancreatic perfusion that can accompany the shock state. Electrolyte derangements may be present if vomiting has been significant and dehydration is severe. A hypokalemic, hypochloremic, metabolic alkalosis is most common. Hypocalcemia may occur in patients with severe pancreatitis. It must be treated urgently with parental calcium because it predisposes to cardiac arrhythmias. The prognosis appears to be related to the degree of hypocalcemia and the ease with which it is corrected. Hypomagnesemia is also common, particularly in alcoholic patients, and the magnesium should be replaced. About 30 percent of patients with acute pancreatitis develop arterial hypoxemia ($P_{aO_2} < 70$ mmHg) and require supplemental O_2 by mask. Because the onset of this is often insidious and may precede any changes on the chest film, it is prudent to determine arterial blood gases every 12 h for the first several days after admission to the hospital in most patients and longer in those who appear to have more serious disease. An occasional patient may develop *adult respiratory distress syndrome (ARDS)* and require endotracheal intubation and mechanical ventilation. Table [30-3](#) lists the changes in cardiovascular function in acute pancreatitis, and Table [30-4](#) lists the changes in respiratory function.

A number of drugs have been shown to exert no beneficial effect in patients with acute pancreatitis. These include antibiotics (which should be used only to treat specific suppurative complications), anticholinergics, gastric acid antisecretory agents, glucagon, somatostatin, and inhibitors of proteolytic enzymes (e.g., Trasylol). Parenteral nutrition has no specific beneficial effect on the course of pancreatitis. It should be used, as in any other patient, when oral intake must be withheld for prolonged periods and enteral nutrition is not possible. Elemental diets introduced by tube into the small intestine do not avoid pancreatic stimulation.

Peritoneal lavage has been used in patients with severe acute pancreatitis and intraperitoneal fluid to remove toxins and various metabolites from the peritoneal cavity and minimize their systemic absorption. Although controlled trials failed to show that this decreased mortality rates, there was some evidence that the incidence of cardiopulmonary complications was lessened. For this reason, and because there are many anecdotal reports where patients rapidly improved during lavage therapy, it is still done. The usual indication is severe pancreatitis that fails to show significant clinical improvement after several days of intensive therapy by more conventional means. The technique requires the insertion of a peritoneal lavage catheter and the rapid infusion and withdrawal of lactated Ringer's solution (1 to 2L/h) for 2 to 3 days or longer. If a response is seen, it is usually during the first 6 to 8 h. If no improvement occurs, surgery may be required.

Biliary pancreatitis is caused by a gallstone that obstructs the pancreatic and bile ducts. In the majority of cases, the gallstone dislodges spontaneously and usually passes into the intestine. When this happens, the patient's clinical course begins to improve, and the episode is generally short-lived. If the stone remains impacted, the attack of pancreatitis may become severe and refractory to the usual treatment. Impaction should be suspected when a patient with pancreatitis is known to have gallstones, when the serum bilirubin concentration is elevated more than 4 mg/dL, when the alkaline phosphatase level is elevated, and when the clinical course does not improve within 24 to 36h. If a skilled endoscopist is available, ERCP with *endoscopic sphincterotomy* to remove the stone may abort the attack. Otherwise, surgical removal of the impacted stone is indicated.

Surgical Treatment

Surgery is contraindicated in uncomplicated acute pancreatitis. Surgery is occasionally performed in patients with abdominal pain in whom the diagnosis is unclear, and the patient is found to have acute pancreatitis. If the pancreatitis is mild to moderate in severity and gallstones are present, a cholecystectomy should be done, and an operative cholangiogram should be obtained. Common duct stones should be removed. The pancreas should not be disturbed. For severe pancreatitis, the biliary tree should be drained if stones are present (cholecystostomy), but cholecystectomy probably should not be done. The common duct should be decompressed with a T-tube if stones are present, and if a stone is impacted at the ampulla of Vater, it should be removed. It is unclear whether drains should be placed to the area of the pancreas; many surgeons would not do this, thinking that the foreign body might predispose to infection. If necrotic pancreas or peripancreatic tissue is found, it should be debrided and drained widely.

In most patients with biliary pancreatitis, the diagnosis is obvious without the need for operation, and the pancreatitis resolves within 2 or 3 days. In these patients, cholecystectomy should be performed during the same hospitalization, usually 4 to 6 days after the original admission. The older practice of discharging the patient and delaying cholecystectomy for 6 weeks was associated with a 60 to 80 percent incidence of recurrence of pancreatitis. On the other hand, patients with severe episodes of biliary pancreatitis should be discharged once they have recovered, and cholecystectomy should be deferred for at least 6 weeks. By this time, the inflammation has subsided and the operation can be done without increased risk.

Complications

Severe episodes of acute pancreatitis may be associated with both systemic and local complications. *Systemic complications* include respiratory insufficiency with ARDS, renal failure, and depressed myocardial function. To some extent, this multiple organ failure may be due to secondary pancreatic infection and the systemic effects of sepsis, as seen in other disorders. Bacterial translocation and absorption of endotoxin from adjacent inflamed colon may be important. There is also evidence that systemic absorption of pancreatic proteases, phospholipase A_2 , and inflammatory cytokines such as interleukin-1, interleukin-6, and tumor necrosis factor, and other noxious agents released from the inflamed pancreas may contribute. This is the rationale for peritoneal lavage: to remove the putative toxins. Inadequate fluid resuscitation accounts for some cases of renal failure. The treatment of multiple organ failure in this setting is no different from that in any other. Resistance to maximal treatment may justify surgical intervention to debride and drain the pancreas, although pancreatic infection is the usual indication.

Local complications include the development of pancreatic or peripancreatic infection and pseudocyst formation. Pseudocysts will be discussed in a separate section. Septic complications are best divided into infection of pancreatic necrosis and pancreatic abscess. *Infected pancreatic necrosis* occurs in at least 40 percent of patients with severe acute pancreatitis and necrosis of more than 50 percent of the pancreas (demonstrated by CT scan). Infection is difficult to recognize clinically. Even with sterile necrosis, patients may appear septic, with high fever, leukocytosis, and abdominal tenderness. Sterile necrosis in a stable patient should be managed expectantly, without surgical intervention. If gas

bubbles (presumed to originate from bacterial fermentation) are seen on the CT image, infection is certain, and surgery should be undertaken (Fig. 30-5). Nevertheless, percutaneous aspiration with Gram stain and culture of the involved material is usually required to prove the presence of infection. Infected pancreatic necrosis is most often diagnosed during the second week of an episode of pancreatitis. The organisms are usually a mixed flora including *Escherichia coli*, *Bacteroides*, *Staphylococcus*, *Candida albicans*, etc., so broad-spectrum antibiotics should be used. Antibiotics, however, are only adjunctive. Surgical drainage is necessary for definitive management. Percutaneous drainage is almost always inadequate, since the material is thick and contains particulate debris that rapidly clogs the narrow lumen of these tubes. At operation, the infected necrotic material should be debrided widely and the involved areas drained to the outside with large sump drains or packed open. Postoperatively, peritoneal lavage may be useful to mechanically remove additional devitalized tissue. Reoperation to drain recurrent or persistent infection is required in at least *one-fifth* of these patients. Associated complications such as fistulas from the bowel or pancreas or bleeding from the retroperitoneal structures are common. The overall mortality rate varies from 10 to 35 percent. Table 30-5 lists the indications for surgery in severe acute pancreatitis.

The decision to operate on a patient with severe acute pancreatitis is often difficult and requires mature clinical judgment. Those indications which are widely accepted include (1) for differential diagnosis when the surgeon is concerned that the symptoms are due to a disease other than pancreatitis for which operation is mandatory, (2) in persistent and severe biliary pancreatitis when an obstructing gallstone that cannot be removed endoscopically is lodged at the ampulla of Vater, (3) in the presence of infected pancreatic necrosis, and (4) to drain a pancreatic abscess if percutaneous drainage is not adequate. Other indications that are less well defined are (1) the presence of sterile pancreatic necrosis involving more than half the pancreas and (2) when the pancreatitis persists or the patient's condition deteriorates despite maximal medical therapy. Some less common indications for operation in these patients (e.g., peritonitis, hemorrhage, biliary obstruction, duodenal obstruction, etc.) have been omitted because the need for operation is generally more *apparent*.

Pancreatic abscesses are localized collections of pus in or near the pancreas and are confined by a capsule. They are usually diagnosed later than infected necrosis, about 3 to 4 weeks after the onset of the attack. Although systemic evidence of toxicity is often less, the patients may appear quite septic. CT scans reveal a fluid collection that may be indistinguishable from a pseudocyst. As with infected necrosis, gas bubbles in the fluid prove infection, but aspiration of the fluid with Gram stain and culture is usually required for the diagnosis. If the fluid is thin in consistency and there is little particulate matter in it, percutaneous drainage may be adequate treatment. Open surgical drainage is indicated if the material is thick, if rapid improvement does not occur after percutaneous drainage, or if there is CT evidence of considerable surrounding tissue necrosis in addition to the abscess. The mortality rate is about 20 percent.

Prognosis

The mortality rate for all patients with acute pancreatitis is approximately 10 percent. Necrotizing pancreatitis associated with infection has a mortality rate of about 35 percent, although there is some evidence that earlier diagnosis of infection and aggressive surgical intervention may lower this figure.

Chronic Pancreatitis

Clinical Manifestations

In 95 percent of patients with chronic pancreatitis, the principal symptom is *abdominal pain*. It is usually located in the epigastrium and is described as cramping, boring, or aching. In more than half the cases, it is also felt in the back. During an episode of acute inflammation, the pain is similar to that felt by patients suffering attacks of acute pancreatitis. In between such episodes, patients have recurrent pain that may last from a few days to several weeks, separated initially by several months or more. Typically the pain-free intervals become shorter and the painful attacks last longer so that pain eventually occurs every day. When the disease is so far advanced that most of the exocrine function is lost, about one-third of patients experience spontaneous pain relief. Another 10 to 20 percent have temporary remissions of pain that last 1 to 3 years. Some pain relief is typically obtained when the patients sit with the knees flexed and a pillow pressed to the abdomen. Most require analgesic drugs and many become addicted to narcotics. Assessment of the severity of pain is often difficult because many of these patients are addicted to alcohol and/or narcotics and have unstable personalities. Consequently, secondary gain may lead to an exaggeration of symptoms. The mechanism of the pain is uncertain, although it may be related to elevated duct and pancreatic tissue pressures and impaired parenchymal blood flow, a kind of compartment syndrome. Table 30-6 lists the potential causes of pain in chronic pancreatitis.

Variable *weight loss* occurs in 75 percent of patients with chronic pancreatitis because food usually aggravates the pain, and thus intake is voluntarily restricted. Significant weight loss from malabsorption is less common. Nevertheless, patients may complain of bulky, offensive, fatty or oily stools.

About two-thirds of patients with chronic pancreatitis have abnormal glucose tolerance, and half of these have *diabetes mellitus*. In these patients some degree of malabsorption is also present, since exocrine and endocrine insufficiency usually develop concurrently. The diabetes is usually easily controlled with insulin, but hypoglycemia may be a problem in alcoholics with irregular eating habits.

As a rule, most patients with chronic pancreatitis have failed to adapt to the stresses of society and have inadequate personalities. The additional stresses created by the disease often result in depression and anxiety. These personality traits are important in evaluating these patients, as well as in choosing specific types of therapy.

In most patients with chronic pancreatitis there are few physical findings. There may be some epigastric tenderness, some evidence of weight loss, and occasionally jaundice. If a mass is palpable, it is most likely a pancreatic pseudocyst.

Diagnostic Studies

These are not especially helpful in the diagnosis of chronic pancreatitis. About two-thirds of patients with alcoholic pancreatitis have abnormal liver function. Mild elevations of serum concentrations of bilirubin (2 to 4 mg/dL), alkaline phosphatase, and serum glutamic-oxaloacetic transaminase (SGOT) and mild depression of serum albumin concentration may occur. Pancreatic exocrine function is impaired in almost all patients, but there is a wide range of normal values for bicarbonate and enzyme secretion and volume flow in response to secretory hormones. Thus mild abnormalities may not be apparent. Pancreatic function tests are rarely indicated in patients with overt pancreatic disease (chronic pancreatitis or pancreatic cancer). Although these tests would confirm the presence of pancreatic insufficiency, they (1) rarely influence therapeutic decisions and (2) cannot distinguish between pancreatitis and cancer. They are of greater value in patients with vague complaints in whom the presence of pancreatic disease is suspected but for which no direct evidence exists. In these patients, these tests are 90 percent sensitive in detecting functional abnormalities. Thus a normal test along with one or two other studies that are also normal (e.g., ERCP or CT scan) reliably rules out pancreatic disease.

In 30 to 50 percent of patients with chronic pancreatitis, plain x-rays of the abdomen reveal pancreatic calcifications (Fig. 30-6). When calcifications are present, the diagnosis of chronic pancreatitis is certain even if there is no clinical evidence. With early chronic pancreatitis, and even in some patients with advanced disease, ductal abnormalities demonstrated by ERCP may be minimal. In the majority of patients with advanced disease, the main duct may be dilated up to 1 cm or more in diameter with intermittent points of obstruction (Fig. 30-7). Strictures, cysts, and ductal calculi may be seen. The common bile duct is also opacified by this technique, and distortion of its anatomy is evident. Chronic pancreatitis characteristically produces angulation of the distal common duct with a smooth, tapered narrowing of its intrapancreatic portion (Fig. 30-8). Generally, all patients with chronic pancreatitis who are being considered for operation should undergo ERCP examination. The study provides critical information about ductal anatomy. Patients with dilated ducts are candidates for drainage procedures; those who do not have ductal dilatation may require pancreatic resection (Fig. 30-9). Occasionally, it is impossible to obtain a preoperative ERCP. In these cases, a pancreatogram can be done at operation by puncturing the duct directly through the anterior surface of the gland.

CT scans also provide useful information. When the duct is grossly dilated, it usually can be seen running the length of the gland (Fig. 30-10). Pseudocysts and cystic communications with the ductal system are evident. Biliary dilatation and the level of bile duct obstruction are defined. CT scans provide the most precise information about the size and configuration of the pancreas, information that is useful in decisions regarding possible pancreatic resection.

Differential Diagnosis

The diagnosis of chronic pancreatitis is usually straightforward. Most patients are chronic alcoholics with recurring episodes of abdominal pain, diabetes, weight loss, pancreatic calcifications, and some evidence of malabsorption. Occasionally, the distinction between chronic pancreatitis and pancreatic cancer may be difficult. This may occur when there is no previous history of pancreatitis, when the patient presents with jaundice and pain is not a prominent part of the picture, or when weight loss is significant. In such circumstances, surgery may be necessary to make the diagnosis.

Complications

The principal complications of chronic pancreatitis are pseudocyst formation, diabetes mellitus, and malnutrition. *Jaundice* is seen sometime during the course of the disease in up to one-third of patients. This is most often the result of transient obstruction of the intrapancreatic portion of the common bile duct during an episode of acute pancreatic inflammation. It subsides when the acute process resolves. As many as 10 percent of patients develop persistent *common duct obstruction* from fibrous constriction of the duct in the head of the gland. The obstruction is almost never complete, which is different from the obstruction caused by pancreatic cancer. *Duodenal or colonic obstruction* occurs in less than 1 percent of patients with chronic pancreatitis and is caused by the inflammation in the adjacent pancreas with subsequent scarring. Gastrojejunostomy is effective in the former case, resection of the involved segment of colon in the latter.

Medical Treatment

The medical treatment of an episode of acute pancreatic inflammation is the same as the treatment of any bout of acute pancreatitis. In patients with *diabetes* from chronic pancreatitis, ketoacidosis is rare, even if blood glucose levels are quite high. Even small amounts of insulin may induce a profound hypoglycemia. This is probably due to concurrent glucagon deficiency. Thus the patient must be instructed thoroughly on the dangers of hypoglycemia, its symptoms, and management. Mild elevations in blood sugar do not require treatment, but fasting elevations in excess of 250 mg/dL should be managed with insulin. Frequently, only 20 to 30 units a day is required. Rigid control of blood sugar is not advisable. Oral agents are almost never successful.

The medical management of *pain* in patients with chronic pancreatitis should include the cessation of all alcohol intake. As many as 50 percent of patients experience some pain relief when they stop drinking. Pancreatic secretion itself may exacerbate the pain. For this reason, patients have been treated with somatostatin analogues, which effectively suppress secretory activity. Efficacy is as yet unproven, but even if the pain is relieved, long-term administration of these drugs is impractical because they can only be given parenterally, they are expensive, and they have significant side effects (e.g., gallstones).

The analgesic management of these patients should begin with small doses of oral medications such as codeine derivatives, either alone or combined with acetaminophen. If more potent narcotic analgesics are required, especially for prolonged periods, serious consideration must be given to surgical pain relief. All too often these patients are denied surgical treatment until after they have become addicted to narcotics and are severely disturbed psychologically as a result of living with constant pain.

Surgical Treatment

The most common indication for surgery is persistent severe pain, although a variety of other complications may require operative intervention (Table 30-7). The indications for surgery for pain relief are impossible to state in precise terms. When the pain interferes substantially with the quality of life, surgery should be considered. Pertinent considerations include obvious effects on health and well-being such as weight loss and nutritional status, need for frequent hospitalization, inability to maintain employment, psychiatric manifestations (usually depression), deterioration of family life, and narcotic addiction.

All patients who are being evaluated for possible surgical intervention should undergo an abdominal CT scan and an ERCP, in addition to the standard preoperative assessment for a major abdominal procedure. Psychiatric evaluation is also appropriate in many patients. Information from these studies helps with decisions about the type of surgical procedure that may be indicated.

There are two categories of operations for the relief of pain in patients with chronic pancreatitis. The first are drainage operations, which attempt to drain more adequately a dilated ductal system that is presumed to be obstructed. The second are the pancreatic resection operations, which remove diseased pancreatic tissue, usually in situations where the pancreatic ducts are normal or narrowed in size. Sphincterotomy or sphincteroplasty of the sphincter of Oddi has no place as the primary treatment of pain in patients with this disease.

Drainage Procedures

When the diameter of the main pancreatic duct in the head and body of the pancreas increases to 7mm or more, a lateral pancreaticojejunostomy (the Puestow procedure) is technically feasible (Fig. 30-11). The duct is opened throughout the length of the pancreas and anastomosed side to side to a Roux-en-Y limb of jejunum. Immediate pain relief is likely in 80 to 90 percent, but pain recurs unpredictably in 25 to 50 percent of patients after 5 years. In some patients, recurrence of pain may be due to stenosis of the pancreaticojejunal anastomosis. This can be proved by ERCP. Reconstruction of the anastomosis may then be beneficial. In the majority of cases, there is no apparent cause for the recurrence of pain, and pancreatic resection must be considered. If the Puestow operation is done in a pancreas with a normal-sized duct, both short- and long-term results are unsatisfactory, perhaps because the anastomosis strictures close in most cases.

The operative mortality of longitudinal pancreaticojejunostomy is about 4 percent, and the morbidity is minimal. Because almost no pancreas is resected, the procedure rarely results in diabetes. Nevertheless, a number of patients eventually will require insulin, since destruction of the pancreas probably continues. Although pancreatic enzymes can now empty freely through the pancreaticojejunostomy, there is rarely any clinical improvement in the degree of malabsorption. However, patients may gain weight because eating no longer produces pain, and therefore, they eat more.

Pancreatic Resections

Pancreatic resection should be considered for pain relief when the pancreatic duct is narrow or normal in diameter, when a previous pancreaticojejunostomy has failed, or when the pathologic changes in the pancreas particularly involve one part of the gland and the rest is less diseased. The standard Whipple pancreaticoduodenectomy (the operation usually done for pancreatic cancer) may be indicated when most of the disease involves the head of the pancreas (Fig. 30-12). A modification of this operation is the pancreaticoduodenectomy with preservation of the stomach and pylorus (Fig. 30-13). The theoretical advantage of this procedure is that it preserves gastric function and avoids some of the nutritional disturbances that may follow the standard pancreaticoduodenectomy. Ninety-five percent distal pancreatectomy entails removal of the spleen and all the pancreas except for a thin rim of tissue that lies within the C loop of the duodenum. Because the spleen is sacrificed and all the patients require insulin after this operation, it is not done frequently. Today, total pancreatectomy is performed rarely because of the uniform requirement for insulin and the significant alterations in digestive and absorptive functions that follow. A new operation, in which the head of the pancreas is resected but the entire duodenum is preserved, may be preferred to the Whipple procedure in those patients in whom the head of the pancreas is the major focus of disease. The theoretical advantages of this procedure include the preservation of gastroduodenal and biliary continuity and function and a low likelihood of the development of diabetes, since most of the pancreas is preserved.

The mortality rate of the various pancreatic resection operations is less than 5 percent. The significant drawback associated with major pancreatic resections is the creation of the diabetic state, which is often difficult to control. For this reason, such procedures should not be done in unreliable alcoholic patients who may be addicted to narcotics. Pain is relieved in 80 to 90 percent of patients in the first several years after these operations, and even after 5 years, about 80 percent continue to be pain-free. Tables 30-8 through 30-10 list late results, mortality rates, and long-term pain relief of pancreaticoduodenectomy for chronic pancreatitis in various hands.

Nonsurgical destruction of the celiac ganglia may have a place in the management of patients with severe pain from chronic pancreatitis. In those in whom the surgeon and gastroenterologist wish to avoid or delay pancreatic resection, percutaneous injection of the ganglia with absolute ethanol may provide some relief for a period of time. Consultation with an anesthesiologist interested in pain management may be helpful. Although permanent pain relief is unusual with this technique, a limited period of improvement may be useful for patient and physician alike, resulting in a better rapport and a better understanding of the patient's commitment to a treatment program. In a few cases, spontaneous improvement in the pain may even occur.

Pancreatic Transplantation

In an effort to avoid diabetes, autografts of the body and tail of the pancreas have been done in a small number of patients who were subjected to total or near-total pancreatectomy. Success has been limited, and more experience is required before this approach can be recommended widely.

Prognosis

In patients with alcoholic chronic pancreatitis, life expectancy is reduced by 10 to 15 years, and the mean age at death is approximately 50 years. The principal causes of death include upper respiratory malignancies (most patients are heavy cigarette smokers), malnutrition, complications of diabetes, and suicide. Abuse of ethanol and/or drugs is often contributory.

PSEUDOCYSTS

Pancreatic pseudocysts are localized collections of fluid with high concentrations of pancreatic enzymes. They occur usually as a complication of pancreatitis, although some occur after trauma. Pseudocysts are located either within the pancreatic parenchyma or in one of the potential spaces that separate the gland from the adjacent abdominal viscera. Most of the time they are found in the lesser peritoneal sac behind the stomach, but they also may occur at great distances from the pancreas (e.g., neck, mediastinum, pelvis). They are called pseudocysts because, unlike congenital cysts or cystic neoplasms, they lack a true epithelial lining.

The pathogenesis of pseudocysts depends on the clinical setting in which they develop. During an attack of acute pancreatitis, fluid extravasates from the pancreas, where it becomes walled off by surrounding structures and fails to reabsorb as the inflammation resolves. In patients with chronic pancreatitis, ductal strictures or stones may obstruct the pancreatic duct or its branches and produce localized ductal dilatations. These areas coalesce and lose their epithelial lining as they enlarge to form a pseudocyst. Some pseudocysts follow trauma and are the result of either pancreatitis or ductal disruption with direct leakage of fluid from the gland. Occasionally, pseudocysts form after the pancreas is injured during splenectomy.

Ethanol abuse and biliary tract disease account for most of the cases, with men outnumbering women 2:1 in most series. The average age at diagnosis is 45 years. About one-third of pseudocysts are located in the head and two-thirds in the body or tail of the gland. Multiple cysts occur in 20 percent of patients.

Clinical Manifestations

Pseudocysts should be suspected when a patient with acute pancreatitis fails to recover after 5 to 7 days of treatment or, after some improvement, begins to deteriorate again. In patients with chronic pancreatitis, there is often no acute episode to herald the appearance of the cyst. Instead, the patient may complain of abdominal pain or symptoms that reflect compression of an adjacent viscus by the cyst. Nausea, vomiting, and weight loss may reflect duodenal or gastric outlet obstruction from a pseudocyst. Jaundice may be present if the cyst compresses the common bile duct.

Diagnostic Studies

Ultrasound (US) and CT scanning are the most useful tests to diagnose pseudocysts. However, CT scans are more sensitive and specific than US and have the additional advantages of superior resolution and more precise localization (Fig. 30-14). Moreover, bowel gas, obesity, and displacement of intraabdominal structures from prior surgery may impair US images. The main disadvantages of CT are the relatively high cost compared with US and exposure to ionizing radiation. CT scan should be used for the initial evaluation of patients suspected of having pseudocysts; US should be reserved for follow-up examinations.

ERCP also has been used before surgery to provide anatomic information about pseudocysts, but it should not be done routinely. ERCP is indicated in jaundiced patients to differentiate between common bile duct compression by the cyst and a stricture of the intrapancreatic portion of the common duct caused by a fibrotic pancreas. In the former case, cyst drainage alone would relieve the jaundice, whereas in the latter instance, a biliary bypass procedure to relieve the biliary obstruction also would be required. Because it may introduce bacteria into a previously sterile cyst, ERCP should be performed only when cyst drainage is to be done within 24 h.

Complications

The most feared complications of pseudocyst are hemorrhage, rupture, and infection. Hemorrhage occurs in approximately 6 percent of patients, rupture into the gastrointestinal tract or peritoneal cavity in 7 percent, and infection in 14 percent. Other complications include obstruction of the gastrointestinal tract (duodenum or stomach) in 3 percent and common bile duct obstruction in 6 percent.

Hemorrhage, from erosion of pseudocysts into major blood vessels (e.g., splenic or gastroduodenal arteries), may be massive. Arteriography is the initial diagnostic procedure of choice if the patient is stable. It offers the opportunity for selective embolization to plug the bleeding vessel. This may be definitive or, by temporarily stopping the hemorrhage, allow resuscitation of a critically ill patient and permit surgery to proceed on an urgent rather than emergency basis. Frequently, blood loss may be so rapid that it precludes diagnostic procedures and requires emergency operation. Both the cyst and the bleeding vessel should be excised, if possible. Even with aggressive surgical management, the mortality rate is at least 30 percent.

If the pseudocyst ruptures freely into the peritoneal cavity, the patient may develop symptoms and signs of acute peritonitis. Treatment of free rupture requires urgent operation with irrigation of the peritoneal cavity and usually external drainage of the cyst. Free rupture of a pseudocyst is a serious complication; the mortality rate is approximately 15 percent. Rupture into the gastrointestinal tract may produce vomiting or diarrhea but is usually not a surgical emergency.

Pseudocyst infection may occur in the early weeks after cyst formation as a complication of an episode of severe acute pancreatitis. Pseudocyst infection must be distinguished

from infected pancreatic necrosis. Less commonly, superinfection of a mature cyst is heralded by fever and leukocytosis in a patient who has not been acutely ill. In either case, infection can be proved by percutaneous drainage of the cyst under US or CT guidance, with Gram stain and culture of the fluid. Infected pseudocysts may resolve permanently with percutaneous drainage for a week or so. However, if percutaneous drainage does not produce prompt clinical improvement, surgery is required.

Treatment

Pseudocysts that develop in patients with chronic pancreatitis are unlikely to resolve spontaneously and may be viewed as mature from the time of discovery. About 40 percent of acute pseudocysts (those which develop during an episode of acute pancreatitis) will disappear spontaneously as the pancreatitis resolves, however. If mature pseudocysts are asymptomatic and are less than 5 cm in diameter, they probably require no treatment, since serious complications generally occur in cysts that are larger. These patients should be followed and the cysts reevaluated at 3- to 6-month intervals with US. The onset of symptoms, growth in the size of the cyst, or development of a complication demands intervention.

Surgery

The three types of surgical procedures used to treat pseudocysts are resection, external drainage, and internal drainage. Elective surgery should be delayed until the pseudocyst is mature, i.e., until the wall is firm enough to hold sutures. This requires 4 to 6 weeks for cysts that arise during acute pancreatitis. For cysts that arise during chronic pancreatitis, no wait is necessary. At operation, the interior of the cyst should be inspected for evidence of tumor, and a biopsy of the cyst wall should be obtained to exclude a cystic neoplasm (e.g., cystadenocarcinoma). *Resection* is the preferred method of treatment but is usually possible only for cysts in the tail of the gland, when the head and body are normal or minimally changed (e.g., traumatic pseudocysts). *External drainage* is preferred when the cyst wall is not sufficiently thickened to allow anastomosis to the gut lumen. A large-bore tube is sewn into the cyst lumen, and the other end is brought out through the abdominal wall. Pancreatic fistulas complicate about 20 percent of external drainage operations and usually close within a few months. Fistulas and recurrent cysts are more common when the cyst communicates with the pancreatic duct. *Internal drainage*, in which the cyst is anastomosed to a Roux-en-Y limb of jejunum (cystojejunostomy), to the posterior wall of the stomach (cystogastrostomy), or to the duodenum (cystoduodenostomy), is the most frequently used operation for pseudocysts. Dependent drainage by cystogastrostomy is preferred for collections adjacent to the stomach, whereas cystojejunostomy can be done for cysts in any location. Obliteration of the cyst cavity usually follows internal drainage within a few weeks. Patients recover rapidly and are usually eating within a week. Infected pseudocysts adherent to the stomach can be drained by cystogastrostomy. Otherwise, drainage should be external because the suture line of a Roux-en-Y cystojejunostomy may not heal in an infected surgical field.

Mature pseudocysts also may be treated nonsurgically, but these techniques are still being evaluated. External drainage may be established with a narrow-lumen catheter placed percutaneously via US or CT guidance. Successful obliteration is claimed in as many as 80 percent of cases; if the cyst persists, surgery should be performed. Nonsurgical management is contraindicated in patients with cyst rupture or hemorrhage or when adjacent vascular structures preclude a percutaneous approach. In addition, if there is ever concern that the "pseudocyst" may actually represent a cystic neoplasm, surgery is required. This should be considered if there is no history of pancreatitis, septations are seen within the cyst on CT scan, or cyst calcifications are present. If any of these features are present, suspicion that the cyst represents a neoplasm should be high. Strong consideration should be given to surgical intervention rather than nonoperative management (e.g., percutaneous drainage). Table [30-11](#) lists the characteristics of cystic neoplasms of the pancreas.

Prognosis

Pancreatic pseudocysts recur at a rate of approximately 10 percent, although most of these probably are new cysts that arise from recurrent episodes of pancreatitis. After external drainage, the rate is approximately 20 percent. Even when pseudocysts do not recur, many patients experience chronic pain, a manifestation of underlying chronic pancreatitis.

Pancreatic Ascites

Ascites may occur when pancreatic secretions collect within the abdomen following disruption of the pancreatic duct or when a pseudocyst continuously leaks into the abdomen. A leak into the retroperitoneum may track some distance from its source, and the fluid collection may appear in the mediastinum, pleural space (pancreatic pleural effusion), or groin. Most patients are male alcoholics with chronic pancreatitis, although ascites also may occur following pancreatic trauma. Symptoms of a distinct episode of acute pancreatitis rarely precede pancreatic ascites. Patients usually present with increasing abdominal girth, weight loss, and muscle wasting. Although the presentation can be confused with cirrhosis, abdominal paracentesis is diagnostic. The combination of high protein content (>3 g/100 mL) and amylase concentration in the ascitic fluid greater than the serum amylase concentration is pathognomonic. The initial therapy should consist of parenteral nutrition, especially if malnutrition is severe. A somatostatin analogue to suppress pancreatic secretions also may be useful. In some cases, the ascites resolves with this treatment, which should continue for 10 to 14 days. If the ascites persists, operation is indicated and should be preceded by endoscopic retrograde pancreatography to show the site of the ductal disruption. If it is located in the distal pancreas, distal pancreatectomy may be appropriate. If the leak is located in the head or body of the gland, internal drainage using a Roux- en-Y limb of jejunum is preferable. The success rate is approximately 85 percent, but in patients who are severely debilitated, postoperative morbidity is considerable.

TUMORS OF THE PANCREAS

Ductal Adenocarcinoma

In the United States, cancer of the exocrine pancreas is the fourth leading cause of cancer death in men and the fifth in women. Each year, about 28,000 cases will be diagnosed, and almost as many patients will die of advanced disease. Ductal adenocarcinoma of the pancreas accounts for approximately 90 percent of exocrine tumors, and it is a characteristically aggressive lesion. At the time of diagnosis, the tumor is confined to the pancreas in fewer than 10 percent of patients, 40 percent have locally advanced disease, and over 50 percent have distant spread. More than 95 percent of patients eventually die of their disease, and even after resection for attempted cure, the median length of survival is only 18 to 20 months.

Clinical Manifestations

Pancreatic cancer occurs in the head of the pancreas in 75 percent of cases and in the body and tail of the gland in the rest. This distinction is important because lesions of the head, close to the bile duct, may produce obstructive jaundice when they are still small and curable. Indeed, biliary obstruction, in the absence of other symptoms, is associated with a better prognosis. Painless jaundice alone (13 percent) is uncommon, however. About 75 percent of patients with carcinoma of the head of the pancreas present with *obstructive jaundice, weight loss* (average 20 pounds), and deep-seated *abdominal pain*. The jaundice is usually unrelenting but fluctuates in about 10 percent of patients. Cholangitis is uncommon (10 percent). Jaundice may be accompanied by pruritus. When back pain occurs (25 percent), it is associated with a worse prognosis, and it implies retroperitoneal invasion by the tumor. The sudden onset of diabetes mellitus is noted in about 20 percent of patients. Body and tail tumors almost always produce symptoms late and are usually far advanced at the time of diagnosis. Very few are resectable, and the prognosis is particularly grim.

Physical examination and routine laboratory tests are primarily of value in supporting the diagnosis of obstructive jaundice and in detecting very advanced disease. Hepatomegaly, present in 65 percent of patients, is a nonspecific finding. Although it can indicate liver metastases, the liver also may enlarge as a result of bile duct obstruction alone. A palpable abdominal mass (10 percent) and ascites (5 percent) suggest advanced disease. *Courvoisier's sign*, a distended and palpable gallbladder in the jaundiced patient, suggests malignant obstruction. It is present in only 25 percent of patients.

Diagnostic Studies

The total bilirubin level tends to be greater with malignant obstruction compared with benign causes such as choledocholithiasis (e.g., mean levels of 15 versus 5 mg/dL). Serum amylase elevations are uncommon (5 percent) in patients with pancreatic cancer.

A number of screening tests for pancreatic cancer have been investigated. Ideally, such a test would be inexpensive, abnormal early in the course of the disease when it was still curable (high sensitivity), and normal in other conditions (high specificity). Of the many *serum markers* that have been evaluated (CEA, POA, GT-II, CA 19-9, DU-PAN-2), the results with CA 19-9 have been the most encouraging. However, even CA 19-9, a monoclonal antibody to the Lewis blood group antigen, has a sensitivity of only 80 percent. The smallest lesions and presumably those most amenable to cure are least likely to be detected. Gastric, colorectal, and other carcinomas also may be associated with elevations in CA 19-9, so the specificity is only about 90 percent.

The diagnostic approach to pancreatic cancer varies depending on whether the patient presents with obstructive jaundice or other symptoms. In those with jaundice, the clinician's attention is directed rapidly toward an assessment of the biliary tree. An abdominal CT scan, perhaps preceded by a US examination, is generally performed early in the evaluation. ERCP also may be done. The presumptive diagnosis usually can be arrived at within several days. In patients without jaundice, where the complaints are weight loss, pain, or other nonspecific problems, a number of other studies may precede a CT scan and/or ERCP examination. Nevertheless, in both groups, CT and ERCP have become the mainstay for diagnosis of this disease.

Ultrasound

US is relatively inexpensive and has a sensitivity of 70 percent and a specificity of 95 percent for the diagnosis of pancreatic cancer. However, the examination is unsatisfactory because of body habitus or overlying bowel gas in 20 percent of patients. When it demonstrates dilated intrahepatic and extrahepatic biliary radicles, it confirms the presence of extrahepatic biliary obstruction. US also helps to identify gallstones as a cause of obstruction.

Computed Tomography

Spiral CT has a sensitivity of 85 percent and a specificity of 95 percent for the diagnosis of pancreatic cancer. Lesions larger than 2 cm are usually detectable (Fig. [30-15](#)). More reliable than US, CT visualizes the entire pancreas and also provides some information about resectability (e.g., liver metastases and major vascular involvement may be detected). Nodal enlargement may reflect metastases, but nodes smaller than 1 to 2 cm may be missed altogether, and inflammatory changes in nodes cannot be distinguished from neoplastic ones. About 20 percent of metastatic liver lesions are overlooked.

Endoscopic Retrograde Cholangiopancreatography

ERCP has a sensitivity of 95 percent but a specificity of only 85 percent. This is due to the difficulty in distinguishing between cancer and chronic pancreatitis. The study can be performed successfully in over 90 percent of patients and detects some tumors that are not visualized by CT.

The characteristic finding that suggests the diagnosis of pancreatic cancer is a constriction of both the pancreatic and bile ducts in the head of the gland. This is the so-called double-duct sign (Fig. [30-16](#)). However, ERCP is superfluous in some patients. If the patient has a typical history for pancreatic cancer and a mass in the head of the pancreas is seen on CT scan that is consistent with the diagnosis, ERCP is not needed. During ERCP, cytologic study of cells obtained directly from the pancreatic duct may reveal cancer in up to 80 percent of cases. It also may be possible to determine the origin of a periampullary tumor (pancreatic versus duodenal, ampullary, or bile duct) with this technique. Therapeutic

ERCP also permits endoscopic stent placement. Although this is not indicated for preoperative biliary decompression, it provides effective palliation of the jaundice for certain patients who may not need an operation.

Fine-Needle Aspiration

Preoperative *fine-needle aspiration (FNA)* for cytologic examination can be obtained percutaneously under CT or US guidance. The technique has a sensitivity of about 85 percent and a specificity of almost 100 percent. *However, the clinician must remember that a negative FNA never reliably eliminates the possibility that cancer is present.* FNA is justified when the knowledge that the diagnosis is pancreatic cancer will unquestionably alter the patient's management. For example, a patient with cancer of the body of the pancreas and retroperitoneal extension of tumor seen on CT scan is not likely to benefit from operation. Proof of the diagnosis by FNA would avert the need for operative biopsy and permit the use of other palliative measures (radiotherapy, chemotherapy, celiac ganglion injection, etc.) (Fig. [30-17](#)). The routine use of FNA for patients with a mass in the head of the pancreas and obstructive jaundice is unnecessary.

Laparoscopy

Laparoscopy has been used to stage patients with pancreatic cancer. At least 20 percent of patients with no evidence of metastases from the preceding tests have small lesions evident on the liver surface, omentum, or peritoneum that can be seen laparoscopically. Such documentation of unresectability could be helpful in the patient with a body or tail lesion who would not benefit from surgery. It has no value in the patient who requires surgical bypass for relief of jaundice and gastric outlet obstruction.

Magnetic Resonance Imaging

MRI appears to offer no advantage over CT scan.

Upper Gastrointestinal Series

An upper gastrointestinal (UGI) series may show displacement of the stomach and widening of the duodenal C loop, but these are relatively late changes and are usually detectable on CT examination. A UGI series may be of some value to assess gastric outlet obstruction but is not performed routinely in the absence of specific indications.

Percutaneous Transhepatic Cholangiography

Percutaneous transhepatic cholangiography (PTC) permits visualization of the proximal biliary tree in patients with ductal obstruction who cannot be studied from below with ERCP. Although PTC is the preferred method for proximal bile duct tumors (e.g., Klatskin tumors), ERCP provides more diagnostic information and is better for periampullary lesions.

Angiography

This study is not performed for diagnosis but may provide information about resectability of the cancer. Rarely, complete occlusion of the celiac, hepatic, and/or superior mesenteric arteries or the portal vein suggests that the lesion is unresectable. More commonly, distortion of the vessels is seen, which is more difficult to interpret. The sensitivity and specificity of angiography to determine resectability are each 70 percent. Preoperative recognition of such vascular anomalies as a right hepatic artery arising from the superior mesenteric artery (25 percent of patients) can facilitate their preservation at surgery. Nevertheless, such anomalies also can be determined at the time of operation. Since modern spiral CT provides similar information, most surgeons do not routinely perform angiography.

Summary

An *algorithm* for the *management* of the patient with suspected cancer is presented in Fig. [30-18](#). Spiral CT scan provides the best overall initial assessment and may be the only test required. If the history and clinical picture suggest pancreatic cancer and CT reveals a mass in the head of the pancreas without evidence of unresectability, laparotomy may be undertaken without additional studies. Exploration provides the best means of confirming resectability; and even if resection is not possible, most patients will benefit from biliary and gastric bypass. If the CT suggests hepatic or peritoneal metastases or vascular invasion that precludes resection, percutaneous FNA may be appropriate if a positive biopsy will avoid an operation. If the biopsy is negative, laparoscopy should be the next step. In some patients with a general medical condition that makes operation too risky, no further evaluation may be necessary. Age alone is not a contraindication to major pancreatic resection. Endoscopically placed biliary stents effectively relieve biliary obstruction and are an alternative to surgical biliary bypass in patients who do not undergo operation.

If the mass involves the body and tail of the gland, usually there is evidence of metastatic disease or local spread. Since these patients seldom benefit from palliative operations, percutaneous FNA is often indicated to provide a tissue diagnosis before initiating adjuvant therapy. Occasionally, these tumors appear resectable or the biopsy is negative or equivocal. In these cases, laparoscopy for further assessment may be reasonable. The patient should undergo laparotomy under the same anesthetic if the attempted biopsy is unsuccessful or the lesion appears resectable.

If the patient with suspected pancreatic cancer has a normal or nondiagnostic CT, ERCP is indicated. If both the CT and ERCP are normal, pancreatic cancer is an extremely unlikely cause of the symptoms that are being evaluated.

Differential Diagnosis

The other periampullary neoplasms (carcinoma of the ampulla of Vater, distal common bile duct, or duodenum) also may present with abdominal pain, weight loss, and obstructive jaundice. While the preoperative studies may suggest the correct diagnosis, sometimes this is possible only at operation. Occasionally it is not apparent even at that time, and the pathologist may have to determine the true origin of the neoplasm. This element of diagnostic uncertainty is acceptable because the treatment (pancreaticoduodenectomy) is the same for each of these lesions. Chronic pancreatitis also can be confused with pancreatic cancer, and it may be impossible to distinguish between the two even at operation. Nevertheless, pancreaticoduodenectomy is an acceptable treatment for this as well.

Treatment

Pancreatic resection for pancreatic cancer is appropriate only if all evidence of gross tumor can be removed with a standard resection, i.e., if cure is considered possible. Resection for palliation is not commonly done. In the absence of evidence of distant metastases (e.g., to the liver, peritoneal surfaces, distant lymph nodes), resectability usually depends on whether the tumor has grown to involve adjacent major vascular structures. These structures are the superior mesenteric and portal veins and the superior mesenteric and hepatic arteries. The vena cava also may be involved. Because the pancreas is so close to the portal vein and superior mesenteric vessels, and because of the propensity of pancreatic cancer to spread early to adjacent structures, only about 20 percent of tumors of the head are resectable at the time of exploration. Fewer than 5 percent of tumors of the body and tail can be removed.

In some patients, the diagnosis has already been determined preoperatively by FNA or other technique. If this is not the case, most surgeons will try to establish a tissue diagnosis at an early stage of the operation. When the cancer has extended to the surface of the gland, a simple “shave” biopsy of the lesion is safe and reliable. If the lesion is located deep within the pancreas, a needle biopsy or FNA is preferred. For cancers in the head of the pancreas, the needle is often inserted into the gland through the duodenum so that if a pancreatic fistula develops, it will drain into the bowel rather than leak into the peritoneal cavity. Although these techniques are accurate and reasonably safe, they are often inconclusive with the small deep-seated lesions that are most likely to be cured by resection. This occurs because the tumor is surrounded by a zone of chronic inflammation and may be missed altogether when the needle is inserted. Thus most experienced pancreatic surgeons are willing to perform a pancreaticoduodenectomy on the basis of clinical inferences, drawn preoperatively and at operation, that suggest the diagnosis of cancer. Biopsy proof is not required.

Pancreaticoduodenectomy (Whipple Resection)

This is the most commonly performed operation for carcinoma of the pancreatic head (see Fig. [30-12](#)). It includes resection of the distal stomach, gallbladder, common bile duct, head of the pancreas, duodenum, proximal jejunum, and regional lymphatics. Truncal vagotomy used to be done to reduce the risk of marginal ulceration, but it is done infrequently today. Instead, many patients are managed with acid antisecretory agents. Restoration of gastrointestinal continuity requires pancreaticojejunostomy, choledochojejunostomy, and gastrojejunostomy. The operative mortality rate in expert hands is less than 5 percent. When death occurs, it is most often due to complications such as postoperative pancreatic or biliary fistula, hemorrhage, and infection.

Pylorus Preservation

About a third of all pancreatic surgeons in the United States now perform a modification of the Whipple procedure in which the stomach and pylorus are preserved (see Fig. [30-13](#)). This operation is preferred by some because it is believed that it avoids the undesirable nutritional sequelae (i.e., weight loss, dumping syndrome, diarrhea) that may follow the standard Whipple procedure. Although the lymphatic dissection is less radical than with the standard operation, there is no evidence as yet that it is associated with a lower survival rate.

In the United States, the 5-year survival rate in all patients who have undergone pancreatic resection for pancreatic cancer is approximately 10 percent. In those patients without lymph node involvement, the 5-year survival rate is about 35 percent.

Total Pancreatectomy

A number of more extensive procedures have been proposed to improve the cure rates associated with the Whipple operation. To date, none has proved more effective. One of the proposed procedures is *total pancreatectomy*. The rationale for this procedure included the observation that in 30 to 40 percent of pancreatic cancer patients, the tumor was multicentric and would not be removed completely by a Whipple resection. However, most series of total pancreatectomy showed no evidence of increased survival, and recent studies have suggested that concerns regarding multifocal disease were unwarranted. However, it may still be appropriate when tumor is present at the planned site of transection of the pancreas or when the gland is too soft to permit a satisfactory anastomosis to the jejunum.

Extended Whipple Resections

These operations involve wider soft tissue and lymph node dissections than the standard Whipple and include resections of the superior mesenteric vessels and portal vein when they are involved with tumor. Some retrospective data have suggested a better outcome with this aggressive approach. Nevertheless, these operations are best done within the framework of an experimental protocol, since the morbidity is greater and increased efficacy has not been proven.

Palliative Operations

Surgical palliation in patients with cancer of the head of the pancreas is directed toward relief of obstructive jaundice, gastric outlet obstruction, and pain. Patients with body and tail lesions are less likely to have jaundice or duodenal obstruction, but pain is often significant. Nonoperative techniques for pain relief are preferred.

At some time in the course of the disease, obstructive jaundice develops in about 70 percent of patients with pancreatic cancer. Cholecystojejunostomy and choledochojejunostomy are both safe and are the procedures of choice to relieve the biliary obstruction (Fig. 30-19). If the gallbladder is chosen for the anastomosis, the surgeon must be certain that the cystic duct is patent and enters the common bile duct away from the tumor. An intraoperative cholangiogram may be necessary to confirm this. Roux-en-Y jejunostomy theoretically is better than simple loop reconstruction because it prevents reflux of intestinal contents into the biliary tree and secondary cholangitis. Since most of these patients die within 7 or 8 months, this is rarely a concern, and the simpler loop reconstruction is usually done. Jaundice is relieved in 85 to 90 percent of patients. When it is not, it is usually because of irreversible hepatic parenchymal damage from high bilirubin levels and long-standing obstruction.

For patients in poor general medical condition or in whom the tumor is known to be unresectable, nonoperative palliation of obstructive jaundice is preferred. Percutaneous or endoscopically placed plastic biliary stents are each effective, but the latter are better accepted by patients. The major problem with either technique is stent occlusion, associated with recurrent jaundice and sepsis. This requires stent replacement every 3 to 4 months in up to two-thirds of patients. Metal stents have a larger diameter and may last longer.

About one-third of patients with cancer of the head of the pancreas eventually develop gastric outlet obstruction, which is treated effectively by a gastrojejunostomy (Fig. 30-20). If obstruction is not present at the time of initial exploration and biliary bypass, there is some debate about whether gastrojejunostomy should be done prophylactically. Because it is difficult to predict which patients will obstruct later in the course of their disease, and it does not appear to increase the operative morbidity or mortality rates, many surgeons do perform gastrojejunostomy in patients who are expected to survive more than a few months.

Pain is a symptom in up to 80 percent of patients with pancreatic cancer, and its severity and persistence correlate well with the stage of the disease. This problem is managed best by chemical destruction of the celiac nerves, achieved by percutaneous injection of 50% ethanol into the region of the ganglia under CT guidance. This also can be done at operation. Adequate pain relief is possible in up to 80 percent of patients, although it often recurs. Celiac injection can be repeated if needed.

Adjuvant Therapy

Neither radiation nor chemotherapy given singly is effective. Radiotherapy combined with 5-fluorouracil given as a radiosensitizer has been shown to increase survival in both patients with resectable and those with unresectable disease.

Prognosis

Most patients with pancreatic cancer die within 1 year of the time that the diagnosis is made (Table 30-12). In patients with carcinoma of the head of the pancreas who have undergone resection, prognosis is related to whether lymph node metastases have already occurred. In patients with positive nodes, the 5-year survival rate is less than 10 percent; in those patients without nodal involvement, the 5-year survival rate is approximately 35 percent.

Carcinoma of the Ampulla of Vater

About 10 percent of periampullary cancers arise from the ampulla of Vater. These tumors produce symptoms that are similar to those seen in pancreatic cancer, although weight loss and pain are less prominent. Obstructive jaundice occurs early and is present in almost all patients because of the location of the lesion near the bile duct. Jaundice may fluctuate because as pieces of the tumor slough into the duodenal lumen, the degree of obstruction of the bile duct varies. The stools often contain occult blood from the ulcerating neoplasm. The lesion is visible by endoscopy if it extends into the duodenal lumen, and a tissue diagnosis can be obtained easily by biopsy. If the neoplasm resides entirely within the ampulla, the diagnosis of ampullary neoplasm may not be possible. It may be suspected if the ERCP or CT scan suggests that the obstruction is at the most distal point in the bile duct. Sometimes it may not be possible to distinguish an ampullary tumor from one that arises from the head of the pancreas, even at the time of operation. Because these tumors grow more slowly and tend to remain localized longer than pancreatic cancers, many more are resectable for cure. The treatment consists of pancreaticoduodenectomy, and the 5-year survival rate is 35 to 40 percent. Occasionally, in patients who are not candidates for the Whipple resection, local excision of the tumor through a duodenotomy may be indicated.

Cystic Neoplasms

Approximately 90 percent of primary malignant neoplasms of the nonendocrine pancreas are carcinomas of duct cell origin. Several types of cystic neoplasms also occur that may be confused with pseudocysts.

Papillary-cystic neoplasm of the pancreas is a rare lesion that occurs almost exclusively in women under 25 years of age. It is often large and may present as a hemorrhagic cystic tumor in the body or tail of the gland. Although it can be locally invasive, metastases are unusual, and resection is generally curative. *Serous (microcystic) cystadenomas* are the most common benign neoplasm of the exocrine pancreas, and they occur with equal distribution throughout the gland. Histologically, they appear as numerous, small cysts lined by flattened cuboidal epithelial cells. The cell cytoplasm contains glycogen but no mucin. The cystic spaces are filled with a serous material. The tumor has been likened to the cut surface of the lung, with multiple epithelium-lined cystic spaces analogous to alveoli. *Mucinous cystadenomas* are premalignant lesions that often evolve into *cystadenocarcinomas*, which are frankly malignant tumors. These are most common in the body or tail of the pancreas and are often quite large (e.g., 10 to 20 cm) when diagnosed. Most patients are women (8:1). The cellular lining consists of mucin-producing columnar epithelial cells. They take the form of papillary projections into the lumen of the cyst, giving it an irregular appearance.

Manifestations

The most common symptom of any of these tumors is vague abdominal discomfort. Obstructive jaundice occurs in fewer than 10 percent of patients. Anorexia, weight loss, and weakness may occur, particularly with cystadenocarcinoma. Radiographic studies may reveal calcification within the cyst (at least 10 percent of patients), which is an uncommon finding with pseudocysts. Barium studies of the stomach and duodenum may reveal displacement by the tumor mass. Angiography characteristically reveals hypervascularity, and arteriovenous shunting is frequently observed. The hypervascularity differentiates these cystic tumors from pseudocysts and the usual type of ductal adenocarcinoma, both of which are hypovascular. US and CT demonstrate multicystic tumors, sometimes with solid components.

Treatment

Surgery is usually indicated because of the concern that a cyst may be malignant and to manage symptoms that may be present. At operation, the diagnosis of malignancy may be difficult. Benign serous cystadenomas typically have a honeycombed appearance on cut section. They may contain hundreds of small cystic spaces that vary in diameter from a few millimeters to a centimeter. These lesions apparently have no malignant potential, but they should be excised if it can be done safely. They should *not* be drained internally. Both the benign and malignant neoplasms may be unilocular or multilocular. The most characteristic feature of the malignant cystadenocarcinoma is the viscid mucinous fluid it contains. This fluid has the gross appearance of raw egg white and can be strung like egg white for 6 to 12 in. between the instrument and the tumor. Examination of a frozen section is helpful but is not reliable in distinguishing benign from malignant stages unless it shows frank carcinoma or papillary projections, which are almost synonymous with malignancy. Whether or not overt malignancy is present, these macrocystic lesions should be excised, since there is a high likelihood that they will become malignant. During the resection of neoplastic cysts, efforts should be made to remove the tumor intact (rupture may disseminate cells). In order to be certain of the diagnosis in patients with presumed pseudocysts, a frozen-section diagnosis should always be obtained on a portion of the cyst wall before proceeding with internal drainage.

Prognosis

Cystadenocarcinomas have a better prognosis than the usual adenocarcinomas of the pancreas. Although precise statistics are unavailable because of the rarity of the lesions, resection of apparently localized tumors probably cures about two-thirds of patients.

Islet Cell Neoplasms

Insulinoma

Insulinoma arises from the beta cells and is the most common islet cell neoplasm. It produces insulin, and the symptoms are all related to hypoglycemia. The tumors may occur at any age and are found equally distributed throughout the pancreas. About 80 percent are solitary and benign, and about 40 percent are less than 1 cm in diameter. Fifteen percent are malignant, and metastases are usually evident at the time of diagnosis. Fewer than 5 percent of insulinomas are associated with the multiple endocrine adenopathies (MEA) syndrome; these are usually multiple and are distributed throughout the pancreas.

Clinical Manifestations

Symptoms related to cerebral glucose deprivation include bizarre behavior, memory lapse, and unconsciousness. Some patients have been mistakenly diagnosed with psychiatric or neurologic illness. Symptoms related to sympathetic discharge include palpitations, nervousness, sweating, and tachycardia. Many patients gain significant weight because symptoms are produced by fasting; consequently, they eat excessively to avoid this. The classic diagnostic criteria (*Whipple's triad*) are usually present. These include (1) hypoglycemic symptoms produced by fasting, (2) blood glucose concentration below 50 mg/dL during symptomatic episodes, and (3) relief of the symptoms during intravenous administration of glucose.

Diagnostic Studies

The most useful diagnostic test, and the one indicated in most patients, is the demonstration of fasting hypoglycemia in the face of inappropriately high levels of insulin in the serum. The patient is fasted, and blood samples are obtained every 6 h for blood glucose and insulin measurements. The fast is continued until hypoglycemia or symptoms appear or for a maximum of 72 h. By 48 h, over 95 percent of the tumors have been diagnosed. If hypoglycemia has not developed after 70 h, the patient should be exercised for the final 2 h. Although the insulin levels are not always elevated in patients with insulinoma, they will be high relative to the blood glucose concentration. A ratio of plasma insulin to glucose greater than 0.3 is diagnostic. Proinsulin, which constitutes more than 25 percent of total insulin in about 85 percent of patients with islet cell tumors, also should be measured. Very high proinsulin levels (>50 percent) suggest that the insulinoma is malignant.

Provocative tests using a variety of drugs that stimulate the release of insulin (e.g., tolbutamide, leucine, arginine) have been used in some patients, but they are diagnostic in only 50 percent of patients and are rarely indicated. A calcium infusion provocative test also may be of value.

Differential Diagnosis

Fasting hypoglycemia occurs with some nonpancreatic tumors as well, and the clinical picture may be identical with that seen with insulinoma. Plasma insulin levels are normal, however. In most cases, the nonpancreatic tumors are large and are readily detected with various diagnostic studies. The majority are of mesenchymal origin (e.g., hemangiopericytoma, fibrosarcoma) and are located in the abdomen or thorax, although hepatoma, adrenocortical carcinoma, and a number of other lesions also can produce hypoglycemia. The principal mechanisms by which these tumors produce hypoglycemia are (1) secretion by the tumor of somatomedins (insulin-like peptides that mediate the effects of growth hormone), which can be measured by specific radioimmunoassay, and (2) inhibition of gluconeogenesis or glycogenolysis. Surreptitious self- administration of

insulin is also seen, most often by individuals who have access to insulin in their job. Insulin antibodies are usually detectable in the serum if these individuals have been taking injections for at least 2 months. Circulating C-peptide levels are normal in these patients but usually are elevated in patients with insulinoma.

Treatment

Surgery should be undertaken promptly, since the tumor may be malignant and, with repeated attacks of hypoglycemia, permanent cerebral damage may occur.

Surgical Therapy

Because 10 to 20 percent of the tumors are not easily identified at operation, preoperative localization studies are helpful. CT scans are usually obtained, but the majority of lesions are not seen because of their small size. Selective angiography demonstrates the tumor in about 50 percent of patients. Percutaneous transhepatic venous catheterization allows sampling of blood draining from the pancreas at multiple sites along the splenic and portal veins. Insulin is measured in the samples, and the point where the insulin concentration rises sharply indicates the location of the tumor. If patients are receiving diazoxide, it should be discontinued for at least a week prior to operation because of the reported association with hypotension on induction of anesthesia.

At operation, the entire pancreas must be fully mobilized and carefully palpated to find the tumor. Intraoperative US also should be done. If the tumor is found, it should be enucleated if it is superficial. If it is deeply situated, it should be removed as part of a partial pancreatectomy. Tumors in the head of the gland can nearly always be enucleated, so a pancreaticoduodenectomy is not often indicated. If the tumor cannot be found, and the results of venous sampling studies suggest its location, that portion of the pancreas should be resected. Under these circumstances, a Whipple resection may be appropriate if these studies indicate that the tumor is in the head of the gland. If no such localizing information is available, it is common practice to remove the distal half of the pancreas and have a pathologist slice the specimen into thin sections to look for the tumor. If it is found, the operation is concluded. If not, more is removed until an 80 percent distal pancreatectomy has been done. Since the tumors are evenly distributed, this confers an 80 percent probability of cure.

For islet cell hyperplasia, nesidioblastosis, or multiple benign adenomas, distal subtotal pancreatectomy is indicated to sufficiently decrease insulin levels so that medical management can be more effective. For malignant insulinomas, both resection of the primary tumor and debulking of metastases are warranted if technically feasible and safe.

Medical Therapy

Frequent feedings supplemented with a slowly absorbed form of carbohydrate such as cornstarch may be helpful, but usually patients require treatment with diazoxide, which suppresses the release of insulin. Fewer than two-thirds of patients are controlled successfully with this agent, and many develop significant adverse reactions that include hirsutism and salt and water retention. For incurable islet cell tumors, streptozocin appears to increase survival, with 60 percent of patients living an additional 2 years. Toxicity of the drug is considerable, however, so streptozocin is not used routinely as an adjuvant to surgical therapy.

Gastrinoma (Zollinger-Ellison Syndrome)

The principal manifestations of the Zollinger-Ellison syndrome are due to gastric acid hypersecretion caused by excessive gastrin production by the tumor. Although the normal pancreas does not contain gastrin-producing cells, most gastrinomas arise in the pancreas. The others are found chiefly in the duodenum. In the pancreas, about 60 percent of gastrinomas are non-beta islet cell carcinomas, 25 percent are solitary adenomas, and 10 percent are microadenomas or hyperplasia. About 5 percent of gastrinomas occur in the first or second part of the duodenum. One- quarter of gastrinomas are associated with the *multiple endocrine neoplasia type I (MEN I)* syndrome, in which case the tumors are usually multiple and benign. Those without MEN I (sporadic gastrinomas) are more often single and malignant. The diagnosis of malignancy requires the histologic demonstration of blood vessel invasion or the finding of metastatic disease, since the cellular characteristics of benign and malignant lesions are similar. Even when malignancy is present, the tumors are often slow-growing and compatible with a long survival, however. Thus the principal threat to life is the severe peptic ulcer diathesis that the hypergastrinemia provokes, and this is the main focus of therapy.

Clinical Manifestations

Symptoms of gastrinoma are chiefly those of gastric acid hypersecretion: severe, refractory peptic ulcer disease and, often, diarrhea. The diarrhea results from fat malabsorption, since the excessive quantities of gastric acid destroy pancreatic lipase. In addition, the acid damages the small bowel mucosa, and the intestine is overwhelmed with large volumes of gastric and pancreatic secretions. About 5 percent of patients present with diarrhea alone. The diagnosis should be suspected when (1) patients present with ulcer disease that is refractory to the usual forms of acid antisecretory therapy (antacids and H₂-receptor antagonists), (2) patients present with multiple ulcers or ulcers in unusual locations (e.g., third-fourth portion of duodenum, proximal jejunum), and (3) ulcers recur after surgical procedures that would normally cure the ulcer diathesis.

Diagnostic Studies

The diagnosis of gastrinoma requires the demonstration of fasting hypergastrinemia (>200 pg/mL of blood) in the face of gastric acid hypersecretion (basal acid output >15 mEq/h with an intact stomach or >5 mEq/h after ulcer surgery). Most patients with gastrinoma have serum gastrin values that exceed 500 pg/mL, and some are as high as 10,000pg/mL. When the serum gastrin level is in the range 200 to 500pg/mL, however, a secretin provocative test is usually done to confirm the diagnosis. Following 2 units/kg secretin (bolus, I.V.), a rise in the serum gastrin level of 200 pg/mL within 15min or a doubling of the fasting gastrin level is diagnostic. An upper gastrointestinal series will reveal the ulcers, and in the stomach, prominent rugal folds and excessive luminal secretions are often seen. The duodenum and small bowel may be hyperactive, with edema in the bowel wall and accelerated transit time. A CT scan should be done in all patients, and it often shows the tumor. If not, angiography with portal venous sampling to find “hot spots” of gastrin may be helpful.

Differential Diagnosis

The differential diagnosis of gastrinoma includes gastric outlet obstruction, retained antrum after a Billroth II gastrectomy, and antral G cell hyperplasia, all of which are characterized by hypergastrinemia and acid hypersecretion. In these conditions, the secretin test is negative.

Treatment

Medical Treatment

The medical treatment of gastrinoma is directed at control of the acid hypersecretion. Although H₂-blocking agents (cimetidine, ranitidine) may be effective, because of its long duration of action, omeprazole (a proton pump blocker) is the drug of choice. The dose should be adjusted to keep the H⁺ output less than 5 mEq during the hour preceding the next dose. Over 90 percent of patients are adequately controlled in this fashion. In patients with malignant gastrinoma, a combination of streptozocin and fluorouracil, with or without doxorubicin, may be effective.

Surgical Treatment

All patients with sporadic gastrinoma should undergo surgical exploration, unless there is evidence of extensive metastatic disease. Because patients with gastrinoma and MEN I syndrome have multiple small pancreatic tumors that cannot all be resected, surgery is generally not indicated (Fig. 30-21).

At operation, if the tumor is found in the pancreas, it should be enucleated if possible. If not, a distal pancreatectomy may be required. Pancreaticoduodenectomy is rarely indicated. Ninety percent of the tumors are found in the triangle defined by the junction of the cystic and common bile ducts, the junction of the second and third portions of the duodenum, and the junction of the neck and the body of the pancreas (“gastrinoma triangle”). US may be useful to detect small lesions deep within the substance of the gland. Intraoperative endoscopy to transilluminate the duodenal wall may help to find duodenal tumors that otherwise would escape attention. These tumors are often very small. Resection of lymph nodes that contain gastrinoma appears to have cured some patients of the syndrome, with serum gastrin concentrations returning to normal. In about a third of patients, the tumor is never identified at operation.

Total gastrectomy should be done in patients whose ulcer disease is refractory to medical therapy and who are unable to undergo resection of the tumor because of multiple hepatic metastases or the MEN I syndrome.

VIPoma (Verner-Morrison Syndrome; WDHA Syndrome: Watery Diarrhea, Hypokalemia, and Achlorhydria; Pancreatic Cholera)

Eighty to ninety percent of cases of WDHA syndrome are caused by an islet cell tumor of the pancreas that secretes VIP. The rest are associated with tumors in extrapancreatic sites such as the retroperitoneal sympathetic chain and the adrenal medulla.

Clinical Manifestations

The syndrome is characterized by a profuse watery diarrhea that contains large quantities of potassium, resulting hypokalemia (often 2 to 3 mEq/L), and profound weakness. Some gastric acid secretion is present in about half the patients; in the rest it is absent. Stool volume can be as great as 10 L/day but averages about 5 L/day during acute episodes. It contains over 300 mEq potassium (20 to 25 times normal) and resembles dilute tea. The diarrhea is “secretory” in nature; it persists during fasting and does not respond to antidiarrheal agents. Severe metabolic acidosis is common from loss of bicarbonate in the stool. Many patients are hypercalcemic, perhaps because the tumor also secretes a parathormone-like substance. Abnormal glucose tolerance may be the result of hypokalemia and altered sensitivity to insulin. Flushing is a prominent feature in about 15 percent of patients, usually in association with a bout of diarrhea. Up to 50 percent of the pancreatic tumors are malignant, and 75 percent of those have already metastasized by the time of diagnosis.

Diagnostic Studies

Patients with severe diarrhea must be studied carefully before attributing the cause to VIPoma. Chronic laxative abuse is the most common explanation. The specific diagnosis requires the demonstration of elevated fasting plasma VIP concentration by radioimmunoassay; plasma VIP levels are normal in other diarrheal states. Plasma PP and prostaglandin levels also should be measured, since some people with the syndrome have had elevations in the level of these substances, while VIP levels were normal. CT scan and then angiography should be used to localize the tumor. Transhepatic portal and splenic venous sampling also may be helpful.

Treatment

The immediate goal of management is the correction of dehydration and electrolyte abnormalities, especially hypokalemia. Once the diagnosis is suspected, a trial of somatostatin is indicated. This should control the diarrhea during the preoperative preparation and performance of the diagnostic studies. At operation, the tumor should be resected, and if metastases are present, they should be debulked. Plasma VIP concentrations should return to normal if all the tumor has been removed. If no pancreatic tumor is identified, the adrenal glands and sympathetic chains should be explored. If no tumor is found anywhere, a distal 80 percent pancreatectomy should be performed. In some patients with diffuse islet cell hyperplasia, symptoms have improved with this strategy.

In patients with inoperable or metastatic malignant tumors in whom complete resection is not possible, streptozocin alone or in combination with 5-fluorouracil may be effective. Almost all tumors respond initially, and a few long-term remissions have been noted. Renal function must be monitored carefully when streptozocin is used; selective arterial administration of the drug is preferred when renal function is impaired. In patients for whom no definitive therapy is possible, symptomatic relief has been reported with several agents, including high-dose steroids, trifluoperazine, and somatostatin. The tumors appear to be radioresistant.

Patients with benign tumors are cured by surgical excision. The average survival rate for patients with malignant tumors is approximately 1 year.

Glucagonoma

These tumors arise from a2cells in the islets. They produce a distinctive syndrome consisting of a migratory necrolytic dermatitis usually involving the legs and perineum, weight loss, stomatitis, hypoaminoacidemia, anemia, and mild diabetes mellitus. Visual scotomata may occur. An increased tendency to venous thrombosis and pulmonary emboli also has been reported. The condition is more common in women, and the age range is 20 to 70 years. The diagnosis is confirmed by demonstrating elevated plasma glucagon levels. The tumor should be localized with CT scans and arteriography, if necessary. About 25 percent of glucagonomas are benign. The majority are malignant and have metastasized by the time of diagnosis, however. The most common sites are the liver, lymph nodes, adrenal gland, or vertebrae. A few cases of the syndrome have been due to islet cell hyperplasia.

Treatment

Surgical excision of the primary tumor and debulking of metastases are indicated. Oral zinc supplements may improve the dermatitis. Somatostatin has been reported to return serum glucagon and amino acid levels to normal, clear the rash, and promote weight gain. Streptozocin and dacarbazine are the most effective chemotherapeutic agents.

Somatostatinoma

Somatostatin-producing tumors of the islet cells are characterized by mild diabetes mellitus, malabsorption and diarrhea, and dilatation of the gallbladder, often with cholelithiasis. Weight loss also may be prominent. Most of the tumors are located in the head of the pancreas and are malignant, with metastases to the liver at the time of exploration. CT scan usually shows a mass in the pancreas, with evidence of spread to the liver. The plasma levels of somatostatin are elevated, confirming the diagnosis.

The clinical manifestations are the result of the physiologic actions of the peptide. Inhibition of insulin release produces the diabetic state. Because glucagon release is also suppressed, the diabetes is easy to control. Cholelithiasis results from inhibition of gallbladder motility. Decreased biliary and pancreatic secretions probably cause the malabsorption.

Surgery may be indicated if the disease appears to be localized. More often, chemotherapy with streptozocin, dacarbazine, doxorubicin, etc. is the only treatment possible.

Miscellaneous Neoplasms

Because the pancreatic islets are part of the APUD (amine precursor uptake and decarboxylation) system, they retain the potential to produce a variety of neoplasms with multiple secretory products. For example, MEN I syndrome (Werner's syndrome) is a syndrome complex associated with tumors or hyperplasia in two or more of the endocrine glands. It involves the pancreas (81 percent), parathyroid (87 percent), pituitary (65 percent), and less frequently, the adrenal cortex (38 percent) and thyroid glands (19 percent). When the pancreas is involved, it usually becomes evident clinically later than does the parathyroid disorder (hyperparathyroidism is diagnosed in most patients before 40 years of age). In the pancreas, gastrinomas are most common, followed by insulinomas. Less often, the pancreatic tumors may secrete VIP, glucagon, pancreatic polypeptide, calcitonin, or 5-hydroxytryptophan (carcinoid-islet cell tumor). MEN I syndrome is present in about 25 percent of patients with gastrinoma and about 4 percent of all patients with insulinoma. Whatever the predominant cell type, the tumors are usually multicentric.

TRAUMA

Because of its protected retroperitoneal position, pancreatic injury only occurs with deep penetrating wounds or blunt trauma of significant force. Injury to the pancreas must be considered, however, in all patients who sustain injury to the upper abdomen. The diagnostic approach varies with the mechanism of injury. With penetrating trauma, almost all patients are operated on, and the diagnosis of pancreatic injury is made at the time of exploration. The diagnosis of blunt injury requires a high degree of clinical suspicion, since there may be few symptoms during the initial phases of management. In the first several days, the majority of patients will eventually develop epigastric and/or back pain, nausea, vomiting, ileus, and tenderness to abdominal palpation. Although the serum amylase concentration may be normal early, it eventually becomes elevated in 80 percent of patients. An elevated amylase level by itself is generally not an indication for surgical exploration, however. The decision to operate should be supported by evidence from physical examination or CT scan. CT may show fluid in the area of the pancreas or pancreatic ductal or parenchymal disruption. Because significant pancreatic injury may be missed by CT, it should be used primarily in patients in whom there is suspicion of injury, and operation is planned only if test results are positive. Diagnostic peritoneal lavage may be valuable, since many patients with pancreatic injury also have associated visceral injury that will result in a positive lavage.

Minor contusions and lacerations of the pancreas without evidence of ductal injury require only the control of hemorrhage and external drainage. Deeper lacerations or crush injuries with ductal injury and/or devitalization of the gland are best managed by distal resection of the injured gland and external drainage. Injuries to the head of the gland with significant damage to the duodenum and adjacent structures occasionally require a pancreaticoduodenectomy. However, even the majority of injuries to this area can be managed more conservatively.

Pancreatic trauma is associated with significant morbidity (e.g., pancreatic fistula in 10 percent and infection) and a mortality rate of about 20 percent. Most deaths are due to massive hemorrhage from erosion of adjacent vessels.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 31 - Spleen

Seymour I. Schwartz

ANATOMY

The spleen arises by mesenchymal differentiation along the left side of the dorsal mesogastrium in the 8-mm embryo. The weight of the spleen in the healthy adult ranges between 75 and 100 g, decreasing somewhat with age. The organ is located in the left upper quadrant, having a superior relationship to the undersurface of the left leaf of the diaphragm and protected anteriorly, laterally, and posteriorly by the lower portion of the rib cage. Its position is maintained by several suspensory ligaments, the major ones being the splenophrenic, splenorenal, splenocolic, and gastrosplenic ligaments. The gastrosplenic ligament normally contains the short gastric vessels, while the remaining ligaments are generally avascular, except in patients with portal hypertension, when collateral veins become apparent. Arterial blood enters the spleen via the splenic artery, a branch of the celiac artery. The major venous drainage courses through the splenic vein, which joins the superior mesenteric vein to form the portal vein.

Accessory spleens have been reported in 14 to 30 percent of patients, with a higher incidence occurring in patients operated on for hematologic disorders. These accessory organs, which receive their vascular supply from the splenic artery, are present, in decreasing order of frequency, in the hilus of the spleen, the gastrosplenic and splenocolic ligaments, the gastrocolic ligament, the splenorenal ligament, and the greater omentum (Fig. [31-1](#)). They also may occur in the pelvis of the female, and functioning splenic tissue has been removed from the scrotum in juxtaposition to the left testicle.

The spleen consists of a capsule that is normally 1 to 2 mm thick and trabeculae that enclose the pulp. The pulp itself has conventionally been divided into three zones: white, marginal, and red. Peripheral to the white pulp is the marginal zone, which contains end arteries arising from the central artery and from the penicillary arteries. The marginal zone contains lymphocytes, macrophages, and some red cells that have exited the terminal arteries. The marginal zone contains the marginal sinus, which filters material from the white pulp. Locally produced immunoglobulins enter this sinus and course to the peripheral circulation. Peripheral to the marginal zone is the red pulp, which consists of cords and sinuses that contain cellular elements of blood in transit.

Blood brought to the spleen via the splenic arteries courses through branches, the trabecular arteries, that leave the trabeculae and enter the white pulp as central arteries (Fig. [31-2](#)). These central arteries give off at right angles numerous arterioles, some of which terminate in the white pulp. The perpendicularity contributes to a skimming effect by which plasma exits while most red cells pass to the red pulp. Other branches cross the white pulp and end in the marginal zone or in the red pulp itself. The branch of the central artery that terminates in the red pulp, known as the "artery of the pulp," breaks up into many branches. Within the red pulp, the blood is collected in splenic sinuses. These large, thin-walled venous spaces drain into the pulp veins, which in turn drain into the trabecular veins and then into the main splenic veins to enter the portal circulation. Thus splenic pulp pressure reflects pressure throughout the portal venous system. The tissue between the splenic sinuses is a reticular, connective tissue meshwork that appears as cords on histologic section, and hence is designated the "splenic cords." At points of passage from cords to sinuses, deformability and flexibility are demanded of the red blood cells so that they can squeeze through. Although these fenestrations are of small diameter (0.5 to 5.0 mm), they are traversed by normal red cells, which easily adjust to these dimensions. Under normal conditions, 10 percent of red cells pass from terminal arterioles through arteriovenous connections; 90 percent of cells course from terminal arterioles into pulp cords and into splenic sinuses after traversing the cordal-sinus wall apertures. The total splenic blood flow averages 300 mL/min.

PHYSIOLOGY AND PATHOPHYSIOLOGY

(Fig. [31-3](#))

Galen is credited with the phrase "The spleen is an organ full of mystery." During the fifth to eighth month of fetal life, the spleen contributes actively to the production of both red cells and white cells that enter the circulation. This function does not continue in the normal adult. The role of the spleen in the immunologic processes of the body is discussed in Chap. [31-3](#). The splenic function that is the focus of surgical attention is related to the organ's reticuloendothelial tissue, which contributes to the removal of cellular elements from the circulating blood. Normally, cells pass through the spleen rapidly, but in the presence of splenomegaly and other disease states, the flow patterns become circuitous, contain more obstacles, and result in pooling of cells within the cords.

Abnormal and aged erythrocytes, abnormal granulocytes, normal and abnormal platelets, and cellular debris may be cleared by the spleen, which apparently is capable of discriminating between these and normal cellular components. In the normal adult, the spleen is the most important site of selective erythrocyte sequestration, and during its 120-day life cycle, the red cell spends an estimated minimum of 2 days within the spleen.

The action of the spleen that results in the pathologic reduction of circulating cellular elements of blood has been attributed to two possible mechanisms: (1) excessive splenic destruction of cellular elements, and (2) splenic production of an antibody that results in the destruction of cells within the circulating blood. Overactivity of splenic function leading to accelerated removal of any or all of the circulating cellular elements of the blood with resultant anemia, leukopenia, or thrombocytopenia, alone or in combination, is referred to as *hypersplenism*.

The normal adult spleen contains about 25 mL of red blood cells, but relatively few of these are removed during a single passage through the organ. The spleen is capable of removing nuclear remnants (Howell-Jolly bodies) from circulating erythrocytes. The postsplenectomy blood smear is characterized by the presence of circulating erythrocytes with Howell-Jolly bodies and Pappenheimer bodies (siderotic granules that stain with Wright's stain) as a result of the loss of the pitting function of the spleen.

During the course of a day, approximately 20 mL of aged red blood cells are removed. Aging changes the biophysical properties of the red cell, making splenic entrapment of red cells that have circulated for 105 to 120 days more likely. Delay in splenic transit of aged or abnormal cells can lead to further cell injury because of the relatively hypoxic, acidotic, and substrate- (glucose-) deprived environment that is present in congested splenic red pulp cords. These environmental conditions lead to further physical and chemical deterioration of the erythrocyte, making it more susceptible to phagocytosis by splenic macrophages and reticuloendothelial cells or to intrasinusoidal disintegration. The central event in cytolysis may be the fall in cellular *adenosine triphosphate (ATP)* to very low levels and the loss of vital cellular functions dependent on ATP, such as sodium and calcium efflux, priming of glycolysis, and maintenance of membrane integrity. To what extent repeated passages through the spleen contribute to normal red cell aging is not certain, but evidence that the red cell loses surface area with aging, combined with the ability of the spleen to remove bits of red cell with surgical precision, raises this possibility. On the other hand, a normal red cell life span in splenectomized subjects suggests that red cell aging occurs independently of splenic presence and at approximately the normal rate. A variety of erythrocytes altered by intrinsic factors (membrane, hemoglobin, or enzymic abnormalities) or extrinsic factors (antibody and nonantibody injury) may be prematurely removed by the spleen. Severely damaged cells may be removed by the reticuloendothelial system at a variety of sites. Minimally altered erythrocytes may require the specific rigors of the splenic circulation for premature destruction and therefore may have normal or near-normal survival after splenic removal.

The neutrophil is removed from the circulation with a half-life of about 6 h. Although the role of the spleen in the destruction of neutrophils under normal conditions is not well quantified, in some hypersplenic states the spleen's role is augmented, with resulting neutropenia. This augmented removal can occur because of splenic enlargement and sequestration of granulocytes or because of enhanced splenic removal of altered granulocytes, as seen in immune neutropenias.

The platelet (thrombocyte), under normal circumstances, survives about 10 days in the circulation. One-third of the total platelet pool is normally sequestered in the spleen, but the role of the spleen in the final removal of normal platelets has not been precisely defined. With splenomegaly, a larger proportion of platelets is sequestered in the spleen (up to 80 percent), and this and accelerated platelet destruction in the spleen account for thrombocytopenia. Increased sequestration and destruction of platelets occur in the absence of splenomegaly in patients with *idiopathic thrombocytopenic purpura (ITP)* as a consequence of immunologic alteration of the platelets. In thrombotic thrombocytopenic purpura (TTP), platelets are sequestered in the spleen because of encroachment of the vascular lumen by increased subendothelial collagen.

Splenectomy results in an increase in platelets, which at times reach levels greater than 1,000,000/mm³. Postsplenectomy thrombocytosis is usually transient but may persist. This is particularly notable in congenital hemolytic states that do not respond well to splenectomy. In these circumstances, continued hemolysis in the absence of the splenic removal mechanism can lead to persistent, extreme thrombocytosis and intravenous thrombosis.

DIAGNOSTIC CONSIDERATIONS

Evaluation of Size

Normally the spleen is not palpable on abdominal examination, but the organ may be felt in about 2 percent of healthy adults. In healthy subjects, no significant dullness is elicited by percussion over the spleen either anteriorly or laterally. As the organ enlarges, dullness may be detected at the level of the ninth intercostal space in the left anterior axillary line, especially on expiration. Thereafter, it becomes first percussible and then palpable below the left costal margin. With increasing splenomegaly, notching may be palpable on the anteromedial surface, distinguishing the spleen from other abdominal masses.

Routine radiologic examination of the abdomen usually provides an accurate estimate of the size of the spleen. Although splenomegaly may be suggested by medial or caudad displacement of the stomach bubble, frequently accompanied by caudad displacement of air in the splenic flexure of the colon, the organ's outline can often be clearly demarcated in the left upper quadrant, corroborating enlargement. Computed tomography (CT) and magnetic resonance imaging (MRI) depict the spleen and define abnormalities in size, shape, and parenchymal pathology, i.e., cyst, abscess (Fig. [31-4](#)), or tumor. Radioisotopic scanning with ^{99m}Tc-sulfur colloid also defines the organ; vascular shunts that allow blood to bypass splenic phagocytes interfere with the procedure.

Evaluation of Function

The functional abnormality of hypersplenism may be manifested by a reduction in the number of red cells, neutrophils, or platelets in the peripheral blood. An increase in the rate of red cell destruction will always result in a compensatory rise in the rate of production unless disease of the marrow coexists. The hallmark of increased red cell turnover (hemolysis) is reticulocytosis in the absence of blood loss. Other tests such as measurement of the plasma bilirubin, haptoglobin, and hemoglobin levels are less sensitive and are dependent on the rate and sites of hemolysis as well as on liver function, and are therefore only of adjunctive value.

Reduced red blood cell survival in patients with hemolytic anemia may be more precisely demonstrated by measuring the disappearance of blood radioactivity after the patient's erythrocytes are labeled with Na₂⁵¹CrO₄. With this technique, the normal half-life of the red cells is about 30 days (i.e., 50 percent of the cells remain in the circulation at that time). A half-life of 15 days or less indicates significantly increased hemolysis. The spleen's role in a hemolytic anemia may be assessed by determining the relative uptakes of ⁵¹Cr-tagged erythrocytes by the spleen and the liver. A spleen-liver ratio greater than 2:1 indicates significant preferential splenic sequestration and anticipates a beneficial effect of splenectomy.

Radioisotopic labeling also has been used to evaluate the survival of neutrophils and platelets. There is no clinically practical method of assessing the decreased longevity of

neutrophils or of measuring the role of the spleen in the destruction of these cells. Platelet destruction in the spleen is such a common mechanism in thrombocytopenia that radioisotopic studies are not often done if the bone marrow aspirate indicates an abundance of megakaryocytes, suggesting adequate platelet production.

Rupture of the Spleen

(see Chap. 6)

HEMATOLOGIC DISORDERS FOR WHICH SPLENECTOMY IS THERAPEUTIC

In 1887 Sir Spencer Wells operated on a patient with the preoperative diagnosis of a uterine fibroid, but instead he noted a “wandering spleen,” which he removed. The patient later proved to have hereditary spherocytosis; thus the first surgical cure of a hematologic disorder was inadvertent and became manifest postoperatively when the patient's anemia and chronic jaundice disappeared. Micheli is generally given credit for introducing the concept of splenectomy as treatment for hemolytic anemia in an article published in 1911. Kaznelson, as a medical student in Prague, proposed to Schloffer, a professor of surgery, splenectomy to treat idiopathic thrombocytopenic purpura in a 36-year-old woman. The case, and the successful elevation of the platelet count to above-normal levels, was reported four weeks after the procedure in 1916.

Hemolytic Anemias

Hemolytic anemias include a broad spectrum of disorders in which there is accelerated destruction of mature red blood cells. The congenital anemias are due to an intrinsic abnormality of the erythrocytes, and the acquired anemias are related to an extracorpuscular factor acting on an intrinsically normal cell. In both types of disorder, the reduced red blood cell survival may be demonstrated by measuring the disappearance of the patient's radioactive erythrocytes (labeled with ⁵¹Cr), and the spleen's role may be evaluated by determining the relative uptakes of this radioactivity by the spleen and liver.

Hereditary Spherocytosis

Hereditary spherocytosis is transmitted as an autosomal dominant trait and is the most common of the symptomatic familial hemolytic anemias. Ovalocytosis and stomatocytosis usually exist as harmless traits, but occasionally, when these forms constitute 50 to 90 percent of the red cell population, clinical manifestations indistinguishable from those noted with hereditary spherocytosis may occur. The fundamental abnormality stems from a defective erythrocyte membrane that causes the cell to be smaller than normal, unusually thick, and almost spherical. These cells also demonstrate increased osmotic fragility, i.e., lysis occurs at a higher concentration of sodium chloride than normal. The role of the spleen in this disorder is related to the inability of the spherocytic cells to pass through the splenic pulp. The cells that escape from the spleen are more susceptible to trapping and disintegration during each successive passage, until cell loss ensues. The precise pathogenesis of the cell injury may be related to a decreased availability of red cell *ATP in the environment of the spleen, combined with a cell membrane that has been shown in in vitro studies to be more susceptible to reduction in ATP levels.*

The salient clinical features of the disease are anemia, reticulocytosis, jaundice, and splenomegaly. It is unusual for the anemia to be extremely severe, and the jaundice usually parallels the severity of the anemia. Periodic and sudden increases in the intensity of the anemia and jaundice may occur, and rare fatal crises have been reported. Cholelithiasis with gallstones of the pigmented variety has been reported in 30 to 60 percent of patients but is rare in children under the age of ten. Leg ulcers are uncommon specific manifestations.

Diagnosis is generally established by peripheral blood smear analysis, which demonstrates that more than 60 percent of the red blood cells are spherocyte-shaped with a mean diameter less than normal and a thickness greater than normal. Increased osmotic fragility of the red blood cells provides diagnostic confirmation, but this test is rarely performed.

Splenectomy is the sole therapy for hereditary spherocytosis and ovalocytosis. It is generally recommended that the operation be delayed until the fourth year of life. Intractable leg ulcers associated with hereditary spherocytosis mandate early splenectomy because they heal only after the spleen has been removed. The results of splenectomy as a method of correcting anemia and preventing hemolysis have been uniformly good. Although the inherent membrane abnormality persists and the spherocytosis and increased osmotic fragility of the cells are not altered, in vivo hemolysis virtually ceases, and following removal of the spleen, the erythrocytes achieve a normal life span and the jaundice, if present, disappears. It is appropriate to perform an oral cholecystogram or ultrasonogram prior to splenectomy, and the gallbladder should always be examined at the time of operation. If gallstones are present, the gallbladder should be removed during the operation.

Thalassemia

Thalassemia (Mediterranean anemia) is transmitted as a dominant trait and primarily derives from a defect in hemoglobin synthesis. The development of intracellular precipitates (Heinz bodies) contributes to premature red cell destruction. The disease is classified as alpha, beta, and gamma types, determined by the specific defect in synthesis rate of the peptide chain. In the United States, most thalassemia patients are of southern European origin and suffer from beta thalassemia, i.e., a quantitative reduction in the rate of beta chain synthesis, resulting in a decrease in hemoglobin A (Hb-A). Thalassemia occurs in two major degrees of severity: homozygous thalassemia (thalassemia major), a severe disorder in which the affected child receives a gene for thalassemia from each parent, and heterozygous thalassemia (thalassemia minor), a mild disorder in which the affected child receives a gene from only one parent. Gradations of thalassemia range from heterozygous thalassemia (minor), often not detected until examination of the blood for an unrelated problem, to homozygous thalassemia, a severe, chronic anemia, with icterus, splenomegaly, and death early in life. In thalassemia minor, Hb-A₂ is always increased, and slight increases in Hb-F occur in 50 percent of patients. In both types of thalassemia, the hemoglobin-deficient cells are small, thin, and misshapen. The cells appear washed out and have a characteristic resistance to osmotic lysis.

The clinical manifestations of thalassemia major usually occur in the first year of life and consist of pallor, retarded body growth, and enlargement of the head. Intractable leg ulcers may be noted, intercurrent infections are common, and gallstones are reported in about one-fourth of patients. The manifestations of thalassemia minor may vary. Most patients with thalassemia minor lead normal lives, but some patients have a more severe expression of their disease (referred to as thalassemia intermedia) and generally present with signs and symptoms attributable to mild anemia, chronic mild jaundice, and moderate splenomegaly.

The diagnosis of thalassemia major is established by the peripheral blood smear, which reveals hypochromic, microcytic anemia with markedly distorted red cells of various sizes and shapes. Nucleated red cells, or “target cells,” are invariably present, and the reticulocyte count is elevated, as is the white cell count. The characteristic feature of the disease is the persistence of Hb-F and reduction of Hb-A levels, demonstrated by the alkali denaturation study. It is important that both parents have evidence of thalassemia minor.

Treatment is indicated only for symptomatic patients. Transfusions are usually required at regular intervals, but as most patients accommodate to low hemoglobin levels, the transfusions should be directed at maintaining the hemoglobin level at 10 g/dL. Although splenectomy does not influence the basic hematologic disorder, it may reduce both the hemolytic process and the transfusion requirements. Marked splenomegaly and/or symptomatic repeated splenic infarction also constitute indications for splenectomy. A postoperative complication rate of 43 percent following splenectomy in thalassemia is contrasted with 3 percent in patients undergoing splenectomy for other reasons; overall, however, the benefit- to-risk ratio favors splenectomy when indications are present.

Hereditary Hemolytic Anemia with Enzyme Deficiency

Included in this category are (1) enzyme deficiencies in anaerobic glycolytic pathways, the prototype of which is pyruvate-kinase (PK) deficiency, and (2) enzyme deficiencies in the hexose monophosphate shunt, the prototype of which is glucose-6-phosphate (G-6-PD) deficiency. These deficiencies render the cells susceptible to increased hemolysis. Splenic enlargement occurs more frequently with PK deficiency, and the spleen is rarely enlarged in patients with G-6-PD deficiency. Specific enzyme assays are employed to define the deficiency.

The majority of patients maintain hemoglobins greater than 8 g/dL, are asymptomatic, and do not require therapy. With significant anemia, blood transfusions are indicated, and the transfused cells survive normally. ⁵¹Cr- tagged red cell studies are not predictive of results with a high enough degree of accuracy to be useful. In severe cases of PK deficiency, splenectomy may be worthwhile. In patients with this disorder postoperative thrombocytosis with consequent hepatic, portal, or inferior vena caval thrombosis may occur if the hemolytic rate is unabated. Splenectomy is not indicated for patients with G-6-PD deficiency.

Hereditary High Red Cell Phosphatidylcholine Anemia (HPCA)

When patients with HPCA have undergone splenectomy, the anemia worsens and the hemolysis is unchanged.

Sickle Cell Disease

Sickle cell anemia is a hereditary hemolytic anemia seen predominantly in blacks and characterized by the presence of sickle- and crescent-shaped erythrocytes. In this hereditary hemoglobinopathy, the normal Hb-A is replaced by the abnormal form of hemoglobin, sickle hemoglobin (Hb-S). Hb-F is usually mildly increased. Combinations of Hb-S with other hemoglobin variants also occur as a result of an abnormal trait inherited from each parent, e.g., Hb-S/Hb-C or Hb-S/thalassemia.

Under conditions of reduced oxygen tension, Hb-S molecules undergo crystallization within the red cell, which elongates and distorts the cell. The sickling phenomenon occurs more readily with higher percentages of Hb-S, with a reduced pH, and under conditions of circulatory stasis that tend to exaggerate hemoglobin deoxygenation. The sickle cells themselves contribute to increased blood viscosity and circulatory stasis, thus establishing a vicious cycle. The primary consequence of this stagnation is thrombosis, which leads to ischemia, necrosis, and organ fibrosis.

The role of the spleen in this disorder is not clear. Early in the course of the disease, splenomegaly occurs, but following varying intervals in most patients, the spleen undergoes infarction and marked contraction with eventual autosplenectomy.

Although the sickle cell trait occurs in approximately 9 percent of the black population, the majority are asymptomatic. However, sickle cell anemia has been observed in 0.3 to 1.3 percent of blacks, who often show remarkable adaptation to the state of chronic anemia and jaundice. This adaptive state may be interrupted at intervals by acute symptoms or crises that are related to vascular occlusion. Depending on the vessels involved, the patient may have bone or joint pain, hematuria, priapism, neurologic manifestations, or ulcers over the malleoli. Abdominal pain and cramps due to visceral stasis are frequent, simulating an acute surgical abdomen. Thrombosis of the splenic vessels may result in the unusual complication of splenic abscess, manifested by splenomegaly, splenic pain, and fever. Many patients with sickle cell anemia die in the first decade of life, but a few survive to the fifth decade. Death may be the result of intercurrent infections or cardiac or renal failure.

The diagnosis is established by the presence of anemia, characteristic sickle cells on peripheral blood smear analysis, hemoglobin electrophoresis showing 80 percent or more Hb-S, and the presence of the trait in both parents. Leukocytosis is often noted, and the platelet count is frequently elevated. There may be modest elevation of the serum bilirubin, and cholelithiasis is a frequent accompaniment.

For most patients only palliative treatment is possible. Sodium cyanate will prevent sickling of Hb-S. Transfusions may be required to maintain adequate hemoglobin levels. Adequate hydration and partial exchange transfusion may help during a crisis. Splenectomy may benefit patients with marked hypersplenism or splenic abscesses. In the circumstance of splenic abscess, incision and drainage of the abscess cavity within the parenchyma of the spleen may be necessary, since removal of the organ is hindered by marked inflammatory and adhesive processes. Splenectomy may be of benefit in a very few patients in whom acute splenic sequestration of red cells can be demonstrated, although the operation does not affect the sickling process. A report on forty-six children who underwent splenectomy for hypersplenism and fourteen patients operated on for acute splenic sequestration with a median follow-up of 6 years noted that two patients died of overwhelming sepsis; neither received prophylaxis, and neither pneumococcal septicemia nor meningitis was confirmed.

Idiopathic Autoimmune Hemolytic Anemia

This is a disorder in which the life span of a presumably normal erythrocyte is shortened when exposed to an endogenous hemolytic mechanism. The causation has not been defined, but an autoimmune process appears to be fundamental. In such patients, antibodies reacting with the patient's normal red cells have been defined, and there is evidence that the spleen may serve as a source of antibody. Both "warm" and "cold" antibodies have been described. Some "warm" antibodies have Rh specificity. Most of these antibodies are hemagglutinins rather than hemolysins. It is believed that the reticuloendothelial system traps and destroys the immunologically altered cells. Sequestration occurs primarily in the spleen. By binding the Fc portion of the IgG molecule to the corresponding macrophage surface Fc receptor, the spherocytes become more rigid and more sensitive to destruction in the splenic circulation.

Although autoimmune hemolytic anemia may be encountered at any age, it occurs more frequently after the age of fifty and twice as often in females. Mild jaundice is often present. The spleen is palpably enlarged in half the cases, and gallstones have been demonstrated in a quarter of the cases. The extent of anemia varies, and hemoglobinuria and tubular necrosis have been reported in severe cases. In this circumstance, the prognosis is serious, as the mortality rate is 40 to 50 percent.

The diagnosis of hemolysis is made by demonstrating anemia and reticulocytosis accompanied by the products of red cell destruction in the blood, urine, and stool. The bone marrow is hypercellular with a predominance of erythroid precursors. A distinguishing feature of the disease is the demonstration by direct Coombs' test of an autoantibody on the patient's red cells.

In some patients the disorder tends to run an acute, self-limited course, and no treatment is necessary. If the anemia becomes severe, corticosteroids or blood transfusions may be required. In the "warm" antibody immune hemolytic anemias, splenectomy should be considered (1) if steroids have been ineffective, (2) if excessive doses of steroids are required to maintain remission, (3) if toxic manifestations of steroids become apparent, and (4) if steroids are contraindicated for other reasons. Excessive splenic sequestration of ⁵¹Cr-tagged red cells offers a guide for the selection of patients who may respond to splenectomy. A favorable response is to be anticipated in about 80 percent of selected splenectomized patients. Patients without prominent spleen-liver sequestration ratios may respond to splenectomy. Late relapses may occur in splenectomized patients.

Idiopathic Thrombocytopenic Purpura

Idiopathic thrombocytopenic purpura (immune thrombocytopenic purpura) (ITP) is an acquired disorder caused by the destruction of platelets exposed to circulating IgG antiplatelet factors. The spleen is the source of these factors. It is also the major site for sequestering sensitized platelets. The term ITP should be reserved for a hemorrhagic disorder characterized by a subnormal platelet count in the presence of bone marrow containing normal or increased megakaryocytes and in the absence of any systemic disease or history of ingestion of drugs capable of inducing thrombocytopenia.

Female patients outnumber males at a ratio of 3:1. The most common presenting signs are petechiae and/or ecchymoses. ITP occurs in some patients who are HIV positive or who have acquired immunodeficiency syndrome (AIDS). In the majority of patients these signs are accompanied by one or several other symptoms, including, in order of frequency, bleeding gums, vaginal bleeding, gastrointestinal bleeding, and hematuria. In some patients the clinical manifestations take an almost cyclic course with exacerbations occurring at the time of menses. The incidence of central nervous system bleeding ranges between 1 and 2 percent and usually occurs early in the course of the disease. The spleen is rarely palpable, and its enlargement should evoke suspicion of the presence of another disease causing thrombocytopenia.

The characteristic laboratory findings include a platelet count generally reduced to 50,000/mm³ or less and in some patients approaching zero. Associated with reduced platelet count, the bleeding time may be prolonged, but the clotting time remains normal. There is usually no significant anemia or leukopenia. When ITP is accompanied by an autoimmune hemolytic anemia it is known as Evans's syndrome. Platelet survival time following the transfusion of ⁵¹Cr-labeled normal platelets is short, but this test is not necessary for establishing the diagnosis of ITP. Bone marrow examination reveals either a normal or an elevated megakaryocyte count with or without a relative increase in small forms. Qualitative changes in the megakaryocytes are characterized by degranulation of the cytoplasm, rounding of cytoplasmic edges, the disappearance of the usual pseudopodia containing granule-free platelets, and a varying degree of vacuolization of the cytoplasm.

Acute ITP has an excellent prognosis in children under the age of sixteen; approximately 80 percent of these patients will make a complete and permanent recovery without specific therapy. Much of the discussion regarding therapy for ITP centers on the relative values and disadvantages of steroid therapy and splenectomy. In most series, the results achieved by splenectomy are significantly more impressive than are the responses to steroids (Fig. 31-5). Between 75 and 85 percent of the total number of patients subjected to splenectomy respond permanently and require no further steroid therapy. The same rates are reported for patients with systemic lupus erythematosus and AIDS. No reliable predictive factors for satisfactory response to splenectomy have been uncovered. Opinions differ regarding a correlation between an initial response to steroid therapy and the efficacy of splenectomy. In most patients the platelet count rises to over 100,000/mm³ in 7 days. Rarely, return to normal levels may take months. Even in patients in whom the platelet count does not return to normal levels, recurrent petechiae or bleeding episodes are rare. A long-term follow-up of splenectomized patients does not support the contention that disseminated lupus erythematosus develops after this procedure.

The generally accepted protocol for managing patients with diagnosed ITP includes an initial 6-week to 2-month period of steroid therapy. Infusions of high doses of intravenous g-globulin and plasmapheresis have been used with limited permanent success. If the patient does not respond with elevation of the platelet count, splenectomy is performed. If the patient does respond, the steroid therapy is tapered off, and if thrombocytopenia recurs, splenectomy is carried out. The same criteria pertain to patients with AIDS or HIV positivity. Any manifestation suggestive of intracranial bleeding demands emergency splenectomy. In one series, five of six patients with ITP and life-threatening intracranial bleeding were saved by splenectomy.

For patients with platelet counts approaching zero, platelet packs should be available for the operative procedure but should not be administered preoperatively. Platelet transfusion is reserved for patients who continue bleeding after removal of the spleen. Occasional patients in whom signs of the disease recurred months or years following splenectomy have achieved permanent cure following removal of an accessory spleen, the presence of which may be defined by technetium scan.

Systemic Lupus Erythematosus (SLE)

Refractory cytopenias associated with SLE may respond to splenectomy. The platelet count returns to normal in about two-thirds of patients, the hematocrit level increases 20 percent or more, and the white cell count returns to normal in almost all patients with SLE following splenectomy.

Thrombotic Thrombocytopenic Purpura

Although thrombotic thrombocytopenic purpura (TTP) is a disease of arterioles and capillaries, there are significant accompanying hematologic changes for which the response to splenectomy may be striking. The causation has not been precisely defined, but immune mechanisms have been suggested. Approximately 5 percent of reported cases occurred during pregnancy. TTP in pregnancy must be distinguished from an idiopathic thrombocytopenia that may develop in the third trimester because this state is reversed by termination of the pregnancy. Histologically there is widespread occlusion of multiple arterioles and capillaries by hyaline membranes, with minimal inflammatory change and limited infarction.

The pentad of clinical features of virtually all cases of TTP consists of fever, purpura, hemolytic anemia, neurologic manifestations, and signs of renal disease. The pertinent laboratory findings include anemia with reticulocytosis, thrombocytopenia, and leukocytosis, sometimes accompanied by an elevated serum bilirubin level, proteinuria, hematuria, casts, or azotemia. The peripheral blood smear reveals pleomorphic, normochromic red cells that are fragmented and distorted. The degree of thrombocytopenia varies during the course of the illness, but a profound decrease in platelet count often develops within hours of onset. The bone marrow usually reveals erythroid and myeloid hyperplasia with a normal or increased number of megakaryocytes.

In the majority of cases, the disease has a rapid onset, fulminant course, and fatal outcome usually due to intracerebral hemorrhage or renal failure. Repeated plasmapheresis usually reverses the process, but in occasional cases in which no response is effected, splenectomy coupled with high-dosage steroids has resulted in cure.

Secondary Hypersplenism

Pancytopenia, thrombocytopenia, and leukopenia or anemia may occur whenever there is splenomegaly or splenic congestion. Splenomegaly with engorgement of the vascular spaces accompanies portal hypertension, resulting in accelerated destruction of the circulating cells within the spleen. The clinical manifestations of thrombocytopenia, i.e., petechiae and spontaneous bleeding, are extremely uncommon in patients with portal hypertension. As no correlation exists between the degree of anemia, leukopenia, or thrombocytopenia and the long-term survival of patients with cirrhosis, hypersplenism per se is not an indication for splenectomy in a patient with portal hypertension. Percutaneous transfemoral embolization of the splenic arterial circulation has been performed in some of these patients. Although splenic hyperfunction was controlled, the development of painful infarction of the spleen and septic splenitis suggests that such an approach has limited value. The more common surgical situation entails the patient with bleeding esophagogastric varices accompanied by significant secondary thrombocytopenia. In the experience of Child and Turcotte with more than 300 shunts, it was necessary to perform only one splenectomy for persistent, clinically significant hypersplenism. Thrombocytopenia is generally improved by a portacaval shunt, and in the few instances in which the platelet count does not return to normal, complications are rarely attributable to thrombocytopenia. Splenectomy alone should not be performed in a patient with portal hypertension and secondary hypersplenism as there is no long-term effect on the elevated portal pressure. In the rare circumstance that splenectomy is required for a severe degree of hypersplenism in such a patient, it should be combined with a splenorenal shunt to decompress the portal circulation.

Myeloproliferative Disorders

Myeloid metaplasia is a panproliferative process manifested by increased connective tissue proliferation of the bone marrow, liver, spleen, and lymph nodes and simultaneous proliferation of hemopoietic elements in the liver, spleen, and long bones. The disease is closely related to polycythemia vera, myelogenous leukemia, and idiopathic thrombocytosis. The causation remains unclear. The spleen may be markedly enlarged, and portal hypertension has been described in some patients, due either to hepatic fibrosis of sufficient degree to be obstructive to the portal circulation or to increased forward blood flow through the splenoportal system in the absence of hepatic involvement.

Clinical manifestations generally become apparent in middle-aged and older adults. The presenting symptoms usually are related to anemia and increasing splenomegaly. Symptoms related to the spleen include the intermittent pain of splenic infarction, generalized abdominal discomfort, and a feeling of fullness after meals. Other symptoms include

spontaneous bleeding, secondary infection, bone pain, pruritus, hypermetabolism, and complications associated with hyperuricemia. The most common physical findings are pallor and splenomegaly. Hepatomegaly is present in about three-quarters of the patients.

The laboratory hallmark of myeloid metaplasia is in the peripheral blood smear. The red cells are characterized by fragmentation and immature forms and poikilocytosis with numerous teardrop and elongated shapes. Characteristically, the patients have anemia of the normochromic type. The white blood cell count is under 50,000 mm⁻³ in the majority of the patients but may reach extremely high levels. Immature myeloid cells are found in the peripheral smear. The platelet counts are normal in about one-quarter of the patients, thrombocytopenia is present in about one-third, and marked thrombocytopenia is present in 5 percent. Thrombocytosis with a platelet count over 1,000,000 mm⁻³ is observed in one-quarter of the patients. The platelets are frequently enlarged and bizarre in appearance. The leukocyte alkaline phosphatase is usually high, and hyperuricemia is frequently present. Radiographs of the bone demonstrate increased density in approximately 50 percent of the patients, particularly in the pelvic region. Marrow biopsy sections show varying degrees of bone marrow replacement by fibrous tissue interposed with small foci of megakaryocytes, erythropoiesis, and myeloid cells.

Treatment is generally directed at the anemia and splenomegaly. It usually consists of transfusions, hormones, chemotherapy, and radiotherapy. Male hormone preparations may be of value in patients with anemia due to marrow failure. Alkylating agents may be effective in reducing splenic size and transfusion requirements as well as for patients whose predominant clinical problem is hypermetabolism. Busulfan is the most commonly used alkylating agent, but cyclophosphamide may also be used in thrombocytopenic subjects since it is less likely to suppress platelet production. Since patients with myelofibrosis are very sensitive to chemotherapy, such agents must be used very cautiously.

Although splenectomy does not alter the general course of the disorder, the procedure is indicated for control of anemia and thrombocytopenia, and for symptoms attributable to splenomegaly. Prolonged partial thromboplastin time, abnormal prothrombin consumption, prothrombin time, and bleeding time are usually normalized by splenectomy. Thrombocytopenia associated with sufficiently reduced megakaryocytes to contraindicate chemotherapy is also a frequent indication for splenectomy, as is the large spleen that causes digestive difficulties or is symptomatic because of multiple infarctions despite chemotherapy and/or local irradiation. In patients with esophagogastric varices portal pressures should be determined before and after splenectomy. In most instances splenectomy alone will effect significant reduction in pressure, causing varices to decrease in size or disappear and obviating the need for a concomitant splenorenal or portacaval shunt.

The old concept that splenectomy resulted in the removal of a significant hemopoietic element has been disproved. Mortality and morbidity rates for patients with myeloid metaplasia undergoing splenectomy are higher than those reported for other hematologic disorders. Splenectomy in patients with myeloid metaplasia is associated with an operative mortality of 13 percent, a morbidity rate of nearly 45 percent, and a late morbidity rate of 16 percent. Postoperative thrombocytosis and/or thrombosis of the splenic vein extending into the portal and mesenteric vein occurs more commonly in these patients. The complication is characterized by intractable ascites, hepatic failure, and renal failure that is often not noted until a week after the operation. The incidence of this complication can be reduced by correction of a thrombocytotic state, if present, preoperatively using alkylating agents and by the use of drugs to prevent platelet aggregation and clotting during the perioperative period.

Hodgkin's Disease, Lymphomas, and Leukemias

Chemotherapy and/or radiation therapy is the standard approach to treatment of these disorders.

In Hodgkin's disease, non-Hodgkin's lymphoma, and chronic lymphocytic leukemia (CLL), splenectomy has improved cytopenia in over 75 percent of cases. In these patients, palliative splenectomy should be performed before the platelet count is excessively low or there are marked clinical abnormalities. Splenectomy may allow patients to begin or maintain chemotherapy. Palliative splenectomy may be indicated for symptomatic splenomegaly.

Hairy cell leukemia, or reticuloendotheliosis, is characterized by malignant cells with filamentous cytoplasmic projections. Patients without symptomatic splenomegaly and few “hairy” cells in the circulating blood do well without any treatment. Patients who live 4 years after the diagnosis has been established have a favorable long-term prognosis and do not require splenectomy. When hairy cell leukemia is accompanied by problems caused by neutropenia, thrombocytopenia, and anemia, splenectomy is effective therapy. A complete response occurs in two-thirds to three-quarters of patients, and a partial response is effected in another 20 percent. The 5- year survival rate has been reported as 61 to 76 percent. The response to splenectomy is unrelated to the weight of the spleen. Survival is not related to the hematologic response to splenectomy. Failures are managed with steroids and chemotherapeutic agents.

Staging of Hodgkin's Disease and Non-Hodgkin's Lymphoma

The diagnosis of Hodgkin's disease is generally established by histologic evaluation of a clinically suspect area of lymphadenopathy or splenomegaly. Demonstration of the typical, large, multinuclear cell, the Sternberg-Reed cell, is regarded as essential for the diagnosis. These cells, however, do not form the bulk of the tumor. Four major histologic types have been defined: lymphocyte predominance, nodular sclerosis, mixed cellularity, and lymphocyte depletion. Survival with Hodgkin's disease is related in part to the histologic type and also to the distribution of disease and the presence or absence of specific symptoms. Stage I disease is defined as limited to one anatomic region; Stage II disease is limited to two or more contiguous or noncontiguous regions on the same side of the diaphragm; Stage III disease refers to disease on both sides of the diaphragm with involvement limited to lymph nodes, spleen, and Waldeyer's ring; and Stage IV refers to involvement of the bone marrow, lung, liver, skin, gastrointestinal tract, and any organ or tissue other than the lymph nodes or Waldeyer's ring.

The indications for surgical staging of Hodgkin's disease have decreased significantly in recent years because of greater reliance on CT scans and the more liberal use of chemotherapy, which negates the need to determine the presence of infradiaphragmatic involvement. Currently the indications for surgical staging focus on Stage I patients in whom the disease is apparently limited to one anatomic region, and on Stage II patients in whom the disease is detected in two regions above the diaphragm and the pathology demonstrates nodular sclerosis.

The staging procedure begins with obtaining a wedge biopsy specimen of the liver before retractors are applied and cause confusing white blood cell migration. Splenectomy is then carried out, followed by removal of representative retroperitoneal, mesenteric, and hepatoduodenal nodes. An iliac marrow biopsy procedure is usually included. Surgical staging was found to upgrade the clinical stage in 27 to 36 percent of cases and to decrease it in 7 to 15 percent, for a total alteration of 42 percent. Current consensus is that surgical staging is not indicated for patients with non- Hodgkin's lymphoma. In these patients the combination of CT scans, marrow biopsy, and laparoscopically directed nodal and liver biopsies offers a reasonable alternative to diagnostic celiotomy.

Miscellaneous Diseases

Felty's Syndrome

The triad of rheumatoid arthritis, splenomegaly, and neutropenia is referred to as Felty's syndrome. Mild anemia and/or thrombocytopenia has been noted in some cases, and gastric achlorhydria is common. An antibody specifically directed against neutrophil nuclei is nearly always demonstrable by fluorescent stains. Corticosteroids and splenectomy have been used to reverse the neutropenia in order to reduce susceptibility to infection. The response to steroids is usually not long-lasting, but the hematologic effects of splenectomy generally are excellent. Splenectomy in these patients should be reserved for neutropenic patients who have serious or recurrent infections, patients who require transfusions for anemia, patients with profound thrombocytopenia, and patients with intractable leg ulcers. There is a sharp rise in the total number of leukocytes in the first 24 h, reaching a peak at about the third postoperative day. Although relative neutropenia may persist, the neutrophilic response to infection in the postsplenectomy state becomes normal. The clinical course of the arthritis is rarely altered.

Sarcoidosis

This disease affects young adults. There are few constitutional symptoms, and fever is unusual, although night sweats have been noted. Cough and shortness of breath may attend mediastinal or pulmonary involvement. Skin lesions appear in about 50 percent of patients, and generalized lymphadenopathy is frequent. Involvement of the liver and spleen may produce hepatomegaly and splenomegaly in about 25 percent of patients. About 20 percent of the patients with splenomegaly develop manifestations of hypersplenism, particularly thrombocytopenic purpura. Hemolytic anemia, neutropenia, pancytopenia, and spontaneous splenic rupture have all been observed.

There is no specific treatment, and spontaneous recovery can be anticipated in the majority of cases. Splenectomy should be considered for patients with splenomegaly when there are complications of hypersplenism, since the operation has been almost uniformly followed by correction of the hematologic abnormality.

Gaucher's Disease

This is a familial disorder characterized by abnormal storage or retention of glycolipid cerebroside in reticuloendothelial cells. Proliferation and enlargement of these cells produce enlargement of the spleen, liver, and lymph nodes. The disease is generally discovered in childhood but may become evident either early in infancy or late in adult life.

The sole clinical manifestation may be awareness of a progressively enlarging abdominal mass, primarily due to splenomegaly and, to a lesser extent, to hepatomegaly. Yellowish-brown pigmentation of the head and extremities occurs in 45 to 75 percent of cases. Bone pain and pathologic fracture may develop in long-standing cases. Many patients develop the hematologic manifestations of hypersplenism as a result of excessive sequestration of formed blood elements. Moderate to severe thrombocytopenia and normocytic anemia are almost always present, and often there is mild leukopenia. In the patients with hypersplenism, splenectomy almost uniformly has been beneficial in correcting the hematologic disorder, but there is no evidence that the operation influences the course of the basic disease. Partial splenectomy has been performed in children for symptomatic splenomegaly and hypersplenism to obviate development of overwhelming postsplenectomy infection (OPSI); 400 to 3800 g of tissue has been removed without complication. There has been no OPSI and no postoperative increase in accumulation of beta- glucocerebroside in the liver or bones. All children had an improved growth rate and hematologic picture.

Erythropoietic Porphyria

This is a congenital disorder of erythrocyte pyrrole metabolism that is transmitted as a recessive trait and characterized by the excessive deposition of porphyrins in the tissues. In the skin this results in pronounced photosensitization and severe bullous dermatitis. Premature red cell destruction within the spleen contributes to severe anemia. When the disease is complicated by hemolysis or splenomegaly, splenectomy is followed by marked improvement in the anemia and decreased concentrations of porphyrins in the red cells, bone marrow, and urine.

Systemic Mast Cell Disease (SMCD)

This disorder is manifested by urticaria pigmentosa and mast cell infiltration of the skin, bone marrow, or gastrointestinal tract. In patients with aggressive disease, including lymphadenopathic mastocytosis and eosinophilia and severe thrombocytopenia, splenectomy may be beneficial, extending survival and improving the ability to tolerate chemotherapy.

OTHER LESIONS

Ectopic Spleen

This unusual condition, also known as wandering spleen, is ascribed to lengthening of the splenic ligaments, which results in extreme mobility of the organ so that the spleen of normal size may be palpable in the lower abdomen or in the pelvis. In some cases, acute torsion of the pedicle occurs, necessitating surgical intervention. Adult women are more frequently affected but males predominate in the thirty-five recorded pediatric cases. Elective splenopexy, wrapping the spleen in Dexon mesh and tacking it to the diaphragm and retroperitoneum in its normal position, is indicated to preserve the organ and prevent acute torsion.

Cysts and Tumors

Cysts of the spleen are unusual. Parasitic cysts are usually due to echinococcal involvement, while the nonparasitic cysts may be categorized as dermoid, epidermoid, epithelial, and pseudocysts. Pseudocysts occur after occult rupture of the spleen caused by trauma. Symptomatic parasitic cysts of the spleen are best managed by splenectomy. Large symptomatic nonparasitic cysts may be widely unroofed to assure permanent cure. This can be performed laparoscopically. Asymptomatic pseudocysts do not require operation but should be followed ultrasonographically to rule out significant expansion.

Primary malignant tumors of the spleen are usually sarcomatous. Primary splenic non-Hodgkin's lymphoma is a rarity which has an excellent prognosis following splenectomy. Autopsy series have refuted the concept that metastases to the spleen are rare. Exploration for undiagnosed splenomegaly rarely reveals unsuspected metastatic deposit in the absence of known generalized metastases but may establish the diagnosis of a hematologic malignancy.

Abscesses

Splenic abscess is an uncommon cause of abdominal sepsis. Primary splenic abscesses occur much more often in the tropics, where they are frequently related to thrombosis of the splenic vessels with infarction in patients with sickle cell anemia. Pyogenic splenic abscesses have been reported with increasing frequency in intravenous drug abusers. Clinical manifestations include fever, chills, splenomegaly, and left upper quadrant tenderness. Diagnosis may be established by ultrasound or CT scan (Fig. 31-6). Removal of the spleen is the operation of choice, but some patients have been treated with splenotomy and drainage when there were gross adhesions or the condition of the patient did not permit splenectomy. Splenic fungal abscesses have developed in immunosuppressed patients and in patients taking steroids or chemotherapeutic agents. Many of these patients have been on systemic antibiotics, and intestinal colonization with *Candida* developed. In some cases the abscesses resolved with treatment with antifungal drugs, but more often success has been achieved when the drugs were combined with splenectomy.

SPLENECTOMY

When elective splenectomy is performed for hematologic disorders, specific considerations arise. Patients with malignant lymphoma and leukemia may develop cryoglobulinemia, and therefore the blood should be administered at room temperature. For patients with thalassemia and, more particularly, acquired hemolytic anemia, typing and crossmatching may be difficult, and sufficient time should be allotted during the preoperative period to accumulate the blood that may be required during the operation. For patients with marked immune thrombocytopenia (ITP), platelet packs are not administered preoperatively, since the platelets are rapidly destroyed by the spleen and thus are not very effective. A nasogastric tube is inserted after endotracheal intubation to decompress the stomach and to facilitate handling of the short gastric veins.

Technique

Although the midline incision is preferred for exposure of a ruptured spleen, either a left subcostal incision or a midline incision may be used for elective resection. The spleen is mobilized initially by dividing the ligamentous attachments, which are usually avascular but may contain large vessels in patients with secondary hypersplenism and myeloid metaplasia (Fig. 31-7 A). The short gastric vessels are then doubly ligated and transected (Fig. 31-7 B), with care being taken not to traumatize the stomach itself. If compromise of blood supply to the fundic portion of the greater curvature of the stomach is a concern, enfolding of this area should be performed to prevent the development of a gastric fistula. Finally, dissection of the splenic hilus with individual ligation and division of the splenic artery and vein is performed (Fig. 31-7 C). During the course of hilar dissection, care should be taken to avoid injury to the tail of the pancreas in order to avoid pseudocyst formation. The technique of initial ligation of the splenic artery by exposure through the gastrosplenic omentum has been used in cases of hypersplenism and splenomegaly. Preoperative occlusion of the splenic artery using steel coils placed selectively in the splenic artery under radiographic control has achieved the same end.

Whenever splenectomy is performed for a hematologic disorder, a careful search should be made for accessory spleens. The splenic bed is not drained routinely, but drains are used in patients with myeloid metaplasia if there is a question of continued oozing from distended collateral veins.

Partial Splenectomy

This procedure is particularly applicable in children with Type I Gaucher's disease to minimize the risk of postsplenectomy sepsis. After transection of the ligamentous attachment to permit complete mobilization of the organ, the hilar vessels to the segment to be removed are ligated and divided. The demarcation of the devascularized segment defines the line for transection; at least 30 percent of the spleen should be preserved (Fig. 31-7 D). Bleeding from the raw surface is controlled by cauterization, argon coagulation, or fibrin glue. The omentum can be used to cover the remaining raw surface.

Postoperative Course and Complications

Following splenectomy, characteristic changes in blood composition occur. Howell-Jolly bodies are present in almost all patients, and siderocytes are common. Generally, leukocytosis and increased platelet counts are observed. In patients with marked thrombocytopenia, the platelet count often returns to normal within 2 days, but peak levels may not be reached for 2 weeks. The white blood cell count usually is elevated the first day and may remain persistently elevated for several months. The most frequent complication is that of left lower lobe atelectasis. Other complications include subphrenic hematoma and abscess, injury to the pancreas causing fistula, or pancreatitis. Excessively elevated platelet counts, particularly in patients with myeloid metaplasia, and increased platelet adhesiveness have been reported. Although these factors have been implicated in the greater incidence of thrombophlebitis following splenectomy, many series can show no good correlation between these complications and the platelet counts. In patients with hereditary hemolytic anemia and associated red cell enzyme deficiency, postoperative thrombocytosis may lead to hepatic, portal, or caval thrombosis, particularly if the hemolytic rate is unabated by splenectomy.

There has been increasing concern regarding infection and sepsis in splenectomized patients. In a large review Singer reported that deaths from sepsis in splenectomized patients were 200 times as frequent as in the population at large. Fulminant postsplenectomy infection is very rare following removal of a traumatized spleen, but the incidence of pneumonia, septicemia, and meningitis was reported to be 166 times greater in patients who had undergone splenectomy for trauma than is expected for the normal population. No neutrophil defect or impairment of serum opsonin or chemotactic activity could be demonstrated after splenectomy for trauma. When the risk of infection in childhood Hodgkin's disease was assessed in 181 consecutive, previously untreated patients, the episode did not correlate with splenectomy alone, but appeared to be related more to the treatment with chemotherapy and radiotherapy. In a series of 221 patients with hematologic disease, late postsplenectomy pulmonary infection was seen in eight patients; all but one were receiving immunosuppressive therapy or had advanced malignant disease.

Splenectomy results in a variety of immunologic defects, including a poor response to intravenous immunization with particulate antigens, a deficiency in phagocytosis-promoting peptide, decreased serum IgM, and decreased properdin. The organisms most frequently isolated from septic splenectomized children are those of diplococcus pneumonia and *Haemophilus influenzae*. It is therefore reasonable to delay splenectomy for hematologic disorders in very young children, especially those under the age of 2 years, to attempt to preserve the traumatized spleen in the pediatric patient and to maintain patients with at-risk diseases on long- term oral antibiotic therapy.

All patients undergoing splenectomy should receive, as prophylaxis, pneumococcus vaccine and a vaccine against *H. influenzae*, if available. Vaccination is best performed about 10 days before elective splenectomy and preoperatively in traumatized patients in whom splenectomy is anticipated, but should be performed postoperatively if preoperative vaccination was neglected. Equivalent response to pneumococcal antigens has been demonstrated for patients vaccinated after splenectomy compared with those vaccinated with the spleen in place. All children whose spleens have been removed should receive penicillin until the age of eighteen. The patients with diseases associated with high infection rates, such as thalassemia, sickle cell anemia, and autoimmune hemolytic anemia and thrombocytopenia, should not be denied splenectomy but should be kept under close surveillance postoperatively.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 32 - Intraabdominal Infections

Joseph S. Solomkin
Dietmar W. Wittman
Michael A. West
Philip S. Barie

INTRODUCTION

Intraabdominal infections are commonly encountered in surgical practice and represent an important problem in decision-making, operative technique, and perioperative management. Improvements in supportive care, diagnostic methods, anti-infective therapy, and interventional techniques have resulted in improved outcomes, decreased mortality, and a decreased incidence of recurrent abscess. Despite state-of-the-art care, infections result in considerable morbidity and mortality because of activation of local and systemic inflammatory responses.

Prompt and effective decision making is critical; delay in treatment for intraabdominal infection risks physiologic deterioration and increased mortality. Diagnostic uncertainty must be balanced against the risk of treatment delay. The precise nature of the operative procedure is a critical determinant of outcome and must be tailored to the pathology encountered.

The patient with findings suggestive of infection, typically with unexplained ileus, in whom no source of nosocomial infection can be identified by routine investigations represents a major diagnostic problem. In patients who have recently undergone abdominal operation, a concern for abscesses is often sufficient to lead to empiric treatment. Such an approach can result in overlooking less common causes of infection, bowel ischemia, pseudomembranous colitis, or pancreatitis.

STRUCTURE AND FUNCTION OF THE PERITONEAL CAVITY

Mesothelium

The total area of the peritoneum is approximately 1.8 m². It is formed by a single layer of mesothelial cells with an underlying supporting layer of highly vascularized loose connective tissue. The mesothelial cells contain microvilli 1.5 to 3.0 mm in length, which greatly increase the surface area of the mesothelial cells. The relative density of microvilli differs throughout the peritoneal cavity. Tight intra-mesothelial cell junctions generally connect these cells, although there also are large intercellular gaps, sometimes as large as 500 angstroms, between these lining cells. Alterations in mesothelial cell metabolism and cellular swelling may influence diffusion across this cell layer.

Mesothelial cells are organized into two discrete populations: cuboidal cells and flattened cells. Gaps (stomata) between neighboring cells of the peritoneal mesothelium are found only among cuboidal cells. Peritonitis increases the width of these stomata. Beneath the mesothelial cells is a basement membrane of loose collagen fibers, which offers little resistance to diffusion of molecules smaller than 30 kD. The basement membrane overlies a complex connective tissue layer that includes collagen and other connective tissue proteins, elastic fibers, fibroblasts, adipose cells, endothelial cells, mast cells, eosinophils, macrophages, and lymphocytes (Fig. 32-1). The capillaries branch and ramify within the peritoneal lining layer. In addition, there is a rich lymphatic network.

Peritoneal fluid is secreted by the peritoneal serosa and has the properties of lymph. Diaphragmatic lymphatic channels provide a means for the entry of peritoneal fluid (and any bacteria and proinflammatory mediators) through the thoracic duct into the venous circulation. Inspiration decreases intrathoracic pressure relative to intraabdominal pressure, creating a pressure gradient favoring fluid movement out of the abdomen. Intravascular entry of proinflammatory substances produces many of the hemodynamic and respiratory findings of severe sepsis. Positive-pressure ventilation likely attenuates this process but has not been well studied as a therapeutic maneuver.

Anatomy

The peritoneal cavity encompasses the potential space defined by the mesothelial serous membrane and extends superiorly from the diaphragm to the pelvis in its most caudad extent (Figs. 32-2 and 32-3). Anteriorly the peritoneal cavity reflects onto the posterior aspect of the anterior abdominal musculature. Posteriorly the peritoneal lining lies superficial to the retroperitoneal viscera, including the aorta, vena cava, ureters, and kidneys. The anterior and posterior peritoneal layers are described collectively as the parietal peritoneum. The visceral peritoneum represents the mesothelial lining cells that are reflected onto the surface of the viscera, including the stomach, small bowel, spleen, liver, gallbladder, ovaries, uterus, and portions of the bladder, colon, and pancreas. The peritoneum covering the intestine is the serosa of the bowel. The peritoneum also lines the lesser sac, which communicates with the remainder of the peritoneal cavity via the epiploic foramen (foramen of Winslow). In women the peritoneal lining is reflected onto the fallopian tubes, which communicate through the open fimbriated ends with the uterus and vagina.

The visceral peritoneal lining of the liver and spleen (the “capsules”) are thicker and can be detached from the underlying organ parenchyma, whereas the visceral lining of the bowel is thin and intimately attached to the lamina muscularis mucosae.

The parietal peritoneum is thicker than the visceral peritoneum and contains a richer capillary network that ramifies extensively within the peritoneal lining. This vascular supply allows the parietal peritoneum to be readily dissected free from deeper structures without altering its viability.

The retroperitoneum, which contains a subfascial plane extending from the pelvis to the posterior pharynx, is located deep to the posterior peritoneal lining. In some types of intraabdominal infection, once the inflammatory process penetrates the serosal lining, it is possible for it to spread laterally, caudad, or cephalad with relative ease.

Under normal circumstances the peritoneal cavity is largely a potential space, as only a thin film of fluid separates the parietal and visceral layers. This fluid layer serves as a lubricant, allowing the abdominal viscera to slide freely within the peritoneal cavity. The capacity of this space is illustrated during peritoneal dialysis, as 2 to 3 L of fluid are instilled into the peritoneal cavity without any patient discomfort. Inflammation of the immense surface area of the peritoneal mesothelium can result in massive fluid shifts. For example, an increase in thickness of the peritoneal lining of 1 mm will sequester around 1.8L of fluid. Peritonitis thus may be regarded as requiring the resuscitation equivalent of an 80 percent burn.

Compartmentalization of the peritoneal cavity by various ligaments and retroperitoneal attachments creates several recesses into which fluid or exudate may pool (see Fig. 32-3).

Inframesocolic Spaces

The area below the transverse mesocolon is subdivided into several gutters, or watersheds, by the ascending and descending colon, the root of the mesentery, and the pelvic mesosigmoid. The *right lateral paracolic gutter* lies between the cecum and ascending colon and the right lateral abdominal wall. The drainage pathway of the right lateral paracolic gutter is in a superior direction when the patient is in the supine position, since the superior portion of this sulcus lies on a more posterior plane than the inferior portion. The right lateral paracolic gutter continues superiorly to the region behind the right lobe of the liver and anterior to the right kidney as the hepatorenal pouch. The hepatorenal pouch communicates through the foramen of Winslow with the lesser sac.

Between the mesentery and the cecum and ascending colon is the *right medial paracolic gutter*. This recess is limited inferiorly by the junction of the mesentery and the jejunum and ileum, where the mesentery meets with the cecum, and superiorly by the transverse mesocolon. The drainage from the right medial paracolic gutter occurs in a superior direction, passing above the root of the mesentery into the *left medial paracolic gutter*. The left medial paracolic gutter is limited by the transverse mesocolon above, by the root of the mesentery on the right, and by the descending and sigmoid colon on the left. Drainage from the left medial paracolic gutter is downward into the pelvis. The *left lateral paracolic gutter* extends superiorly from the phrenicocolic ligament to the sigmoid colon inferiorly and from the descending colon on the right to the left lateral abdominal wall. Fluids in the left lateral paracolic gutter tend to remain there.

Supramesocolic Spaces

With the patient in the supine position, drainage on the right tends to flow superiorly toward the right subphrenic area. This area in turn is divided into two portions, a right and a left, by the falciform ligament, which extends from the anterior abdominal wall and inferior surface of the diaphragm to the anterosuperior surface of the liver. The right subphrenic space is further subdivided by the right lobe of the liver into a suprahepatic portion and an infrahepatic portion. The left subphrenic area is divided by the left lobe of the liver and the horizontally placed stomach and lesser omentum. There are two major subdivisions of this space. The larger area lies anterior to the stomach, the spleen, and the left lobe of the liver and is limited by the diaphragm and the abdominal wall anteriorly. The smaller area, the lesser sac, lies behind the stomach and lesser omentum.

With the body in the supine position, the most dependent recess is in the pelvis. In the upper abdomen, the posterosuperior extension of the subhepatic space known as Morison's pouch also collects fluid in recumbent individuals. Morison's pouch connects via the foramen of Winslow (bounded by the free border of the gastrohepatic omentum and the posterior parietal peritoneum) to the lesser sac, which is the largest recess in the peritoneal cavity. Because of the limited communication of the lesser sac to the main peritoneal cavity, abscesses in the lesser sac may not produce physical signs on examination of the anterior abdominal wall. The left subhepatic and subphrenic spaces communicate freely around the small lateral left hepatic lobe. The falciform ligament limits communication between the right and left subphrenic spaces. This may explain the clinical observation that fewer than 15 percent of subphrenic abscesses are bilateral.

Innervation

Innervation of the peritoneal lining figures prominently in clinical recognition of intraabdominal infections. There are some nerve fibers within the visceral peritoneum, but they are much less numerous than those in the parietal peritoneum. They arise from the visceral structures beneath the mesothelial lining. Stretching of the visceral peritoneum produces nausea and poorly localized pain. Most intraperitoneal visceral structures arise from the embryologic foregut, and, accordingly, their nerve supply arises from the visceral branches of the vagus nerve. Pain usually is localized to the dermatome distribution of the associated visceral organ (Table 32-1). For example, the gallbladder arises from the upper thoracic dermatomes; hence, with stretching of the visceral peritoneum overlying the gallbladder, pain is initially perceived as arising in the epigastrium and right upper back. When the inflammatory process also involves the adjacent diaphragmatic structures, pain also may be perceived in the neck or shoulder because the diaphragm arises from and is innervated by the cervical dermatomes. Visceral peritoneal irritation, usually from distention of a hollow viscus, causes dull, poorly localized, often periumbilical, and often crampy pain. Visceral structures are generally completely insensitive to surgical transection, cauterization, heat, cold, or pinching.

The nerve supply of the parietal peritoneum, the somatic afferents, arises from branches of cutaneous nerves in the anterior abdominal wall. In contrast to the visceral peritoneum, the parietal peritoneum is exquisitely pain sensitive to stretch as well as light touch and cutting, particularly anteriorly. Pain arising from irritation of the parietal peritoneum can usually be very precisely localized by patients and is the basis for the clinical findings of “peritoneal signs.” Anterior parietal peritoneal pain may be associated with tenderness and involuntary muscle spasm (guarding). “Rebound tenderness” can sometimes be elicited from the taut peritoneum even in the absence of inflammation, such as in ileus. Irritation of the peripheral diaphragmatic peritoneum is perceived as pain in the adjacent body wall, whereas central irritation manifests itself as ipsilateral shoulder pain. Diaphragmatic irritation

may arise not only from direct involvement by adjacent inflammation but also from the movement of infected fluid into the subphrenic spaces.

The blood and lymphatic supply of the peritoneum generally follow the distribution of the innervation. The blood supply of the visceral peritoneum arises from the adjacent visceral organ.

PHYSIOLOGY OF THE PERITONEUM

Peritoneal Fluid Exchange

The mesothelial lining cells of the peritoneum secrete serous fluid that circulates within the peritoneal cavity. The peritoneal cavity normally contains 50 to 100 mL of fluid with solute concentrations nearly identical to that of plasma. The protein content of peritoneal fluid is somewhat less than that of plasma, about 3 g/dL. Fluid is absorbed by the peritoneal mesothelial lining cells and subdiaphragmatic lymphatics. Mesothelial cells also absorb solutes by the continuous process of endocytosis. Exchange of solutes across the peritoneal lining is the basis for peritoneal dialysis. Most solutes with a molecular weight less than 30 kD traverse the mesothelial lining with relative ease. Splanchnic blood flow and factors that alter membrane permeability affect the efficiency of fluid exchange. Peritoneal permeability is markedly increased by intraperitoneal inflammation.

Peritoneal Fluid Movement

The routes of normal fluid movement within the peritoneum have been defined by injection of water-soluble contrast material into normal individuals. The right paracolic gutter is the main conduit between the upper and lower peritoneal cavities, because the left gutter is obstructed by the phrenicocolic and splenorenal ligaments. Fluid introduced into the right upper quadrant gravitates inferiorly via the subhepatic space into the right subphrenic space. From there it may travel inferiorly into the pelvis via the right gutter before returning to the upper abdomen. Because of the obstruction on the left, fluid from the left upper quadrant migrates into the left subphrenic space. Fluid introduced into the lower abdomen gravitates first to the pelvis, and thereafter ascends, even with the patient in the upright position, into the subhepatic and subphrenic spaces.

Two primary forces govern the movement of fluid within the peritoneal cavity: gravity and the negative pressure created beneath the diaphragm with each normal respiratory cycle. Fluid flux within the peritoneal cavity is dramatically altered by the presence of adhesions, fibrin, paralytic ileus, or mechanical ventilation. Movement of fluid into the pelvis was in the past an important component of the surgical treatment of intraabdominal infections. Earlier generations of surgeons routinely positioned patients in Fowler's position—almost a sitting position—to facilitate dependent movement of purulent material and formation of a pelvic abscess, which could then be drained transrectally without laparotomy. Today the Fowler position is seldom used and pelvic abscesses are relatively rare. Intraabdominal abscesses are most frequently encountered beneath the liver or diaphragm and between loops of bowel.

Subphrenic purulent fluid collections occur because a relatively negative pressure is created beneath the diaphragm with each exhalation. Intraperitoneal pressure measurements show that the pressure is lowest beneath the diaphragm during expiration. The diaphragm raises during exhalation, producing a transiently larger space in the upper abdomen. As the volume increases the pressure decreases. Lower pressures are encountered whether patients are breathing spontaneously or on mechanical ventilation. With positive-pressure mechanical ventilation there is a significantly impaired capacity of the peritoneal cavity to clear particulate debris.

Peritoneal Response to Injury

Peritoneal Injury and Repair

Any inflammatory event in the peritoneal cavity results in local peritoneal irritation with loss of regional mesothelial cells. The defect in the mesothelial lining is repaired by “metastasis” of nearby mesothelial lining cells. Peritoneal defects heal everywhere simultaneously. A large peritoneal defect heals in the same amount of time as a small defect, usually 3 to 5 days. This process is rapid and usually reconstitutes the peritoneal continuity without adhesion formation. The origin of the migrating mesothelial cells remains obscure; they may arise from submesothelial stem cells.

Adhesion Formation

Fibrin polymerization occurs when platelets and fibrin come into contact with the exposed basement membrane, an automatic accompaniment of intraperitoneal inflammation. Normally, well-oxygenated intact mesothelial cells produce plasminogen activator, which lyses the fibrin clots that form after acute injury. In experimental wounds fibrinolytic activity has been found to be minimal 3 days after injury, before reconstitution of the mesothelial integrity. Fibrinolytic activity increases to supranormal levels 8 days after injury. With hypoxia the fibrous adhesions are invaded by fibroblasts and there is stimulation of angiogenesis and collagen synthesis. Fully developed fibrous adhesions are seen at 10 days and become maximal 2 to 3 weeks after peritoneal injury. This explains the difficulty of reoperation 2 to 4 weeks after an acute insult. With time, fibrous adhesions undergo remodeling and usually become progressively attenuated. In patients with intraabdominal soiling complications, such as patients with enterocutaneous fistulas, mortality is 20 percent in operations performed between 10 and 120 days but 10 percent before 10 days and after 120 days.

Host Defense Against Intraabdominal Infection

The three major intraperitoneal defense mechanisms are (1) mechanical clearance of bacteria via lymphatics, (2) phagocytic killing of bacteria by immune cells, and (3) mechanical sequestration. Experimentally, bacterial clearance occurs rapidly, before influx of phagocytic cells. Particulate material within the peritoneal cavity is cleared through stomata that overlie large lymphatic channels (lacunae) on the subdiaphragmatic surface. Diaphragmatic movement with respiration produces influx of fluid into the lacunae with exhalation, and respiratory contraction empties the lacunae into the thoracic duct and ultimately the left subclavian vein. Bacteria (1 to 2 mm) and red blood cells (7 to 8 mm) readily pass through the stomata (8 to 12 mm) into the thoracic duct and thence into the central venous blood. This rapid access of bacterial products to the systemic circulation explains why patients who develop an acute gastrointestinal perforation often present with rigors and fever.

PATHOPHYSIOLOGY OF THE LOCAL AND SYSTEMIC RESPONSE TO INTRAABDOMINAL INFECTIONS

The cellular defense mechanisms of the peritoneal cavity help to explain the specific pattern of response. Elements of these defenses include resident peritoneal macrophages and large recruitable pools of circulating neutrophils and monocytes, cell types that participate in abscess formation. Ingestion of microorganisms by these cells may result in secretion of a variety of proinflammatory molecules that generate the systemic response (Table [32-2](#)).

The peritoneal host defense system also serves to localize and contain bacterial contamination. Hyperemia and exudation of fluid follow activation of immune cells (macrophages, and probably mast cells) and sloughing of peritoneal mesothelium. Histamine and other permeability-increasing substances mediate this response. Several events favor the deposition of fibrin, including the activation of mesothelial and macrophage-mediated procoagulant activity acting on fibrinogen in reactive peritoneal fluid, coupled with the loss of plasminogen activator from mesothelial cells. The combined effect is the deposition of fibrinous exudates that serve to wall off and trap bacteria. Ileus and adhesion formation accentuate the process, but extensive loculation may impede immune cell migration to sites of infection.

The defense system's recognition that microorganisms are present within the peritoneal cavity occurs primarily through phagocytosis of the microorganisms by resident peritoneal macrophages. This results in activation of the local macrophages, with secretion of proinflammatory signaling molecules that recruit from a large circulating pool of phagocytes.

Cells Involved in the Inflammatory Response in the Peritoneal Cavity

At least four major cell types play key roles in the inflammatory response. These include resident as well as recruited macrophages, mesothelial cells, adjacent capillary endothelial cells, and recruited neutrophils. While the macrophage has been traditionally considered the first line of cellular defense, developments in our understanding of cytokines and the immunologic functions of mesothelial cells suggest that the inflammatory response is actually initiated by interactions between these two cell types. It has been suggested that the mesothelial cell could even be considered an immune cell, if this designation were not dependent on its lineage. The first step in the response to contamination is probably induced by a host response to microbial cell products. In the case of gram-negative organisms, this is most likely *lipopolysaccharide (LPS)*. For gram-positive organisms, the cell wall product teichoic acid and specific cell wall glycans induce a macrophage response. Lipopolysaccharide, present on gram-negative organisms, interacts with monocytes and other cell types, including subpopulations of neutrophils via the CD14 receptor. Interaction with the CD14 receptor is mediated by lipopolysaccharide binding protein (LBP), and leads to macrophage activation, with subsequent production of *tumor necrosis factor- α (TNF- α)*, interleukin (IL)-1, IL-6, IL-8, granulocyte-macrophage colony-stimulating factor (GM-CSF), prostaglandin E₂ (PGE₂), and nitric oxide. LPS is a very potent immunomodulator, with concentrations in the few-picograms-per-milliliter range inducing near maximal responses from human cells.

Macrophages

Macrophages, including resident peritoneal macrophages and those recruited from the circulation, serve as an early line of cellular defense and coordinate the inflammatory response. Macrophage products include IL-1, IL-6, IL-8, IL-10, IL-12, *TNF- α* , GM-CSF, granulocyte colony-stimulating factor (G-CSF), monocyte chemoattractant protein-1 (MCP-1), monocyte inflammatory protein α (MIP- α), and several eicosanoids. These molecules recruit and activate mesothelial cells as secretors of inflammatory molecules. Mechanisms of macrophage microbicidal activity and associated parenchymal injury include the respiratory burst and release of proteolytic enzymes into intracellular vesicles and into the extracellular space. These enzymes include plasminogen activator, collagenase, elastase, gelatinase, acid phosphatase, and cathepsin D.

Mesothelial Cells

Macrophage secretion products and microbes and their products activate and injure mesothelial cells. This mesothelial injury is a prominent feature of most types of inflammation. In response to these products mesothelial cells produce a variety of inflammatory molecules that recruit additional cells to the peritoneal cavity and activate them there. Mesothelial cells secrete IL-8, a neutrophil chemoattractant and activator. Macrophage-secreted components of complement, upon activation, also recruit neutrophils and monocytes. Mesothelial cells also express various adhesion molecules that promote leukocyte–mesothelial cell adherence.

Neutrophils

Neutrophils are the key effector cells in the inflammatory response and are responsible for the parenchymal destruction seen in inflammation. Neutrophils possess an array of granular enzymes that are capable of digesting basement and interstitial proteins, facilitating directed migration and allowing for digestion of injured tissue. Neutrophil-derived proteases also can degrade a variety of functionally important molecules found at sites of inflammation, including immunoglobulins and intermediates of the kinin, complement, fibrinolytic, and clotting cascades. Neutrophil azurophilic granules also contain the cationic antibacterial protein bactericidal permeability increasing (BPI) factor. BPI factor has sequence homology with lipopolysaccharide-binding protein and binds to the lipid A portion of lipopolysaccharide, preventing its interaction with the CD14 receptor and thereby inhibiting its toxic effects. BPI factor is also specifically cytotoxic to gram-negative bacteria. Neutrophil microbicidal activity and cytotoxicity are in part determined by a nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, which produces reactive oxygen intermediates. NADPH oxidase, a membrane-bound enzyme, produces reactive oxygen species such as hydrogen peroxide. Neutrophils also secrete *TNF- α* , IL-1, and IL-8, which activate cells and recruit additional neutrophils to sites of inflammation; in turn, they are highly responsive to TNF- α , IL-1b, IL-6, and IL-8.

Host defense systems have an obligatory requirement for molecular oxygen to enable the humoral and cellular immune systems to function optimally. Normally, after opsonization a microorganism is phagocytosed by a neutrophil. The phagosome containing the bacteria then fuses with lysosomal granules within the neutrophils. In the presence of molecular oxygen, toxic oxygen radicals, such as hydrogen peroxide or superoxide, are generated, which result in lysis of bacteria. In the presence of devitalized tissue and in many shock states, delivery of oxygen to the neutrophils is inadequate. In these instances, even though phagocytosis may occur, bacteria may not be destroyed and infection develops or persists.

Endothelial Cells

Macrophage, neutrophil, and mesothelial cell products also injure the closely approximated endothelial cell lining and can thereby increase movement of mediators into the vascular space. It is only upon activation or injury, mediated by *TNF- α*, IL-1, or oxygen radicals, that the endothelial cells secrete inflammatory molecules, which enhance leukocyte-endothelial adhesion and leukocyte transendothelial cell migration. Thus activation of the vascular endothelium is a key mechanism of inflammation.

Endothelial cells produce a variety of adhesion molecules, including members of the selectin and immunoglobulin families, that mediate leukocyte trafficking to sites of inflammation. Endothelial cells also produce IL-8 and *platelet-activating factor (PAF)*, which localize on the endothelial surface in association with various extracellular-matrix proteins. Endothelial cell-associated IL-8 and PAF contribute to the arrest of neutrophil locomotion and also activate them. IL-8 also is thought to mediate neutrophil emigration from the vasculature, in part by forming a chemotactic gradient from the luminal side to the basal side of the endothelium.

Other Cells

Additional cell types also are recruited in the acute inflammatory response. Platelet activation and adherence to sites of injury results in the formation of platelet plugs. During activation, platelets release several preformed molecules into the area of evolving inflammation, including platelet basic protein and its proteolytic derivative *neutrophil-activating peptide 2 (NAP- 2)*, a neutrophil chemoattractant. Platelets also produce *PAF*, tumor growth factor b (TGF-b), prostaglandins, thromboxane A₂ (TXA₂), lipoxin A₄ (LXA₄), and nitric oxide, which can modulate the inflammatory response. Vascular smooth muscle cells and fibroblasts also can release cytokines that can modulate the immune response. Additionally, T and B cells participate in the specific inflammatory response by recognition of antigens presented by monocytes, epithelial cells, and endothelial cells.

Mechanisms of Cell Recruitment to Sites of Injury: The Chemokines and Other Chemoattractants

Nonchemokine Chemoattractants

An early epithelial and endothelial event is contraction of these cells, allowing the influx of several important classes of molecules, including components of the complement system. The complement system is a series of proteins whose ordered polymerization results in activation of the coagulation system, the release of inflammatory molecules, recruitment of leukocytes (C5a), opsonization of microorganisms (C3b), and induction of microbial and host cell lysis (membrane attack complex). Released C5a also induces polymorphonuclear cell activation, adherence, and aggregation. C5a can induce the production of *TNF- α*, IL-1, and IL-6 by monocytes. In the lung C5a increases vascular resistance and induces hypoxemia. Complement also synergistically enhances the effects of TNFα in producing hemorrhagic necrosis.

Other chemoattractants recruit cells to the peritoneal cavity, including leukotriene B₄ (LTB₄), formyl-methionyl-leucyl-phenylalanine (FMLP), and platelet activating factor (PAF). LTB₄ is generated by macrophages (and other cells) in response to such stimuli as *LPS*. LTB₄ is chemotactic for neutrophils and increases neutrophil adhesion to endothelium. It also promotes IL-1, IL-6, *TNF- α*, and H₂O₂ production by monocytes. Additionally, LTB₄ increases vascular permeability. FMLP is derived from bacterial cell walls and induces neutrophil and monocyte chemotaxis, degranulation, superoxide production, and phagocytosis. *PAF*, released by neutrophils, monocytes, platelets, and endothelial cells, also stimulates macrophage and neutrophil chemotaxis. PAF also enhances superoxide production by neutrophils at the endothelial surface, increases endothelial permeability, alters vascular reactivity, enhances platelet aggregation, and stimulates production of eicosanoids. PAF release from endothelial cells can be induced by TNF, IL-1, *LPS*, thrombin, phorbol myristate acetate, histamine, leukotrienes, and reactive oxygen species.

Chemokines

Chemokines, or chemotactic cytokines, are the key mediators of leukocyte recruitment to sites of inflammation. Two broad categories of chemokines are the CXC chemokines and the CC chemokines, where “C” refers to the location of cysteine residues in the molecules.

CXC Chemokines

Important cytokines in this category include IL-8, *NAP-2*, and melanoma growth stimulatory activity/growth related oncogene (MGSA/GRO). IL-8 is an 8-kD protein that is produced by macrophages, mesothelial epithelial cells, neutrophils, endothelial cells, fibroblasts, T cells, and hepatocytes. IL- 8 and, to a variable extent, NAP-2 and GRO induce neutrophil chemotaxis, prime neutrophils for enhanced superoxide production, modulate adhesion molecule expression, and enhance transendothelial migration. IL-8 also has some chemotactic activity on lymphocytes.

Several studies have identified elevated plasma levels of IL-8 in various patient groups. However, IL-8 is only functional when bound to matrix and cell surface proteins. Free IL-8 is bound by circulating autoantibodies and to a low-specificity erythrocyte receptor, the Duffy antigen, which may serve as a clearing mechanism. The cell surface-IL-8 interaction is dependent on specific heparin molecules, which provide a high degree of specificity to the response. Heparin sulfate proteoglycan chains self-associate and bind to various extracellular matrix proteins, growth factors, and cell surface enzymes. Heparinase activity releases growth factors from their heparin sulfate storage sites in the extracellular matrix, providing a mechanism for the induction of migration of diverse cell types in normal or pathologic situations. Neutrophils themselves possess and secrete heparinases.

GROα is secreted by endothelial cells, monocytes, and fibroblasts after stimulation with *LPS*, IL-1, or *TNF- α*. *NAP-2* is derived from cleavage of a platelet secretory product, platelet basic protein. Macrophages and neutrophils can cleave it from its precursors, and NAP-2 serves to modify subsequent neutrophil function.

CC Chemokines

The second set of chemokines, CC chemokines, include monocyte inflammatory protein (MIP), monocyte chemoattractant protein (MCP), and RANTES. MCP-1 is a 76 AA polypeptide, produced by monocytes, endothelial cells, lymphocytes, and fibroblasts, that induces monocyte chemotaxis and activation. Mononuclear phagocyte-derived MIP-1α is chemotactic for T cells and monocytes. MIP-1 is produced by monocytes, lymphocytes, and fibroblasts and mediates both T cell chemotaxis and adhesion. RANTES is induced by TNFα and IL-1 in lung epithelial cells and is chemotactic for T lymphocytes.

Leukocyte–Endothelial Cell Adhesion and Emigration from the Vasculature

Binding of the recruited leukocyte to the endothelial cell, emigration from the vasculature and into the parenchyma, and subsequent activation are mediated via a series of adhesion molecules. These include members of the immunoglobulin (intercellular adhesion molecule, ICAM), integrin, and selectin families. A well-accepted model of neutrophil adhesion to endothelium involves at least three sequential events.

Selectin-Mediated Leukocyte Rolling

The first stage, rolling, occurs constantly and involves the transition of the leukocyte from the circulating state, in which it is moving in the center of the vessel, to a slower rolling state, in which it is tumbling along the wall of the capillary or venule. Rolling leukocytes may adhere firmly or may detach and rejoin the mainstream of flowing blood. The prevailing wall sheer rate in postcapillary venules determines in part the level of leukocyte rolling and of firm adhesion and dictates the contact area between rolling leukocytes and the endothelial cell surface.

Rolling involves the selectins, which are calcium-dependent glycoprotein receptors that have a common molecular structure. The selectins are named after the cell type on which they are primarily found: E-selectin (endothelial cell), P-selectin (platelet), and L-selectin (leukocyte). L-selectin mediates rolling in part by presenting neutrophil carbohydrate ligands to the endothelial E- and P-selectins. E-selectin (endothelial leukocyte adhesion molecule-1, ELAM-1) is found exclusively on endothelial cells activated by *TNF- α*, IL-1, *LPS*, interferon-γ, and substance P. It allows endothelial cells to bind to neutrophils, monocytes, and certain memory T cells. L-selectin is expressed by neutrophils and monocytes as well as certain populations of lymphocytes, natural killer (NK) cells, and hematopoietic progenitor cells. It is concentrated on the microvilli projections of unstimulated neutrophils. Rolling also allows neutrophils to pick up endothelial-borne chemical signals such as IL-8 and *PAF*, which then promote integrin-mediated adhesion and subsequent migration of neutrophils.

Integrin-Mediated Leukocyte Activation and Transendothelial Migration

The second event in neutrophil–endothelial cell adhesion is activation, mediated by chemotactic agents that are released from or are attached to the endothelium. Leukocyte surface expression of the β₂ integrin CD11B/CD18 integrin is increased. This activation enables the third stage of firm adhesion and transendothelial migration to occur by promoting the function of integrin adhesion molecules on the leukocyte surface. Subsequent neutrophil transmigration to the extravascular space involves adherence via other leukocyte integrins.

Cytokines in the Systemic Response to Infection

Liver dysfunction is common during the course of intraabdominal infection and occasionally progresses to fatal hepatic failure. There is considerable evidence that various macrophage products, including IL-1, IL-6, and TNF- γ, substantially alter hepatocyte function. Aside from conversion of hepatic synthetic function to acute phase reactants, serum chemistry studies reveal evidence of ductal epithelial cytotoxicity, including elevated levels of alkaline phosphatase and bilirubin. The large number of fixed tissue phagocytes (Kupffer cells) in the liver capable of responding to endotoxin absorbed from systemic or mesenteric blood vessels represents a potentially important source of cytokines and other hepatocyte regulatory substances, although portal endotoxemia has not been detected in humans.

BACTERIOLOGY OF PERITONITIS AND INTRAABDOMINAL INFECTION

Normal Bowel Flora

The vast majority of the bacteria in the colon are anaerobic species that, so far as is known, contribute little to clinical intraabdominal infection. The most common bacteria isolated in clinical infections are *Escherichia coli* and *Enterobacter*, *Klebsiella*, and *Pseudomonas* species (Table 32-3), which make up less than 0.1 percent of the normal colonic flora. Even the most common anaerobic pathogen, *Bacteroides fragilis*, accounts for only 1 percent of the colonic flora. The presence of large numbers of nonpathogenic bacteria provides a measure of protection to the host by suppressing the growth of potentially pathogenic bacteria. Overgrowth of pathogenic gram-negative aerobic bacteria is commonly seen after treatment with broad-spectrum antibiotics. Many of these bacteria are associated with nosocomial infections that are much more difficult to eradicate.

Effect of Level of Gastrointestinal Perforation

The morbidity and mortality of intraabdominal infections vary dramatically with the level of gastrointestinal tract perforation, because the number and type of microorganisms vary throughout the gastrointestinal tract. Under normal circumstances the stomach contains fewer than 10^3 bacteria per cubic millimeter, largely because of the action of hydrochloric acid and the very low pH in the stomach. When patients receive H_2 -receptor blockers and proton-pump inhibitors that raise gastric pH, the number of bacteria rapidly approaches the levels seen in the proximal small bowel. The number of bacteria per cubic millimeter increases with distance down the gastrointestinal tract. In the proximal small bowel there are approximately 10^4 to 10^5 bacteria/mm³, whereas the terminal ileum contains more than 10^9 /mm³. The highest absolute numbers of bacteria are found in the colon, where there are between 10^{10} and 10^{12} bacteria/mm³.

The type of bacteria changes with distance down the gastrointestinal tract as well. In the upper gastrointestinal tract facultative gram-negative aerobic bacteria predominate, whereas the colon contains many more anaerobic than aerobic bacteria. In the colon, gram-negative and gram-positive anaerobic bacteria are present, along with facultative aerobes. The number of species of bacteria isolated also is highest in the colon, with more than 500 species of bacteria and fungi.

Virulence Factors

Virulence factors provide pathogenic bacteria with a selective advantage in that they thwart normal host defense mechanisms by impairing opsonization or phagocytosis. In other cases, extracellular products such as coagulase and catalase are secreted by *Staphylococcus aureus*. In the case of *B. fragilis*, the presence of a polysaccharide capsule is strongly associated with abscess formation. In experimental studies, the injection of a component of the polysaccharide capsule is as efficient as the entire bacteria for causing abscesses.

Microbial Adherence to Peritoneum

Microbial colonization of the peritoneal mesothelial surface is a rapid phenomenon after penetration injury of the distal bowel. Bacteria adherent to the peritoneum are resistant to removal by peritoneal lavage, in contrast to the bacteria present in peritoneal fluid, which are significantly decreased after lavage. Enterobacteriaceae (aerobic *E. coli* and similar organisms) rapidly colonize serosal mesothelium and become the predominant flora within 4h after induction of peritonitis. After 8 h, the *B. fragilis* group is the predominant organism adherent to the peritoneum and is unaffected by extended saline lavage.

Antimicrobial lavage produces an immediate but transitory decrease in the numbers of bacteria; the microbial population exceeds prelavage levels 24 h later. This favors reexploration at 24-h intervals when operative management of intraabdominal infection is done by staged abdominal repair.

Role of Aerobic and Anaerobic Bacteria

Among aerobic bacteria, gram-negative species of the Enterobacteriaceae family (particularly *E. coli*) and the *Klebsiella* and *Enterobacter* genera predominate. Among anaerobic bacteria the most common isolates are *Bacteroides* species. In experimental models that have been reported, when pure cultures of *E. coli* were injected into animals, peritonitis developed in all animals, with a high incidence of *E. coli* bacteremia and significant mortality. Injection with *E. coli* alone did not result in any intraabdominal abscesses, which are usually seen in patients who survive the diffuse inflammatory phase of peritonitis. By contrast, when *B. fragilis* was injected, there was almost no mortality, and very few *B. fragilis* were recovered from the bloodstream, but the incidence of intraabdominal abscesses was almost 100 percent. Injection of the combination of *E. coli* and *B. fragilis* resulted in a picture similar to that seen in human patients, with significant mortality and intraabdominal abscesses in most survivors.

The LPS present on the surface of gram-negative organisms, including those of the Enterobacteriaceae family, is a key virulence mechanism, because it interacts with a broad range of human cell types to induce an inflammatory response. The initial response includes thrombosis, restricting blood supply to an area of contamination. An influx of cells that are profoundly cytotoxic, both for the bacteria and resident tissue, follows. These responses reduce the numbers of organisms needed to establish an infection, a paradoxical consequence of the inflammatory response. This paradox may be explicable if the inflammatory response is a means of generating abscesses with subsequent extrusion to an external surface.

Microbial Synergy

The consequences of synergy are most evident in necrotizing soft tissue infections. Characteristic histologic findings include extensive local thrombosis and dense growth of microorganisms responding only to wide excision with normal tissue. Synergistic interactions between anaerobes, most notably *B. fragilis*, and endotoxin-bearing gram-negative organisms suppress local host defense mechanisms and facilitate the establishment of infection. *B. fragilis* produces a capsular polysaccharide that interferes with complement activation and inhibits leukocyte function. Delivery of phagocytes to the site of infection is restricted, permitting a more rapid rate of bacterial growth.

Aerobic bacteria lower the oxidation-reduction potential, thus favoring the growth of anaerobic bacteria. Anaerobic bacteria elaborate short-chain fatty acids that interfere with neutrophil function. Either type of bacteria may enhance the growth of other bacteria by provision of nutrient factors or destruction of antibiotics. High concentrations of bacteria inhibit the ability of most antimicrobial agents to function optimally. Extracellular products such as proteases facilitate bacterial invasion.

Host Effects on Bacterial Growth

The host neurohumoral response to infection may enhance bacterial growth in that gram-negative bacterial growth is dramatically enhanced by physiologic concentrations of norepinephrine, an effect mediated by a receptorlike mechanism within the bacteria. Some bacterial pathogens may have evolved mechanisms to exploit the inflammatory milieu. This may partially explain the well-known effect of traumatic injury on enhancing the lethality of intraabdominal infection.

Adjuvant Substances

The inoculum of bacteria needed to establish experimental infections is much less if subjects simultaneously are inoculated with adjuvant substances. Adjuvants increase bacterial virulence or interfere with host defenses. In gastrointestinal perforation, a surgical procedure, or trauma, the bacterial inoculum is invariably accompanied by adjuvant substances (Table 32-4).

Clinically, the most important adjuvant substance is blood. Hemoglobin, fibrin, and platelets in blood all impair peritoneal defenses. Hemoglobin decreases the median lethal dose of bacteria in standard rodent models of peritonitis by five orders of magnitude. Hemoglobin increases bacterial growth rates, perhaps because of the liberation of iron, an essential growth factor for *E. coli*. Iron may depress neutrophil migration and phagocytic function. Stroma-free hemoglobin is directly toxic to neutrophils as well. Platelets impair clearance of bacteria from the peritoneal cavity, presumably by physical obstruction of the diaphragmatic lymphatic channels. Fibrin promotes bacterial trapping and abscess formation but may impede the ability of neutrophils to reach the bacteria. Fibrinolytic agents instilled intraperitoneally in experimental animals were found to lyse clots and prevent abscess formation but were associated with higher bacteremia rates and higher lethality.

Other substances that may gain access to the peritoneal cavity during infection include bile salts, gastric mucin, pancreatic secretions, urine, and chyle. Bile salts aid in digestion of lipids because of their detergent action. Bile salt also can result in destruction of leukocytes and impaired host defense. Some bacteria may be digested by bile salts, but others, for example, *Enterococcus faecalis*, can readily grow in media that is 40 percent bile salts—which is the basis of the most common microbiologic test to identify *E. faecalis*. Pancreatic secretions contain proenzymes that can become activated by bacterial infection, resulting in tissue digestion and increased bacterial invasion.

Foreign Materials

Foreign materials may play an important role in the pathogenesis of infection. Microorganisms adherent to foreign materials are far more difficult for phagocytes to kill, and the foreign body may itself induce an inflammatory reaction, reducing the inoculum needed to establish infection. Foreign materials that play a role in the pathogenesis of abdominal infections may be macroscopic or microscopic (Table 32-5). Macroscopic foreign bodies such as drains, suture material, microscopic fragments of cotton gauze sponges shed from laparotomy pads, and hemostatic agents such as collagen or cellulose potentiate infection. Microscopic foreign materials include talc, barium sulfate, necrotic tissue, and fecal material. Some foreign materials may be directly toxic to neutrophils. Mucin and fiber can activate complement. Nylon, including the nylon in sutures, can cause neutrophil degranulation and impairment of oxidative microbial killing. Barium sulfate, a particularly virulent adjuvant, produces its own acute inflammatory insult. The combination of barium and feces is more lethal than either alone. The chemical injury to the peritoneum by barium increases its permeability. Free barium activates the coagulation system by way of the intrinsic pathway, resulting in severe fibrinous peritonitis that carries a mortality of 53 percent even when surgically controlled.

EVOLUTION OF INTRAABDOMINAL INFECTION

Peritoneal Contamination

When peritoneal injury is accompanied by bacterial contamination, degranulation of peritoneal mast cells dramatically increases the permeability of peritoneal endothelial and mesothelial cells. The resultant exudation of protein-rich fluid contains complement, opsonins, and fibrinogen. Increased vascular permeability also results in a significant fluid influx into the peritoneal cavity. This fluid comes at the expense of depleted intravascular volume, resulting in hypotension.

The systemic response to peritonitis includes the release of catecholamines, glucocorticoids, aldosterone, and vasopressin (antidiuretic hormone). These responses, combined with hypovolemia from the copious fluid sequestration within the peritoneum, can create hemodynamic instability with hypotension, reduced cardiac index, increased peripheral resistance, and increased peripheral oxygen consumption. The patient may manifest the hyperdynamic hemodynamic picture of septic (distributive) shock, i.e., increased cardiac output, decreased peripheral resistance, and decreased arteriovenous oxygen difference. Fluid resuscitation may change the patient's condition from a hypodynamic to a hyperdynamic state.

After a free perforation into the intestinal cavity, the contaminating bacteria migrate with the peritoneal fluid to the subdiaphragmatic area, where large numbers are cleared into the systemic circulation via the subdiaphragmatic lymphatic “pump.” The bacteria empty into the systemic circulation through the thoracic duct and into the left subclavian vein within 6 min of inoculation. More than 90 percent of circulating bacteria are cleared by phagocytosis in the reticuloendothelial system, particularly hepatic Kupffer cells.

Concept of the Decisive Period

After peritoneal contamination the ultimate outcome, resolution, or establishment of infection becomes a race between bacterial multiplication and bacterial destruction by the host defenses. Resident peritoneal macrophages are the initial leukocytic cell on the scene capable of phagocytosing bacteria. Normal peritoneal fluid contains less than 300 cells/mm³, with more than 50 percent macrophages, 40 percent lymphocytes, and 5 to 10 percent mast cells or eosinophils. During the first 1 to 2 h after bacterial contamination, macrophages and mast cells are the only phagocytic host defense cells in the peritoneal cavity. Macrophages are capable of phagocytosis on the peritoneal surface even without prior opsonization. Bacterial opsonization, with IgG or C3b, improves phagocytic efficiency. Prompt clinical recognition and appropriate treatment can drastically shift this dichotomy. Miles and Burke suggested the concept of a “decisive period” for bacterial infection. This period refers to the time required for bacterial numbers in fluid or tissue to exceed 10^5 /mm³ (or per gram of tissue) and establish an infection. Surgeons must deal with the infection before bacterial numbers reach these levels or remove the focus of infection so that after

operation the residual numbers of bacteria are less than $10^5/\text{mm}^3$ (Fig. [32-4](#)).

CARE FOR PATIENTS WITH INTRAABDOMINAL INFECTIONS

Diagnosis of Intraabdominal Infection

The history and physical examination of patients with a suspected intraabdominal infection are pivotal in defining the need for and urgency of intervention and the use of diagnostic techniques. A brief history should determine how long the patient has been ill; whether fever or chills are present; whether pain is present, and if so where it is located, what its character is (crampy, dull, burning, etc.), whether it has changed location, character, or intensity, and whether it is associated with anorexia, vomiting, or ileus. A pertinent past medical history, including recent hospitalizations, medications, chronic disease diagnoses, and prior operations, is important.

Stretching of the visceral peritoneum produces nausea and poorly localized pain. Most intraperitoneal visceral structures arise from the embryologic foregut, and, accordingly their nerve supply arises from the visceral branches of the vagus nerve. Pain usually is localized to the dermatome distribution of the associated visceral organ. For example, the gallbladder arises from the upper thoracic dermatomes; hence, stretching of the visceral peritoneum overlying the gallbladder produces epigastric and right upper back pain. When the inflammatory process involves the adjacent diaphragmatic structures, pain may be perceived in the neck or shoulder (fourth cervical nerve), because the diaphragm arises from the embryologic cervical dermatomes.

Visceral peritoneal irritation, usually from distention of a hollow viscus, causes dull, poorly localized, often periumbilical, and often crampy pain. Visceral structures are generally completely insensitive to surgical transection, cauterization, heat, cold, or pinching. Most symptoms result from inflammation of the visceral peritoneum, which receives afferent innervation only from the autonomic nervous system and is relatively insensitive. Visceral afferent nerves respond primarily to traction or distention, but less well to pressure, and are not capable of mediating pain and temperature sensation. Hence stimuli are perceived as poorly localized, dull discomfort.

The root of the small bowel mesentery and the biliary tree are relatively better innervated than the rest of the visceral peritoneum, accounting for more intense and better localized responses to stimulation of these areas. The term *biliary colic* refers to the sensation transduced by visceral afferents from an overdistended gallbladder undergoing muscular spasm because of an obstructing stone. The small bowel itself is not well innervated, and so stimulation of its visceral afferents produces only a vague, dull ache in the midabdomen. Stimulation of the colonic afferents results in a similar dull pain located in the lower abdomen. Pain may start abruptly with a perforation when a highly irritating agent such as pancreatic or gastric juice diffusely inflames the somatic nerve endings of the parietal peritoneum.

A good example of the effects of the dual mechanism of pain perception is in the progression of symptoms in the development of acute appendicitis. Initially inflammation of the appendix is projected via autonomic nerve fibers to the region of the umbilicus and epigastrium, with symptoms of diffuse epigastric pain, nausea, and vomiting. As the somatic pathways of the parietal peritoneum and adjacent structures become involved, the pain seems to “migrate” from the epigastrium and umbilicus to the right lower abdomen and to grow in intensity, and then to localize at McBurney's point. When fully developed, pain in parietal peritonitis is steady, unrelenting, burning, and aggravated by motion, and usually it is most intense in the region of the most advanced peritoneal inflammation.

Assessment of Hemodynamic and Respiratory Status

There are two potential causes of hypotension in patients with intraabdominal infection. Fluid loss occurs through a variable combination of decreased intake, increased respiratory fluid losses from the tachypnea accompanying fever, and intraabdominal fluid sequestration. The extent of fluid sequestration that accompanies infection is a result of both the local peritoneal response and the diffuse capillary leak that accompanies major sepsis. Hypotension also may be caused by the cardiac and vascular dysfunction that is part of septic shock.

Immediate differentiation between these two distinct entities may not be straightforward. Hypotensive patients should receive boluses of 500 mL lactated Ringer's solution until an adequate blood pressure response occurs. If 3 L have been given and a sustained blood pressure response has not occurred, particularly in an elderly patient, a pulmonary artery catheter should be placed to determine hemodynamic function and to monitor fluid and pressor therapy.

Physical Examination

In addition to a review of the vital signs, the general appearance of the patient, including tachypnea or labored breathing, agitation, or evidence of dehydration, should be carefully noted. Each part of the examination should be explained to the patient before it is performed, with reassurances that it will not be painful. The scleras and conjunctivae may reveal icterus or pallor. The chest should be percussed and auscultated to rule out pneumonia, particularly of the lower lobes, which may produce the abdominal findings. In a hypotensive patient, rolling the patient from side to side rather than sitting him or her up is appropriate. A cardiac examination should rule out congestive failure or valvular disease, since aggressive fluid therapy may be required and because acute congestive failure with hepatic distention and capsular stretching may produce abdominal pain.

Inspection and Auscultation

The abdomen should be inspected for distention and any scars noted. The patient should be assured at this point that the examination will cause minimal discomfort but is needed to determine the need for diagnostic measures and therapy. The patient is asked to point with one finger to the area of most severe pain in the abdomen. Auscultation is begun in the quadrant diagonally opposite to this point. Auscultation is performed to simply determine whether bowel sounds are diminished, normoactive (determined according to when the patient last ate), or hyperactive. Hyperactive bowel sounds suggest an obstructive element either as a primary disease process or as part of a localized inflammatory process. Often an inflammatory focus is partially walled off by adherent small bowel. A local ileus occurs, resulting in a functional obstruction. Abdominal distention is a result of concomitant ileus or ascites. Hyperresonance caused by accumulating gas in the distended intestines usually can be demonstrated by percussion. As inflammation spreads, small-bowel sounds disappear and the abdomen becomes nearly silent.

Palpation

The parietal peritoneum is innervated by somatic and visceral afferent nerves and is quite sensitive. The anterior parietal peritoneum is the most sensitive, and the pelvic peritoneum the least sensitive. The ability to localize an inflammatory stimulus affecting the parietal peritoneum, particularly of the anterior abdominal wall, underlies much of our ability to diagnose acute intraabdominal infection.

In addition to the perception of pain as localized tenderness, local inflammation adjacent to the parietal peritoneum leads initially to protective voluntary muscular guarding. If the inflammatory process continues, reflex muscular spasm of the anterior abdominal wall occurs, progressing to various degrees of rigidity. Gentle palpation of the abdomen, the last step in the examination, is intended to determine whether the pain-producing intraabdominal process has resulted in parietal peritoneal inflammation. Palpation starts at the quadrant most removed from the identified point of maximal pain. The purpose of abdominal palpation is to confirm the location of maximal tenderness and the resting tone in various portions of the anterior abdominal wall. To do this, the examiner begins in a part of the abdomen that is asymptomatic and gently presses until the tone of the underlying musculature is determined. In patients with acute peritonitis, a rigid abdomen will be found with no substantial differences from one quadrant to another. For patients with localized tenderness, and a localized increase in abdominal wall tone, voluntary versus involuntary guarding must be distinguished.

Rigidity of the abdominal muscles is produced by voluntary guarding initially and also by reflex muscle spasm. Reflex muscle spasm can become so severe that boardlike abdominal rigidity is produced and the abdominal wall may become scaphoid, as is typically seen after chemical peritonitis caused by peptic ulcer perforation.

Rectal and Vaginal Examinations

Rectal and vaginal examinations are essential to locate the extent of tenderness and whether a pelvic mass or pouch of Douglas abscess is present. Vaginal examination of the cervix may provide clues to the origin of the inflammatory process.

Laboratory and Other Tests

Leukocytosis is common in acute intraabdominal infection, but the total leukocyte count taken alone without a differential count can be very misleading. A leukocyte count of more than $25,000/\text{mm}^3$ or leukopenia with fewer than $4,000/\text{mm}^3$ are both associated with higher mortality. The differential count provides evidence of the presence of acute inflammation by showing relative lymphopenia and moderate-to-marked leftward shift, even if the leukocyte count is normal or subnormal.

Plain radiographs of the abdomen may reveal free air, a uniform indicator of visceral perforation in the absence of recent prior intervention. Free air may be visible on an upright abdominal or lateral decubitus film if a ruptured hollow viscus is the cause of peritonitis. Air beneath the diaphragm may be noted on radiographs of the chest if the patient remains in an upright position for 5 min or more before the film is made. Other findings from plain radiographs that support the diagnosis of intraabdominal infection include intestinal pneumatosis, bowel obstruction, and a mass effect. There are rare benign causes of pneumatosis. More dramatic but less common findings are air in the portal vein or extraluminal gas collections indicative of an abscess; these radiographic signs are sufficiently specific to justify immediate intervention.

The radiologic picture of intraabdominal infection otherwise mimics that of paralytic ileus. Inflammatory exudate and edema of the intestinal wall produce widening of the space between adjacent bowel loops on a plain film of the abdomen. Peritoneal fat lines and the retroperitoneal psoas shadow may be obliterated because of edema. The fat lines in the pelvis may be obliterated, suggesting fluid in the pelvis.

Preoperative Preparation

Acute perforations of the gastrointestinal tract with peritonitis mandate initial treatment in an intensive care environment. In addition, patients with serious intraabdominal infections often require intervention to control infection.

Perforations of the upper gastrointestinal tract cause impressive physical findings of peritonitis but rarely manifest evidence of septic shock. Conversely, perforations of the colon result in massive bacterial contamination and are accompanied by hypotension. In either setting, progressive clinical deterioration will continue until soiling of the peritoneal cavity is terminated. This puts important temporal limits on resuscitation prior to operative intervention. Patients with acute perforations of the gastrointestinal tract should be sufficiently resuscitated to survive the induction of anesthesia, but resuscitation should continue during operation. Rapid volume loading will counter the vasodilatory effects of anesthetics. More refined parameters of completed resuscitation, such as maximization of cardiac output or oxygen delivery, should await the completion of successful intervention.

In addition to intravascular volume loading, other maneuvers are of value in protecting patients from intraoperative hypotension. Infusing low doses of dopamine improves renal blood flow and reduces the incidence of acute renal failure. Positive end-expiratory pressure (PEEP) should be used sparingly, as it decreases venous return and can compromise cardiac output. High inspired oxygen concentrations maintain adequate arterial oxygen saturation until intravascular volume has been satisfactorily restored.

Preoperative preparation of patients with peritonitis involves fluid resuscitation, administration of antibiotics and oxygen, nasogastric intubation, urinary catheterization, and monitoring of vital signs and biochemical and hemodynamic data. Major metabolic derangements and fluid deficits should be corrected before operation, although an inordinate delay is deleterious to outcome. Oxygen is administered to correct the mild hypoxemia in patients with peritonitis secondary to intrapulmonary shunting and the mechanical impairment of pulmonary ventilation by the distended, tender abdomen. Assessment of respiratory function should include apparent tidal volume and work of breathing. If impairment is suspected, ventilatory volume and arterial blood gases should be measured. Ventilatory support should be initiated whenever any of the following is present: (1) inability to maintain adequate alveolar ventilation as evidenced by a Pa_{CO_2} of 50 mmHg or greater; (2) hypoxemia, with a Pa_{O_2} below 60 mmHg and a fraction of inspired oxygen of 1.0; or (3) shallow, rapid respirations, muscle fatigue, or the use of accessory muscles of respiration.

Vital signs—temperature, blood pressure, pulse rate, respiration rate—are continuously recorded. Preoperative biochemical evaluation should include measurement of serum levels of electrolytes, creatinine, glucose, bilirubin, and alkaline phosphatase, and a urinalysis.

Nasogastric intubation is performed to evacuate the stomach, to prevent further vomiting, and to reduce accumulation of additional air in the paralyzed bowel. Urinary catheterization is used to record initial bladder urine volume and to monitor subsequent urinary output. Renal dysfunction requires early corrective measures. Patients with peritonitis develop renal failure as a result of hypovolemic shock, septic shock, increased intraabdominal pressure, and nephrotoxic drugs such as the aminoglycosides. Diuretics are used but may be toxic to tubular cells. If there is cardiopulmonary dysfunction, a Swan-Ganz catheter allows measurement of pulmonary capillary wedge and pulmonary artery pressures, systemic vascular resistance, and cardiac output as indices for pharmacologic intervention.

Once a decision to operate has been made, pain should be relieved with potent narcotics. Morphine is preferred, and should be given intravenously in doses of 1 to 3 mg, repeated every 20 to 30 min or even more frequently, to maintain comfort. Small intravenous doses of morphine are safer than the traditional larger intramuscular doses, particularly in unstable patients whose absorption of intramuscular medication may be unpredictable and variable.

Administration of pharmacologic doses of steroids to septic patients has been advocated in the past, but controlled trials have not shown this treatment to be efficacious. Recent clinical studies support the concept that large doses of steroids may delay mortality in sepsis, particularly if given before the onset of shock or very soon after, but they do not alter the eventual outcome. Steroids are not recommended in the management of bacterial peritonitis.

Patients with hyperthermia or hypothermia should have their temperature corrected toward normal before operation. Administration of acetaminophen is often effective in reducing fever. If not, or if the patient is hypothermic, a cooling or warming mattress should be used.

MANAGEMENT OF DIFFUSE PERITONITIS

The history and physical examination usually are sufficient to make the diagnosis, and no further testing is necessary. Even if infection spreads diffusely throughout the peritoneal cavity, if the source is controlled by early surgical intervention, peritonitis usually responds to vigorous antibiotic and supportive therapy. On the other hand, death nearly always follows continued peritoneal soiling.

Operative management of peritonitis involves immediate evacuation of all purulent collections, with particular attention to subphrenic, subhepatic, interloop, and pelvic collections. The perforated bowel should be resected. Evidence has accumulated from studies of several decades of mortality rates following surgical treatment of perforated diverticulitis that resection with end colostomy significantly decreases acute mortality as compared to transverse loop colostomy and drainage. Primary anastomosis is not recommended in the face of purulent peritonitis; in one study, mortality from primary anastomosis was 23 percent, with most deaths attributed to anastomotic leakage. Controversies in the operative management of peritonitis focus primarily on wound closure techniques and scheduled relaparotomies. Patients with diffuse peritonitis secondary to colonic perforation or anastomotic dehiscence typically develop abdominal wall edema as part of a generalized syndrome of increased capillary permeability. This syndrome is exacerbated by the aggressive restoration of intravascular and extravascular volume. Under these circumstances, primary closure of the abdominal incision may be difficult or even unwise, as increased intraabdominal pressure can result in compression of mesenteric and renal veins, leading in some instances to acute renal failure or bowel necrosis. A fascial prosthesis prevents an overly tight closure. A variety of materials have been employed, including Marlex, Silastic, and polytetrafluoroethylene (PTFE). Each material has its own virtues and problems. Impermeable materials can exacerbate peritonitis and should be used only if a relaparotomy is planned.

Recurrent abscesses and persistent peritonitis largely account for the high mortality in patients with diffuse peritonitis. A continued septic state with renal and/or hepatic dysfunction should raise suspicion of diffuse peritonitis, but there is little satisfaction with current approaches for operative management. A variety of mechanical approaches have been proposed to reduce mortality, generally based on the notion that failure of host defense mechanisms of the peritoneal cavity leads to persistent infection. Planned relaparotomy at present seems most attractive. Patients treated in this fashion undergo standard operative management of their infection, but the fascia is not closed; instead, a prosthetic material is sewn to the fascia as a temporary closure, and at intervals of 24 to 48 h the mesh is opened and the peritoneal cavity debrided and irrigated. Early reports were quite positive, but used historical controls. More recent reports have not found that survival is substantially improved or that the incidence of late abscess formation is reduced. Occasional fistulas have occurred.

We have used planned relaparotomy in conjunction with prosthetic fascial closure for a relatively small number of patients with septic shock and colon-derived peritonitis. The obvious benefits are absence of intraabdominal and wound tension and the ability to obtain sequential microbiologic data to guide antimicrobial therapy. Once the prosthetic mesh is in place, relaparotomy is performed on a scheduled basis, or as dictated by the patient's clinical course and adjunctive tests, particularly computed tomography. The latter, more conservative approach seems worthwhile for patients who recover from their initial episode of shock and who do not manifest progressive organ failure. In such patients, abdominal wall and visceral edema typically resolves over the first week, at which time definitive closure of the abdomen is performed.

Management of Specific Forms of Peritonitis

Perforation of the Stomach and Duodenum

Infection after peptic ulcer perforation presents acutely; the patient is commonly able to give the exact time at which the perforation occurred. This form of peritonitis is initially chemical, but with time it becomes infected. The patient usually seeks help early because of the severe pain, and operative repair often is possible before significant bacterial overgrowth has occurred. The proper management is simple closure of the perforation; some authors have recommended definitive treatment of the ulcer disease in addition if the perforation is less than 12 h old. Parietal cell vagotomy and a Graham patch of the perforation is favored by many.

Small-Bowel Perforation

Infection after small-bowel perforation falls into two categories. In the first, bowel obstruction is present, so that ileus precedes peritonitis. Colic and other features of bowel obstruction are the presenting signs and symptoms, gradually changing to those of localized or diffuse peritonitis, with fever and leukocytosis after perforation. In the second form, bowel wall necrosis as a result of inflammation or impaired vascular supply leads to perforation. Peritonitis may be diagnosed at a late stage because of the lack of initial symptoms, and hence mortality rates can exceed 50 percent.

Colon Perforation

About 20 percent of cases of peritonitis have their origin in the colon. More than half of these are due to inflammatory disease such as diverticulitis and other forms of colitis. The remaining cases are due to perforation proximal to or at stenosis caused by luminal bowel obstruction (tumor) or external bowel obstruction such as incarcerated hernia, intussusception, and volvulus. In colonic obstruction, the colon distends proximally and ruptures as a result of either increased intraluminal pressure or diminished vascular supply to one segment. A malignant growth usually does not cause peritonitis directly but may lead to bowel obstruction with either perforation of dilated segments or bowel ischemia, and/or bacterial migration through the necrotic bowel wall. Carcinoma of the colon, if the ileocecal valve is competent, may lead to diffuse fecal peritonitis as a result of rupture of the cecum or ascending colon.

A ventral midline incision is preferred, as it is quick to open and close and allows easy exploration of the entire abdomen, especially in cases in which the organ source of peritonitis is not clear. Complete control of the source of infection is essential for therapeutic success. The elimination of the infectious source or focus reduces the infectious inoculum to a magnitude that is treatable successfully by antibiotic therapy coupled with active intrinsic host defense mechanisms. When the infectious focus cannot be adequately controlled, the prognosis is poor.

Peritonitis of Genitourinary Origin

Ruptured perinephric abscess and ruptured chronic cystitis after radiation therapy for female reproductive tract cancer can cause peritonitis. A colon anastomosis should be avoided in peritonitis associated with a urine leak. Pelvic peritonitis due to sexually transmitted infection is common in young women; usually there is acute, severe lower abdominal pain, typically just cephalad to the inguinal creases. The condition is diagnosed by Gram stain of cervical discharge. Treatment with antimicrobials is successful in most cases, and operation is reserved for the occasional patient with persisting tubo-ovarian abscess.

Postoperative Peritonitis

Postoperative peritonitis is usually the result of an anastomotic leak and is often discovered only after some delay, typically between the fifth and seventh postoperative day. Delay contributes to a high mortality rate. After duodenal operations, the high mortality of leakage is explained by the fact that the duodenum is fixed in the retroperitoneum and cannot be exteriorized, and the source of infection often cannot be adequately controlled or closed. Consequently, infective material and proteolytic enzymes continue to soil the peritoneal cavity, sustaining the infection. Drainage, controlled fistula formation, exteriorization, repair, or resection and reanastomosis are performed as indicated.

Peritonitis After Trauma

Peritonitis may develop after blunt trauma in patients who have unrecognized intraabdominal injuries such as a ruptured mesentery with loss of the vascular supply to the small or large bowel, or a frank unrecognized bowel perforation. This type of intraabdominal infection usually is severe because the diagnosis is delayed or masked by other injuries elsewhere in the body, particularly head injury. Management in such instances is no different from that of other forms of peritonitis.

Patients suffering penetrating trauma, typically gunshot wound or stab wound, usually are operated on, and the wound is repaired or controlled immediately; these injuries lead to peritonitis only when the initial repair leaks. Perforation of a bacteria-containing hollow viscus by blunt trauma is less frequent but more of a clinical diagnostic problem since symptoms can be masked by pain from the injury. Contamination of the abdominal cavity seen after penetrating abdominal trauma is usually not considered an intraabdominal infection. Only one-third of patients with penetrating trauma to the colon actually sustain documented contamination of the peritoneal cavity.

Cleaning the Abdominal Cavity

Various approaches to cleaning the abdominal cavity include suction, swabbing, lavage, radical debridement, and postoperative irrigation. There is broad agreement about the utility of suction and swabbing to remove pus and necrotic material, followed by local irrigation/lavage and limited debridement—measures typically used in every case. Controversy remains over the efficacy of high-volume lavage, radical debridement, and postoperative irrigation.

Intraoperative High-Volume Lavage

This treatment involves extensive intraoperative lavage of the entire abdominal cavity. Initially, 1 to 1.5 L of physiologic saline or Ringer's solution is placed into the abdominal cavity. The viscera are manipulated so that the fluid reaches all parts of the abdominal cavity to wash out pus, feces, and necrotic material, and then the fluid is suctioned away. This procedure is repeated until the suctioned fluid runs repetitively clear. The total volume of lavage should be at least equal to the estimated volume of the peritoneal cavity, which in adults is 8 to 12 L.

Povidone-Iodine

This should not be used in peritoneal lavage fluid. Iodine is highly toxic to the mesothelial cells of the peritoneum and it kills the phagocytic cells needed to clear bacteria. In animal experiments, use of povidone-iodine increases mortality, and in a controlled clinical trial the outcome with povidone-iodine was worse than with saline.

Radical Debridement

This is a poorly defined technique because it is almost impossible, without causing significant bleeding, to radically debride all necrotic tissue in intraabdominal infections, including intraabdominal abscesses. This may be one explanation for there being no significant outcome advantage in using radical debridement over other techniques. Surgical judgment is the critical element, since the drawbacks of radical debridement must be balanced against the fact that bacteria may be trapped in fibrin, and that necrotic tissue is an excellent medium for bacterial growth. Removal of such foci is the primary goal of the operative procedure, but given the adjuvant effect of hemoglobin, the procedure should not cause additional intraperitoneal hemorrhage.

Continuous Postoperative Irrigation

Four to six Tenckhoff catheters are placed in the abdominal cavity and irrigation fluid infused for 1 h. The abdominal cavity is drained, and the cycle is repeated. The drains often occlude, and there is no way to ensure that all diseased areas of the abdominal cavity are lavaged. Drains also may erode into the intestine to cause new complications. Controlled studies comparing continuous postoperative irrigation to standard therapy have yielded differing results, and in general the numbers of cases examined are small and do not permit any firm conclusions about efficacy.

OTHER FORMS OF PERITONITIS

(Table 32-6)

Primary Peritonitis

Primary peritonitis is an inflammation of the peritoneum from a suspected extraperitoneal source, often via hematogenous spread. It occurs in children and in adults and can be a life-threatening illness, particularly in patients with cirrhosis. The spectrum of bacteria causing this syndrome and the population primarily affected have changed over recent decades. Spontaneous bacterial peritonitis is now more common in adults than in children and shows no differential sex incidence. Children with nephrosis, formerly the group most commonly affected, have been replaced by adults with cirrhosis or systemic lupus erythematosus.

Spontaneous peritonitis in adults is seen most commonly in patients with ascites and is a monomicrobial infection (i.e., only a single species of bacteria is present), in contrast to the polymicrobial infection of typical suppurative (secondary) peritonitis. Although pneumococci formerly were the most frequent infecting organisms at all ages, coliforms now are the chief pathogens in adults, accounting for 70 percent of infections, with *E. coli* being the most common isolate. Gram-positive cocci account for 10 to 20 percent of cases, and anaerobes are seen in 6 to 10 percent. In children the common pathogens are hemolytic streptococci and pneumococci. Two peaks of incidence are characteristic, one in the neonatal period and the other at age four to five. There may be a history of previous ear or upper respiratory infection. Children with nephrotic syndrome and systemic lupus erythematosus are disproportionately affected.

The route by which bacteria are transmitted to the peritoneal cavity is not known. The hypothesis of transmural migration of bacteria from the intestine is supported by the presence of endotoxin in ascitic fluid and blood in many cirrhotic patients with decompensation, even though the bacteria themselves are not present in the ascitic fluid and patients have no clinical evidence of peritonitis. The frequent clinical association of spontaneous peritonitis with urinary tract infection harboring the same organisms, and cases of simultaneous spontaneous pulmonary empyema and pericarditis associated with spontaneous peritonitis, favor the hypothesis that transmission is hematogenous.

Clinical symptoms usually are of short duration in children; onset is more insidious in ascitic adults. Most patients complain of abdominal pain and distention; vomiting, lethargy, and fever are more prominent in children. Diarrhea is typical in neonates but seldom seen in adults. Bowel sounds are variable. Free air usually is not seen on abdominal radiographs. The clinical picture may be nonspecific.

Paracentesis is the most useful diagnostic test. Fluid is examined for neutrophil cell count and pH; a Gram stain should be done and a specimen sent for culture. The neutrophil cell count has the highest sensitivity and specificity in making the diagnosis; a neutrophil count greater than 250 cells/mm³ is positive. Studies of ascitic fluid in patients with suspected spontaneous peritonitis have shown that a wide arterial–fluid pH gradient also correlates with a positive diagnosis. Ascitic fluid pH is low in spontaneous bacterial peritonitis, while in sterile ascitic fluid the pH is the same as in serum. In only one-third of patients with positive fluid cultures are organisms seen on a stain of centrifuged peritoneal fluid, which is a reflection of the low concentration of bacteria in many cases of primary bacterial peritonitis. If the stain shows only gram-positive cocci, spontaneous peritonitis is strongly suggested. If a mixed flora of gram- positive and gram-negative bacteria is present, intestinal perforation is more likely and exploratory laparotomy is necessary. The presence of only gram-negative bacteria on the stain is consistent with either primary or secondary bacterial peritonitis.

When the diagnosis of spontaneous bacterial peritonitis is confirmed, antibiotic therapy should be started and the patient initially managed nonoperatively. Usually a cephalosporin or ampicillin-sulbactam is appropriate, since 90 percent of the organisms causing spontaneous peritonitis are sensitive to these antibiotics.

Peritonitis Related to Peritoneal Dialysis

Peritonitis is the dominant complication of *continuous ambulatory peritoneal dialysis (CAPD)* in patients with end-stage renal disease. Peritonitis occurs more frequently with CAPD than with intermittent peritoneal dialysis. Catheter-related infection is the most common mechanism, with no correlation with catheter location. Other causes of peritonitis in CAPD patients are tunnel infections and cuff extrusion. The bacteriology and treatment of dialysis-related peritonitis are considerably different from those of patients with other types of peritoneal infections. The incidence of dialysis-related peritonitis varies from center to center, but an overall average of 1.3 episodes per patient per year is estimated. In contrast to other causes of intraabdominal infection, dialysis-related peritonitis usually is caused by a single organism, and fungal infections are more common in this group. Two-thirds of patients with positive cultures have a gram-positive coccus as the causative organism, usually *Staph. aureus* or *Staph. epidermidis*. Gram-negative bacteria usually are found only in patients with recurring episodes of peritonitis, which may reflect antibiotic administration for prior episodes. Patients with dialysis-related peritonitis rarely have a positive blood culture, in contrast to patients with other causes of peritonitis. Anaerobes rarely are recovered. The presence of anaerobes or a mixed flora suggests intestinal perforation or other intraabdominal disease (e.g., diverticulitis or cholecystitis). Yeasts and tubercle bacilli are rare. Culture-negative peritonitis in dialysis patients ranges from 10 to 27 percent. *Pseudomonas aeruginosa* has been reported to cause around 3 percent of cases and generally cannot be cured by antibiotic therapy alone; the dialysis catheter must be removed in these patients.

Turbidity of the dialysate is the earliest sign, and the only finding in one- fourth of cases. If accompanied by abdominal pain or fever, prompt diagnostic laboratory studies and therapy are required. The diagnosis is established when any of the following is present: (1) positive culture from the peritoneal fluid; (2) a cloudy dialysate effluent; and (3) clinical signs of peritonitis. The stain of dialysis effluent is frequently negative in dialysis catheter-related peritonitis. Attention to sterile technique, use of topical povidone-iodine about the catheter wound, and use of sterile dialysate prepackaged in plastic bags all reduce the incidence of infection.

The initial treatment of dialysis-related peritonitis is administration of antibiotics and heparin in the dialysate as well as an increase in the dwell time of dialysate fluid. The indications for catheter removal include persistence of peritonitis after 4 to 5 days of treatment, the presence of fungal or tuberculous peritonitis, fecal peritonitis, or a severe skin infection at the catheter site.

Patients with a chronically implanted peritoneal catheter may develop sclerosing peritonitis, with a reported incidence of 1 to 6 percent. The cause is unknown, although a relationship with antecedent infection is possible. Affected patients usually have had frequent previous bouts of peritonitis. Treatment is directed at such complications as bowel obstruction.

A related form of peritonitis, due to infection of a ventriculoperitoneal shunt, exhibits clinical signs and symptoms similar to those of peritoneal dialysis catheter peritonitis. Peritonitis should be suspected in any patient having such a shunt who develops any abdominal symptoms.

Tuberculous Peritonitis

Tuberculous peritonitis, formerly frequently fatal as a manifestation of uncontrolled generalized tuberculosis, now is undergoing a renaissance in Europe and North America because of the increasing prevalence of acquired immunodeficiency syndrome (AIDS) due to human immunodeficiency virus (HIV) infection and of other immunocompromised states. It has always been a serious problem in India, Southeast Asia, Africa, and Latin America. Some cases are the result of reactivation of latent peritoneal tuberculosis established by hematogenous spread from a pulmonary focus during an earlier episode of acute disease. The tubercle bacillus presumably gains entry to the peritoneal cavity transmurally from diseased bowel, from the fallopian tube in tuberculous salpingitis, from the kidney in nephritis, or from some other distant source. Clinically, most patients lack an obvious source; however, a primary focus can be identified in all patients at autopsy. Patients usually have a positive tuberculin skin test.

Clinical Manifestations

Clinically, tuberculous peritonitis is insidious, presenting with fever, anorexia, weakness, and weight loss. Some ascites almost always is present, and more than half of affected patients have dull, diffuse abdominal pain.

In the past the disease was classified into “moist” and “dry” phases. The moist phase refers to the early, subacute stage with fever, ascites, abdominal pain, and weakness. Ascites is progressive and may become massive. The dry form presents later without ascites, following resolution, during which dense adhesions are formed. On examination, the abdomen is somewhat tender, but the classically described “doughy abdomen” is rarely seen today. Clinical manifestations of generalized tuberculous infection are seen in about one-third of

patients and include anorexia, weight loss, and night sweats.

Diagnosis

A peritoneal fluid tap will show mostly lymphocytes. Tubercle bacilli can be retrieved from ascitic fluid 80 percent of the time if more than 1 L of fluid is cultured. The ascitic fluid has an increased protein concentration, lymphocytic pleocytosis, and a glucose concentration below 30 mg/dL. If the fluid tap analysis does not establish the diagnosis, peritoneoscopy and direct biopsy of the peritoneum are recommended. On laparoscopy, the appearance of tuberculous peritonitis includes stalactite-like fibrinous masses hanging from the parietal peritoneum in the lower part of the abdomen. A directed, percutaneous needle biopsy specimen of a granulomatous lesion as well as samples of peritoneal fluid for direct smear examination and injection into a guinea pig should be obtained. As a last resort, exploratory laparotomy may establish the diagnosis. At laparotomy a peritoneal biopsy should be taken; the placement of drains or exteriorization of bowel should be avoided.

Treatment

Treatment is administration of antituberculous drugs. Given reports in recent years of resistance to commonly used tuberculostatic drugs, triple antituberculous drug therapy is instituted early in the course of the disease, generally with good results. Therapy should be continued for at least 2 years after the patient becomes asymptomatic. Tuberculous peritonitis may heal with formation of dense fibrous adhesions; treatment with prednisone during the initial few months of antituberculous drug therapy may reduce adhesion formation and subsequent obstruction.

Operation should be reserved for diagnosis if needle biopsy fails, or for the treatment of such complications as fecal fistula. Because tuberculous peritonitis heals with formation of dense fibrous adhesions, patients with this disease are prone to developing an intestinal obstruction.

Typhoid Perforation

Typhoid perforation usually is seen in the third week of infection with *Salmonella typhi* in patients with acute disease. The disease is rarely seen in Europe or North America but is endemic in regions with poor hygienic conditions and water contamination, such as India, Pakistan, and countries in South America and Africa. Typhoid bacilli are thought to penetrate the Peyer's patches of the intestine wall, mainly in the distal ileum. These collections of lymphoid cells hypertrophy, leading to hemorrhage and then perforation.

Perforation often is not appreciated in an already severely diseased patient, and it is superinfection resulting from leakage of intestinal bacteria that leads to the full-blown picture of suppurative bacterial peritonitis. Treatment consists of closure of the punched-out lesions and evacuation of pus from the peritoneal cavity. Staged abdominal repair may be indicated if abdominal toilet is incomplete during the first operation. Trimethoprim- sulfamethoxazole is the treatment of choice for uncomplicated typhoid, but in patients with peritonitis cefotaxime plus metronidazole is preferable.

Amebic Perforation

Entamoeba histolytica infection of the intestine usually causes a dysentery-like illness, but sometimes liver abscess or perforation of the large bowel occurs. Liver abscesses also can perforate secondarily and cause diffuse peritonitis. The clinical picture is that of bacterial peritonitis. Treatment consists of resection of the diseased bowel segment with anastomosis, application of the general principles of treating peritonitis, and the administration of metronidazole in combination with a third- generation cephalosporin.

Aseptic/Sterile Peritonitis

This form of peritonitis develops from irritant material (talc or foreign bodies) that has gained entry to the peritoneal cavity. Aseptic peritonitis can be caused by a variety of substances. They all act as adjuvants when secondary bacterial contamination occurs, promoting proliferation of microorganisms and transition to suppurative peritonitis. The soilage initially is sterile or nearly so, which distinguishes chemical peritonitis at onset from suppurative peritonitis. In chemical peritonitis following peptic ulcer perforation, cultures are initially sterile, but as time progresses, positive cultures are recovered. After 24 h, all cultures are positive and the patient has developed suppurative peritonitis. A similar shift from initially aseptic to bacterially infected peritonitis can occur with all forms of chemical peritonitis.

Foreign-Body Peritonitis

Foreign bodies may be deposited in the peritoneal cavity during operations (sponge or instrument inadvertently left behind) or may result from penetrating injuries or perforation of the intestines following ingestion. A larger foreign body can lead to the formation of an abscess in the presence of bacteria, but otherwise foreign bodies are sealed off and encapsulated.

Periodic Peritonitis

Recurrent episodes of abdominal pain, fever, and leukocytosis occur in certain population groups in and around the Levant, notably in Armenians, Jews, and Arabs. The disease appears to be familial. The major point for the surgeon is that laparotomy is not required in these episodes. Laparotomy is often performed for the first episode, since an acute intraabdominal process requiring surgical cure cannot be ruled out. At operation, the peritoneal surfaces may be inflamed and there is free fluid, but no bacteria. The appendix should be removed even if normal to eliminate the possibility of acute appendicitis in the differential diagnosis of future episodes. Colchicine is effective in preventing recurrent attacks, and a favorable response to chronic administration of colchicine is a definitive diagnostic test.

Drug-Related Peritonitis

Administration of isoniazid and erythromycin estolate have been reported to cause acute abdominal symptoms mimicking peritonitis, but not development of true peritonitis. A number of cases have been reported in which beta-blocking drugs have resulted in striking thickening of the visceral peritoneum. The most frequent clinical presentation is as a typical small- bowel obstruction, often subtle at onset, associated with weight loss and with an abdominal mass on physical examination. The operative findings are striking: the whole small bowel usually is caught up in a thick sac, which sometimes can be lifted as a single mass from the peritoneal cavity. The agglomeration of the small bowel produces the mass that is palpable preoperatively.

Lead Peritonitis

Lead peritonitis has the same clinical picture as acute intermittent porphyria. It is associated with lead intoxication (occurring in painters, smelter workers, pica in children), and a careful history will lead to the correct diagnosis. Unlike true intraabdominal infection, in this disorder sudden stretching of the peritoneum by rebound palpation usually does not cause peritoneal irritation and guarding.

Hyperlipemic Peritonitis

Abdominal pain mimicking peritonitis may be seen in patients with type I and type V hyperlipoproteinemia, a group of heterogeneous disorders resulting from increased concentration of chylomicrons or very-low-density lipoproteins in the blood. These disorders usually lead to pancreatitis. If erroneously operated on during early stages, the abdominal cavity is found to be full of chylous milky material. A careful family history will clarify the differential diagnosis.

Porphyric Peritonitis

This is a condition of perceived abdominal pain rather than an inflammation of the peritoneum. It is seen in patients with acute intermittent porphyria, who suffer from acute attacks that cause nervous system damage, especially to the autonomic system. The pain may be localized or general and is often accompanied by vomiting and constipation. It may be colicky or constant and is associated with abdominal tenderness. With low-grade fever and mild leukocytosis the disorder further imitates intraabdominal infection. The diagnosis is established by the demonstration of porphobilinogen in the Watson-Schwartz test.

Talc Peritonitis

Peritoneal inflammation, exudation and formation of pseudotumors (chronic inflammatory omental tumors), and formation of dense adhesions may follow contamination of the peritoneal cavity by glove lubricants (talc, lycopodium, mineral oil, cornstarch, rice starch) or by cellulose fibers from disposable gauze pads, drapes, and gowns. The reaction, particularly that to rice starch, is largely a hypersensitivity response.

The clinical features include migratory abdominal pain, fever, physical signs of peritonitis, and, often, the presence of an abdominal mass, all developing within 3 weeks after an otherwise uncomplicated abdominal operation. The surgical wound may appear normal or may be moderately indurated. Plain abdominal films are nonspecific. The total white blood cell count is normal; eosinophilia of 4 to 9 percent may be present.

When the diagnosis remains unclear, laparoscopy may be helpful. If talc peritonitis is recognized, reoperation may be avoided and corticosteroids or nonsteroidal anti-inflammatory agents administered. Eventually the peritonitis resolves. If laparotomy is undertaken because the diagnosis is obscure, a thickened peritoneum studded with white nodules is found. Histologically, nodules that contain starch granules are doubly refractive under polarized light and are surrounded by granulomatous foreign-body inflammation.

The cornerstone of management of this problem is prevention. Talc should not be used in glove manufacture or as a glove lubricant. Current techniques of wiping and washing gloves do not remove all the starch; nonetheless, gloves should be washed or wiped before gloved hands are put into the peritoneal cavity, and care should be taken to avoid spillage of glove contents should a glove be torn during an operation. Sodium bicarbonate has been used successfully as a gloving agent, but it requires special sterilization measures. Silicones also have been suggested.

Tertiary Peritonitis

Patients in whom peritonitis and sepsis initially have been controlled operatively and in whom bacteria have been eliminated by successful antibiotic therapy may progress to tertiary peritonitis, a state in which host defense systems produce a syndrome of continued systemic inflammation.

The clinical picture is one mimicking occult sepsis, as manifested by a hyperdynamic cardiovascular state, low-grade fever, and general hypermetabolism. The patient has the clinical picture of sepsis without the presence of a focus of infection. Such patients sometimes are subjected to laparotomy in an attempt to provide drainage of anticipated recurrent or residual collections of infected fluid. On operation no pathogens are present. Empiric anti-infective therapy is of no value.

DIAGNOSTIC IMAGING FOR SUSPECTED INTRAABDOMINAL INFECTIONS

In the absence of physical findings of diffuse peritonitis, diagnostic imaging with computed tomography (CT) or ultrasonography should be routinely performed in seriously ill patients with suspected intraabdominal infection. The urgency of investigation is dictated by the degree of hemodynamic instability present. Most patients should be evaluated within hours of presentation. Interventional radiology has replaced operative treatment for many localized processes, including diverticular abscesses.

CT is the single best modality for fully evaluating the extent of disease. Ultrasonography is also quite versatile and has the added advantage of being portable, allowing certain procedures to be performed at bedside. Ultrasonography is limited by bowel gas, body habitus, and a lower sensitivity for retroperitoneal processes and parenchymal infection. Usually the choice of modality is based on the experience and preference of the interventional radiologist.

When feasible, nonoperative (i.e., percutaneous) drainage of purulent collections is preferable to open surgical intervention, because of the initial deterioration that almost always occurs after operative manipulation in the presence of intraabdominal infection. The exact basis for this is unclear, but a substantial proportion of patients undergoing emergency operation for intraabdominal infection suffer acute hemodynamic compromise in the early postoperative period. When used for appropriate indications, *percutaneous abscess drainage (PAD)* is at least as effective as operation and is associated with less morbidity. Percutaneous drainage of an intraabdominal abscess usually is successful if the following criteria are met: (1) there is a well-established, unilocular fluid collection; (2) a safe percutaneous route of access is available; this often means that the abscess is adjacent to the body wall; (3) joint evaluation by a surgeon and a radiologist is done so that correct judgments and decisions are made; and (4) there is immediate operative backup available in case of failure or complications.

OPERATIVE MANAGEMENT OF ABSCESES

The indications for open surgical drainage are failure of percutaneous drainage, inability to safely drain percutaneously, the presence of a pancreatic or carcinomatous abscess, an association with a high-output bowel fistula, the involvement of the lesser sac, or the presence of multiple, isolated interloop abscesses. Additionally, exploration and open drainage are undertaken whenever the presence of an abscess is suspected clinically but cannot be localized by CT or ultrasonography. Abscesses in the pelvis usually are drained directly through the rectum or vagina, obviating the need for either percutaneous drainage or an abdominal procedure.

If the abscess is in contact with the abdominal wall or the diaphragm, a direct and, when possible, extraperitoneal approach is preferred. Subphrenic abscesses on either side and right subhepatic abscesses are best approached via the lateral extraserous route. A left subphrenic abscess also can be drained posteriorly through the bed of the twelfth rib. The point at which the abscess is encountered during dissection is usually much deeper than expected.

A lesser sac abscess is approached directly via an upper abdominal incision. Drains are placed dependently; sump-suction drains are frequently helpful. Interloop abscesses are explored through a midabdominal incision. Each abscess cavity is thoroughly debrided, but drains are not generally used. If exploration of the abdomen is being done to find an occult abscess, a midline approach is more efficient, but a transverse incision may be better tolerated. The initial exploration should be limited to the area either above or below the transverse colon and mesocolon, the choice depending on clinical suspicion of the origin of infection.

Confirmation of the presence of pus is obtained by needle aspiration. It is essential to obtain specimens of the abscess content for stain, aerobic and anaerobic culture, and sensitivity studies. All of the abscess contents should be evacuated by suction. The cavity should be thoroughly explored digitally and all loculations within it broken down to create a single residual space. The cavity is irrigated and debrided of nonviable tissue. Multiple drains then should be brought from the abscess cavity to the exterior as directly and dependently as possible. If the abscess cavity is particularly large, or if thorough dependent drainage cannot be established, sump-suction drains should be employed in addition.

After percutaneous or open drainage of an abscess, the drains are left in place until external drainage stops or is clear. The drainage tract should then be irrigated and a sinogram obtained to verify collapse of the cavity before the drains are moved. It may take 2 to 3 weeks for a large cavity to become small enough to permit drains to be slowly advanced, allowing the drainage tract to seal as they are withdrawn.

Left Subphrenic Abscess

Formerly uncommon, these are now the most common variety of upper abdominal abscess after peritonitis or leakage from a viscus. Left subphrenic abscess also follows splenectomy, particularly when the splenic fossa is drained, and also is a consequence of pancreatitis. Unlike the subdiaphragmatic space on the right, which is divided by the liver into suprahepatic and subhepatic spaces, on the left side all these spaces are contiguous. The physical signs are costal tenderness on the left, sometimes pain in the shoulder (Kehr's sign), the presence of a left pleural effusion, and limitation of diaphragmatic motion noted by physical examination or radiography.

Lesser Sac Abscess

Technically the lesser sac abscess is a variety of left subhepatic-subphrenic abscess, since the lesser peritoneal sac anatomically is a portion of the left subhepatic space. However, the anatomic features of this abscess—and the surgical maneuvers required for drainage—are distinctly different. Lesser sac abscess is an unusual complication of diseases of the stomach, duodenum, and pancreas. The most common cause is a pancreatic abscess or a secondarily contaminated pancreatic pseudocyst that involves the lesser sac by direct extension. Perforation of a gastric ulcer or, less commonly, of a duodenal ulcer, and occasionally rupture of an ulcerating malignant gastric tumor, also may result in formation of a lesser sac abscess.

Diagnosis of a lesser sac abscess can be difficult, since much of this space is overlapped anteriorly by the left lobe of the liver. Tenderness to palpation in the midepigastrium usually is present but is a nonspecific sign. Ultrasonography is a most useful diagnostic test. Radiographic studies may show displacement of the stomach; occasionally plain radiographs show fine gas bubbles within the lesser sac, indicative of the presence of purulent material.

Lesser sac abscess has been associated with a poorer prognosis than abscesses located in other portions of the abdominal cavity. This abscess must be drained adequately; the location often dictates the use of suction drains; dependent drainage may not be easily established.

Right Subphrenic Abscess

The right subphrenic space is a potential space between the liver and the diaphragm, above the posterior attachments of the coronary and triangular ligaments to the diaphragm and posterior body wall, and extending anteriorly to the costal margin. Because the potential space is so limited in the vertical dimension, a right subphrenic abscess tends to be localized to only a portion of the total potential space, loculating either anteriorly or posteriorly. Abscesses within this space are most frequently secondary to rupture of a hepatic abscess and sometimes to operations on the stomach or duodenum.

Signs and symptoms of a right subphrenic abscess may be quite minimal. Pain is occasionally reported in the upper part of the abdomen or the lower part of the chest, sometimes referred to the back or to the right shoulder. Chest radiographs show a pleural effusion or platelike atelectasis in the right lower lung in 90 percent of patients, and the diaphragm is elevated and shows reduced motion on inspiration in sniffing or physical examination in two-thirds of patients. An air fluid level can be demonstrated in about one-fourth of those affected and establishes the diagnosis. An abscess in the right subphrenic space cannot be drained through a posterior approach unless the route of drainage transgresses the pleural space, a step that is not desirable.

Right Subhepatic Abscess

The right subhepatic space lies under the liver, bounded inferiorly by the hepatic flexure and the transverse mesocolon, medially by the duodenum and the hepatoduodenal ligament, and laterally by the body wall. The most posterior (deepest) part of this space is Morison's pouch. A gastric procedure, especially an emergency operation for complications of ulcer disease, is the most common antecedent event. Biliary tract procedures are second in etiologic frequency. Appendicitis has declined in importance and accounts for only 8 percent of right subhepatic abscesses. Complications of colonic surgery are increasing in importance. A right subhepatic abscess usually produces some tenderness in the right upper quadrant, and the patient may complain of pain, particularly exacerbated by coughing or similar activities that produce visceral motion in the region.

Interloop Abscess

Interloop abscesses often are multiple abscesses that arise as loculations between loops of bowel, mesentery, abdominal wall, and omentum. The transverse mesocolon acts as a barrier to superior extension so that interloop abscesses usually do not involve the upper part of the abdomen. They are commonly associated with a simultaneous abscess in the pelvis.

There are no reliable symptoms or signs. A huge abscess containing more than 1 L of pus may occur without any significant physical findings. Occasionally an interloop abscess may produce a palpable, enlarging abdominal mass. The possible presence of an interloop abscess must be suspected if there has been a preceding episode of peritonitis with incomplete clinical resolution. Occasionally abdominal films show edema in the wall of loops of bowel involved in the loculation or separation or fixation of involved structures. CT examination is the most reliable diagnostic tool.

Halasz emphasized the serious problem of failure to recognize synchronous or multiple abscesses and recommended transperitoneal exploration as preferable to more limited exploration. In his series of 43 patients, one-fourth had synchronous abscesses elsewhere in the abdomen. Failure to find and drain the concomitant abscess resulted in therapeutic failure. Today, with CT diagnosis to demonstrate whether the multiple loculations are interconnected or separate, a choice between percutaneous and open drainage is more easily made. If the multiple pockets do not interconnect, open drainage is preferable; otherwise, initial percutaneous drainage is preferred.

Pelvic Abscess

This form of abscess most often follows a ruptured colonic diverticulum, pelvic inflammatory disease, ruptured appendix, or drainage into the pelvis during resolution of generalized peritonitis. Unless the abscess involves the anterior abdominal wall, few symptoms or physical signs are present on examination of the abdomen. The patient may complain only of poorly localized, dull, lower abdominal pain. Irritation of the urinary system and the rectum produces symptoms of urgency and frequency or diarrhea and tenesmus.

A localized purulent collection in the pelvis is easy to diagnose. The abscess usually can be palpated directly by rectal or vaginal examination. The typical pelvic abscess bulges as a tender mass into the anterior rectal wall. In general, pelvic inflammatory masses involving the fallopian tubes do not bulge into the rectum and tend to become less tender and to resolve over serial examinations, whereas a true pelvic abscess tends to enlarge and finally to rupture.

Pelvic drainage should be accomplished, whenever possible, directly through the rectum or vagina. Drainage should be delayed until formation of the pyogenic membrane has effectively excluded the small bowel and other intraabdominal viscera. Readiness for drainage is apparent when the most prominent aspect of the abscess presenting vaginally or rectally begins to soften. Using a speculum or anoscope, the abscess should be exposed and the presence of pus confirmed by needle aspiration. The needle is left in place as a guide; sharp incision with a knife is preferred to blunt digital entry into the abscess cavity. To ensure continued drainage and obliteration of the cavity, daily dilation of the tract digitally or with an instrument should be done until the cavity becomes obliterated. Alternatively, a Penrose drain may be placed under direct vision.

Retroperitoneal Abscess

Abscesses within the retroperitoneal spaces are not common. An abscess located in the upper retroperitoneum usually is secondary to infection of the pancreas. Retroperitoneal abscesses in other locations may be caused by primary or secondary infections of the kidneys, ureters, or colon, or by osteomyelitis of the spine, or they may be secondary to trauma. They are usually insidious in onset, thereby causing a delay in diagnosis and therapy. Commonly fever is present and there is tenderness over the involved site.

CT is helpful in defining the anatomic location and the extent of the abscess. Treatment consists of intravenous antibiotic therapy and surgical drainage, preferably via an extraperitoneal approach. Percutaneous catheter drainage with CT or ultrasound guidance may be attempted and serves as a temporizing measure in patients who are poor candidates for prompt open surgical drainage. Despite aggressive therapy, overall mortality remains higher than 50 percent in the presence of multiple organ system failure.

Percutaneous Drainage Procedures for Intraabdominal Abscesses

Percutaneous abscess drainage (PAD) and open operative intervention are best viewed as complementary rather than competitive techniques. There are many situations for which PAD is the definitive procedure of choice, others for which surgery alone is indicated, and some for which both techniques are applicable, alone or in conjunction. Inflammation may present as a phlegmon (viable inflamed tissue), a liquefied abscess, infected necrotic (nonviable) tissue, or a combination. Liquefied abscesses are drainable, whereas phlegmonous or necrotic tissue is not. Decisions regarding which modality to employ are largely based on CT findings and require experience, clinical judgment, and careful consideration of underlying and coexistent disease processes. Close cooperation between the surgeon, the interventional radiologist, and other physicians involved in the patient's care is mandatory. Specific indications for PAD have expanded significantly over the past decade and now include many conditions previously thought undrainable, such as multiple or multiloculated abscesses, abscesses with enteric communication, and infected hematomas.

The basic requirements for *PAD* include a safe route of percutaneous access and the presence of a fluid collection of drainable consistency. Bleeding dyscrasias are a relative contraindication. It is generally possible to distinguish drainable fluid from phlegmonous or necrotic tissue using a combination of imaging and fine-needle aspiration. Not all fluid collections require drainage, although it is generally required for infected collections and for sterile collections that cause symptoms because of their mass. The determination of whether drainage is required is made on a case-by-case basis. A *cure* is achieved when the abscess is resolved by the drainage procedure. *Temporization* refers to resolution of an abscess and clinical improvement, with operative intervention needed to treat the underlying cause. *Palliation* refers to improvement in the patient's condition as a result of abscess drainage.

The possibility of underlying neoplastic disease must be considered in enteric perforation, especially in the elderly. Significant soft-tissue thickening of the bowel wall, especially if localized and noncircumferential, should raise the possibility of underlying tumor, as should the demonstration of potential metastatic disease such as adenopathy or liver lesions. A "target" appearance, with circumferential low-attenuation submucosal thickening sandwiched between the enhancing mucosa and submucosa, is likely to be specific for inflammatory disease. To fully exclude the possibility of neoplasia, follow-up imaging is used to verify resolution of the abscess, or confirmatory tests such as barium studies or endoscopy may be performed.

Technical Aspects

Imaging permits precise localization and characterization of disease, appropriate access route planning, and, after drainage, immediate assessment of whether the procedure was successful. Imaging also is needed for adequate follow-up to identify any problems and to gauge outcome. The drainage route should not cross a sterile fluid collection or other infected space, because of the risk of cross-contamination. Crossing the pleural space for thoracic and upper abdominal drainage carries the risk of empyema formation. Collections in the upper abdomen thus often require an angled subcostal or low intercostal approach. It is acceptable to cross the peritoneal space in order to drain an extraperitoneal abscess. Placement of a catheter through the small bowel or colon should always be avoided. Transgastric drainage of lesser sac pseudocysts has been advocated by some authors and appears to be safe, although this approach remains controversial. Lesser sac collections also can be approached transhepatically through the left lobe of the liver, although traversing solid organs should be avoided whenever possible. It is important to be aware of and to avoid major vascular structures.

In most cases drainage is performed after fine-needle (18–22 gauge) aspiration, with the aspirate being used to document infection and to gauge the viscosity of the fluid. Immediate Gram stain of the fluid is useful to determine the need for drainage. In some situations single-step aspiration of the fluid may suffice, without the need for tube placement. Examples include clearly aseptic collections, small abscesses (2 to 3 cm) into which tube placement would be difficult, and relatively nonviscous collections that can be completely evacuated. For most collections, a drain should be placed to ensure complete evacuation and to minimize the chances of recurrence. The aspiration needle can be used for placement of a guide wire or as a guide for tandem insertion of the drain. Antimicrobial therapy should be initiated before the drainage procedure to minimize any infectious complications of contaminating sterile tissue and the systemic consequences of drainage.

A wide variety of techniques and equipment are available for catheter placement, the majority falling into two categories. The *trocarr technique* involves the one-step insertion of a catheter that is loaded coaxially onto a stiffening cannula and needle. This is the technically simpler approach and is applicable to most superficial and large collections. The *Seldinger technique* is preferable for deep abscesses with difficult access, for drainages whose access route crosses exceptionally firm tissue, and for placement of catheters larger than 14F. A needle is first inserted into the collection. A guide wire is then passed through the needle, and the needle is withdrawn. The tract is sequentially dilated over the guide wire until it is possible to insert the catheter.

A variety of catheters are available for percutaneous insertion. The choice of catheter size is determined primarily by the viscosity of the fluid to be drained. In the majority of cases, 8F to 12F drains are sufficient. Larger drains may be needed for collections containing debris or higher-viscosity fluid, although the likelihood of success decreases. If needed, larger drains can be placed by exchange over a guide wire. There is no absolute limit on the number of drains that can be placed. While most abscesses can be drained with a single catheter, there should be no hesitation in placing as many drains as are needed to effectively evacuate the abscess(es).

After catheter placement, the cavity should be evacuated as completely as possible and irrigated with saline or antibiotic solution until the fluid is clear. Initial manipulation of the catheter(s) and irrigation should be done as gently as possible to minimize the induction of bacteremia. Immediate imaging determines the need for repositioning of the catheter, for placing a larger-bore catheter, or for placing additional drains. For cavities that are completely evacuated at the initial drainage and for which there are no abnormal communications to viscera, simple gravity drainage generally suffices. For larger or more viscous collections and for those with ongoing output from fistulous connections, suction drainage with sump catheters is more effective, although success is less likely. Thoracic drains always should be placed with water-seal suction.

Proper catheter management after the initial placement is a critical determinant of success and requires the interventional radiologist to remain an active member of the management team. Drains should be checked regularly (at least daily) to monitor the volume and nature of the output, to ensure adequate function and clinical response, and to recognize early and quickly correct any catheter-related problems. Periodic irrigation of the drains, once or several times per day, with sterile saline solution is recommended. In general, irrigation with proteolytic agents (e.g., acetylcysteine) or antibiotics is of no value, although fibrinolytic agents may be useful for evacuation of fibrinous or hemorrhagic collections. There is no standard protocol for follow-up imaging. Serial imaging studies and catheter injections are frequently used to monitor progress and identify problems. Occasionally it is necessary to replace or reposition tubes or to add additional catheters. The need for follow-up imaging studies should be determined on a case-by-case basis by monitoring clinical progress and drainage output.

Catheters should be removed when criteria for abscess resolution are met. The clinical criteria of success include the resolution of symptoms and of indicators of infection (fever, leukocytosis). The catheter-related criteria include a decrease in daily drainage to less than 10 mL and a change in the character of the drainage from purulent to serous. Radiographic criteria include verification of abscess resolution and of closure of any fistulous communications. If catheters are maintained until these criteria are satisfied, the likelihood of recurrence of the abscess will be minimized. Although some authorities recommend gradual catheter removal over several days, we usually remove the drain in one step and have had no significant problems with recurrence. For sterile fluid collections, the drain should be removed as soon as possible, generally within 24 to 48 h, to minimize the risk of superinfection.

Causes of Failure

A number of factors are consistently identified as causes of *PAD* failure. Prominent among these are fluid that is too viscous for drainage, or the presence of phlegmonous or necrotic debris. Technical modifications such as increasing the drain size and using irrigation can salvage some of these procedures. When phlegmonous or necrotic tissue is recognized on follow-up imaging studies, attempts at PAD may be ended in favor of open drainage, or the goals and expected outcome for PAD may be revised. Multiloculated collections and multiple abscesses are another cause of failure that can be minimized by using an adequate number of catheters along with mechanical disruption of adhesions with a guide wire. Fistulous communications, either unrecognized or persistent, are yet another potential cause of failure, as is drainage of a necrotic tumor mistaken for an abscess in imaging studies. Recognition of a significant soft-tissue component, maintenance of a high index of suspicion, and the use of percutaneous biopsies can minimize the risk of failing to appreciate the presence of tumor. Suspicious fluid also may be sent for cytologic examination. The success rate for PAD tends to be lower in immunocompromised patients. Lambiase and associates reported a cure rate of 53 percent in immunocompromised patients (including those with alcoholism, AIDS, diabetes, renal failure, or steroid use) as compared to 73 percent in immunocompetent patients.

MANAGEMENT OF SPECIFIC INTRAABDOMINAL INFECTIONS

In cases with abscesses complicating diverticulitis, *PAD* usually permits stabilization and allows time to prepare the patient optimally for operative therapy. Subsequent operation is required in most patients and is generally simplified to a one-step procedure. In some patients who remain asymptomatic after drainage, such as those with other, ultimately fatal diseases, subsequent colectomy may be avoided. It is important to perform follow-up radiographic studies to exclude the possibility of a perforated neoplasm. Percutaneous drainage for abscesses complicating Crohn's disease is less successful. Patients without fistulous communications to the bowel are usually cured by *PAD*, whereas those with fistulas generally require bowel resection. Among patients requiring operation, initial PAD usually leads to significant clinical improvement.

Low pelvic abscesses in contact with the rectum or vagina may be treated surgically by incision and drainage through these organs. The same approach can be taken using ultrasound guidance, and recent advances in endoluminal ultrasound techniques have facilitated such procedures. Ultrasound-guided transrectal and transvaginal drainage are effective and well tolerated. Good results also have been achieved in the management of tubo-ovarian abscesses complicating pelvic inflammatory disease that are refractory to medical management. In many such cases the need for hysterectomy and oophorectomy has been obviated.

The distributions of bacterial isolates in clinical intraabdominal infections as reported in several studies are listed in Table [32-7](#).

Infections Complicating Acute Pancreatitis

Acute necrotizing pancreatitis is the antecedent cause of pancreatic abscess in the majority of cases. Secondary infection of a pancreatic pseudocyst and abdominal trauma with pancreatic injury are other important causes. Abdominal pain, nausea, vomiting, distention, and absent bowel sounds are frequently present. Abdominal tenderness and fever with a temperature higher than 39°C also are commonly present.

Pancreatic abscesses are polymicrobial in nature, the common organisms being aerobic representatives of the fecal flora, predominantly *E. coli* and aerobic hemolytic streptococci. Staphylococci usually are not involved in a primary pancreatic abscess but are frequently recovered in abscesses that follow abdominal exploration in the presence of pancreatitis.

Percutaneous drainage has a high failure rate with pancreatic abscesses, and therefore open surgical drainage is preferred. The approach is transperitoneal above the transverse mesocolon and involves radical debridement of all necrotic tissues, followed by irrigation with saline and placement of large-bore sump-suction drains. Reexploration with further debridement is frequently needed, and the staged abdominal approach has been helpful.

Infections superimposed on acute pancreatitis are among the most difficult intraabdominal infections to manage, and no consistently successful approach exists for lesser sac collections with pancreatic necrosis. Current surgical therapy consists of repetitive scheduled debridement at 48-h intervals.

For pancreatic debridement, a bilateral subcostal (chevron) or transverse incision is used to stay superior to the transverse mesocolon and avoid exposure of the small bowel. The lesser sac is entered and the stomach and colon retracted cephalad and caudad, respectively. Blunt dissection is used to remove necrotic pancreas and to open loculated abscesses. The pancreatic bed is irrigated and often packed with gauze. The value of using packing material has not been established by clinical studies and probably has little merit. Foreign material incites an inflammatory response and prolongs the time needed to clear the infection. Failure to remove packing within 48 h results in recurrent fever and other findings of infection. The use of topical antibiotic solutions is not recommended since parenteral antibiotics penetrate sites of inflammation well and topical solutions may produce unwanted adverse systemic effects.

Three or four reexplorations suffice to remove necrotic tissue and produce a granulating wound. After the final reexploration prosthetic material used to facilitate abdominal reentry should be removed and the fascia closed. Patients should be monitored closely for evidence of recurrent infection and undergo CT scanning if infection is suspected.

For localized (acute or chronic) fluid collections, percutaneous drainage is successful in most cases. Fistulous communications to the pancreatic duct are commonly present but may be difficult to detect radiographically. To minimize the risk of recurrence with pancreatic fluid collections, it is especially important to verify complete cessation of drainage (< 25 mL/day) before the drains are removed. *Endoscopic retrograde cholangiopancreatography (ERCP)* is valuable in determining patency of the pancreatic duct, since fistulas associated with obstruction are unlikely to heal and generally require operation.

CT is the imaging tool of choice for localizing and characterizing complications of acute pancreatitis, and fine-needle aspiration is invaluable in determining and evaluating infection. Percutaneous drainage is a therapeutic option for evacuation of infected fluid but is not capable of removing infected necrotic tissue. Percutaneous drainage can be used to temporize and allow for a delayed definitive necrosectomy. Factors that would militate against this approach include the presence of multiple small lesser sac abscesses or concern about erosion of the inflammatory mass into the colon or major blood vessels. Two recent series detailing the use of *PAD* for severe complicated pancreatitis reported success rates of only 37 and 47 percent. Although the fluid collections associated with pancreatitis usually can be drained percutaneously, operation is required for debridement of infected necrotic tissue. Drainage of central collections (in the pancreatic bed and lesser sac) is less often successful than is drainage of peripheral collections because of the frequent presence of phlegmon and necrosis in the central regions.

A combination of percutaneous drainage and surgical debridement is used for infections complicating pancreatitis. One possible approach that has already been mentioned is to perform *PAD* initially, followed by operation for necrosectomy. This carries the risk of unduly delaying definitive surgical therapy in critically ill patients. A preferable approach is to perform early surgical debridement of the central necrotic tissues and to use *PAD*, if needed, for peripheral or residual fluid collections. Antimicrobial chemotherapy for patients with infections complicating pancreatitis should be guided by cultures of specimens from the pancreatic bed. Often patients are sequentially infected with gram-negative enteric flora, then gram-positive methicillin-resistant organisms, and finally *Candida* species.

Intraabdominal Infections in Postoperative Patients

Postoperative peritonitis generally is a consequence of anastomotic dehiscence. This is a highly lethal condition, in part because it often is diagnosed late because of reluctance to entertain the possibility of a suture-line leak. This diagnosis should be considered in any patient with signs of sepsis who has undergone a gastrointestinal anastomosis. Typical findings of diffuse abdominal tenderness may be masked by incisional pain. Because laparotomy itself introduces free air into the abdominal cavity, pneumoperitoneum is a nonspecific finding in patients during the first few days after celiotomy. Ultrasonography or CT will reveal peritoneal fluid, which, if present, should lead to ultrasound-guided aspiration for diagnostic purposes. A stain that reveals white blood cells or bacteria is an indication for immediate laparotomy. Surgical treatment should include resection and reanastomosis in the absence of purulence (small bowel) or end colostomy (colon). Postoperative abscesses are managed as detailed above.

Fistulas

Intestinal fistulas present some of the more difficult diagnostic and therapeutic problems following intraabdominal operation. The most common source is the small intestine, followed by colon, stomach, duodenum, biliary tract, and pancreas. The initial finding in most patients is occult sepsis, and the systemic response is caused by the inflammation surrounding the nascent fistula. Radiographic evaluation commonly suggests an abscess.

Abscesses with fistulous communication to the alimentary canal, biliary tree, or pancreatic duct represent a special problem for percutaneous drainage. Fistulas are loosely characterized as high-output (more than 500 mL/day) or low-output (less than 200 mL/day). Low-output fistulas may be managed with *PAD*, but most fistulas require sump drainage.

In several series an average of only 26 percent (range 17 to 40 percent) of fistulas were identified at the time of initial drainage. Presumably the abnormal communication is initially occluded with debris, or adequate maneuvers to uncover the leak are not performed, in the interest of minimizing tissue manipulation. A sudden change in the character of drainage or persistent output greater than 50 mL/day should alert the clinician to the presence of a fistula. Injection of contrast medium into the drainage catheter provides accurate imaging; other contrast studies (upper gastrointestinal contrast study, barium enema, *ERCP*, or radionuclide biliary scan) are less likely to be useful.

A catheter should be placed as close as possible to the site of leakage, using additional catheters as needed for abscess drainage. Suction drainage should be used to gain control of the leak. Slowly withdrawing the drain may promote tract closure after drainage stops. Some authors recommend a trial of capping the drain before removal, followed by imaging, to exclude fluid reaccumulation.

In five reported series the success rate of *PAD* with fistulas ranged from 67 to 85 percent. For a significant number of patients who were not cured, the procedure served to temporize, simplifying subsequent surgery to cure the underlying disease. When the involved bowel or duct is otherwise normal, as commonly occurs with postoperative fistulas, the vast majority of drainage procedures are successful. The presence of active underlying inflammatory disease (e.g., Crohn's disease, diverticulitis), ischemia, or neoplasia is associated with a higher rate of failure, and temporization in these cases is often a more reasonable goal. Downstream obstruction must be excluded, as this invariably prevents closure of the fistula. Maintenance of nutrition (intestinal tube feeding or parenteral nutrition) is a critical determinant of success. Some low-volume distal colonic fistulas can be managed by elemental diets.

If possible, enteric nutrition should be provided through catheters placed distal to the fistula, but it can also be given proximal to the fistula if there are 4 feet of bowel between the feeding tube and the fistula. For high- output fistulas, parenteral nutrition is often required. Somatostatin appears useful in the management of patients with pancreatic and biliary fistulas, but not enteric fistulas.

ANTIMICROBIAL THERAPY FOR INTRAABDOMINAL INFECTIONS

The goals of antibiotic therapy for intraabdominal infections that will be treated by either percutaneous or operative intervention are to hasten the elimination of infecting microorganisms and thereby shorten the clinical manifestations of infection and minimize the risk of recurrent intraabdominal infection. Since the surgical wound is heavily contaminated by the infecting microorganisms, effective antimicrobial therapy should be begun prior to operation. Necrotizing fasciitis and other forms of extension of infection to the surgical wound represent catastrophic failures of antimicrobial treatment.

In patients with localized abscesses, antibiotics reduce fever and other manifestations of systemic response, but only over an interval of 24 to 36 h. Antibiotics should be administered after fluid resuscitation has been initiated to restore adequate visceral perfusion and provide better drug distribution. Particularly in the case of aminoglycosides, nephrotoxicity is exacerbated by impaired renal perfusion.

Antimicrobial agents are begun when the diagnosis of intraabdominal infection is suspected. This is often before an exact diagnosis is established and before results of appropriate cultures are available. Accordingly, the clinician must anticipate the pathogens most likely to be encountered at the site of infection. Antibiotics used for intraabdominal infections should be active against enteric gram-negative facultative and obligate anaerobic bacilli. The identity and density of microorganisms depend on the site of the gastrointestinal tract perforation. In general, gastric, duodenal, and proximal jejunal perforations release small numbers of gram-positive aerobic and gram-negative anaerobic organisms into the peritoneal cavity. These organisms are generally susceptible to b-lactam antibiotics and are rapidly eradicated by defense mechanisms in intact hosts. *Candida albicans* or other fungi are cultured from about 20 percent of patients with acute perforations of the gastrointestinal tract. Even when fungi are recovered, antifungal agents are unnecessary unless the patient has recently received immunosuppressive therapy for neoplasm, transplantation, or inflammatory disease, or has recurrent intraabdominal infection.

Microbiologic Specimen Collection

Immediately after the abdominal cavity has been opened, a syringe is used to aspirate pus and fluid free of air. Pus is the best transport medium to maintain the viability of bacteria. The syringe containing pus is taken immediately to the laboratory for processing. If the sample cannot be transferred to media expeditiously, 1 to 2 mL of pus is injected into a sealed anaerobic tube (Port-a-cult) and kept at room temperature. Most facultative aerobes survive transportation and holding in an oxygen-free environment together with anaerobes.

The techniques used in selecting and obtaining microbiologic specimens for culture have a marked effect on the organisms recovered. The specimen collected should be representative of the material associated with the clinical infection. There must be sufficient material to permit complete microbiologic analysis. In many cases this is not possible when a specimen is submitted on a swab. When plastic syringes are used they should be transferred to the laboratory for appropriate inoculation into anaerobic medium within 30 to 60 min or oxygen will diffuse through the plastic. In some cases it may be more appropriate to send a sterile tissue sample. Specimen collection on a cotton-tipped swab is the least desirable method because only a small amount of material can be collected and because many organisms will not grow from swabs. If a cotton-tipped swab is used it should be immediately inoculated into a transport medium.

Bacterial Isolates

The organisms recovered from peritonitis and intraabdominal abscess are very similar and include a mixture of aerobic and anaerobic species. Polymicrobial clinical isolates are obtained in more than two-thirds of clinical series. The number of bacteria isolated depends on the meticulousness of laboratory technique as well as on the quality of the clinical specimen. In most clinical settings, two or three aerobic species are identified, and one or two anaerobic species. In the setting of a research-oriented microbiology laboratory, seven to ten aerobic organisms are identified along with 10 to 15 anaerobic species. The clinical utility of such extensive microbiologic efforts is dubious, but they help us to understand the pathogenesis of these infections. One of the most striking findings is that the bacteria populations from peritonitis or abscess are qualitatively and quantitatively different from those found within the gut lumen.

Of the wide spectrum of microorganisms in the intestines, only a few are capable of causing infection in human beings, and even fewer participate in polymicrobial intraabdominal infections. Comparison of the relative occurrence of bacteria in the normal intestine and their relative recovery in intraabdominal infections reveals distinctive selection patterns that reflect the survival capacity of certain bacteria outside the bowel lumen.

Obligate anaerobic bacteria comprise more than 99 percent of the intestinal flora. After bowel perforation most of these anaerobes can be isolated by appropriate culture techniques. They die off quickly, however, in conditions outside their natural environment. Only as contamination of the abdominal cavity progresses to established infection does selection of the most pathogenic organisms take place, resulting in the typical spectrum of bacteria found in an intraabdominal infection.

Cultures from patients with distal small-bowel perforations grow gram-negative facultative organisms with variable density. Perforations of the distal small bowel often evolve to localized abscess formation and present with peritonitis only after rupture of the abscess. Colonic anaerobes such as *B. fragilis* are sometimes present. In patients with colon-derived intraabdominal infections the peritoneal cavity is contaminated with large numbers of facultative and obligate anaerobic gram-negative organisms.

Subsequent decisions regarding antimicrobial therapy should be guided by the results of intraoperative cultures. Although the appropriate role of antienterococcal therapy is controversial, most authorities believe that specific antienterococcal therapy should be given only when enterococci are the only organisms isolated or when they are isolated from blood. Isolation of enterococci as part of a mixed gram-positive and gram-negative flora should not prompt the addition of ampicillin or vancomycin to the antibiotic regimen. The incidence of treatment failure for patients harboring enterococci and not treated for it is the same as for patients treated with imipenem or other agents effective against enterococci. Enterococci are very low-level pathogens, meaning that they incite little host response and do not cause invasive infection in intact hosts. Patients who have had one major episode of sepsis are sufficiently immunosuppressed that isolation of enterococci from a second infectious site (including recurrent infection within the abdomen) should mandate specific antienterococcal therapy. If the smear reveals gram-negative bacilli, failure to isolate either facultative or obligate anaerobes on culture does not obviate the need to continue providing antimicrobial agents against both. Antimicrobial susceptibility patterns within each hospital should be heeded in selecting initial empiric therapy.

Empiric Therapy

Antimicrobials are often begun if infection is even a remote possibility. The wisdom of this approach depends on the evidence suggesting an infection, taken against the background risk of harm to the patient. Antimicrobial treatment is not without potential adverse effects, but these generally are perceived as temporally remote and often take the form of changes in hospital- or unit-specific susceptibility patterns. For patients with suspected community-acquired infection, recent onset of malaise, symptoms referable to a specific anatomic area, and findings of fever and localizing tenderness, mass, or evidence of organ dysfunction are often sufficient to begin antimicrobial therapy after appropriate cultures are obtained. Except in certain forms of meningitis, initiation of antibiotic therapy is not time-critical. Antibiotics should be administered after fluid resuscitation has begun, so that distribution of the drug(s) is adequate. As noted earlier, the nephrotoxicity of antibiotic agents, especially of the aminoglycosides, is exacerbated by impaired renal perfusion. In the case of suspected community-acquired infection, knowledge of the endogenous flora giving rise to infection is a valuable guide to the selection of antibiotic therapy while awaiting the results of cultures.

Infections in hospitalized patients, particularly in the intensive care unit, require a different set of criteria for beginning antimicrobial therapy. Such patients often have undergone extensive operative procedures and have numerous noninfectious potential causes for fever. Conversely, infection may be masked by immunosuppression associated with the condition (or treatment) that resulted in hospitalization. Such patients become colonized, and cultures of sputum, wounds, and drain tracts are often positive in the absence of infection. Interpretation of culture results is therefore quite difficult unless positive cultures are obtained from a routinely sterile site such as blood or pleural or peritoneal fluid.

Value of Precise Identification of Infecting Organism(s)

In vitro data, especially antimicrobial susceptibility tests, are predictive of the in vivo response of infecting bacteria to antibacterial agents; this allows selection of a specific agent or combination regimen and obviates the need for broad-spectrum therapy. Equally important information is gained from stains of infected fluid collections or other direct examinations such as KOH preparations of cerebrospinal fluid for *Cryptococcus neoformans* or bronchoalveolar lavage fluid for other fungi. While a variety of susceptibility testing techniques are available, automated testing is appropriate for bacteria isolated from most infections except in extraordinary circumstances. Certain organisms do not require routine susceptibility testing. Nonenterococcal streptococci are routinely susceptible to penicillin. Because of the technical problems of performing anaerobic susceptibilities, it is best for such isolates to be identified and then batch-tested on a yearly basis for changes in otherwise predictable susceptibilities. This particularly applies to *B. fragilis* group isolates. Local or regional hospital antimicrobial susceptibility patterns determine initial empiric therapy.

Fluid collections, particularly if purulent, should be obtained for smear culture. Specimens of infected fluid are sent to the laboratory in a capped airless syringe with no needle or collected in appropriate separate aerobic and anaerobic transport media. If the stain reveals a predominance of gram-positive cocci, which may indicate that the enterococci or other fecal streptococci are significant copathogens at the site of infection, the clinician should consider alterations in the antibiotic regime to include agents specifically active against *Enterococcus* species.

If the smear reveals gram-negative bacilli, failure to isolate either facultative or obligate anaerobes on culture does not obviate the need to continue to provide antimicrobial agents against both. Conversely, stain in the presence of purulence will not demonstrate a bacterium if organisms are present at a density less than 10^5 /mL. A negative stain should not prompt discontinuance of antibiotics until culture results are available and a clinical response has been achieved.

The disk-diffusion method provides qualitative data about the susceptibility of an organism to a given agent. Quantitative data require serial dilutions of the antimicrobial agent in media. The lowest concentration of the agent that prevents visible growth after an incubation period of 18 to 24 h is known as the *minimum inhibitory concentration (MIC)*. The *minimal bactericidal concentration (MBC)* can be determined by subculturing on antibiotic-free medium those tubes that show no growth in the determination of the MIC.

There are a number of common errors in the interpretation of sensitivity results from hospital laboratories. Certain agents such as nalidixic acid, nitrofurantoin, and norfloxacin are tested at concentrations achievable only in urine. Susceptibility results for such agents are not meaningful for the management of systemic infections, and using them may result in treatment failure. Conversely, some nontoxic antibiotics for urinary tract infections, such as ampicillin, are not tested at the high concentrations achieved in urine and may be effective even though "resistance" is reported. Intravascular infections such as endocarditis or catheter-related bacteremias may be inadequately treated by *bacteriostatic* agents despite in vitro susceptibility results suggesting efficacy. Concentrations of bacteria employed in susceptibility testing are considerably below those found in clinical situations. Antibiotics effective at the lower bacterial concentrations may not be effective at the higher inoculum. This dependence of antibiotic efficacy on bacterial concentration, termed the *inoculum effect*, is prominent for the b-lactam antibacterials.

Treatment Response

Treatment response is defined as a diminution in the physical signs of infection, including decrease in fever, in tachycardia, and in local findings of tenderness and organ dysfunction, such as ileus. The more localized the initial infection, the more rapidly response will occur. Conversely, extensive infections such as peritonitis or multilobar pneumonia may respond only slowly. As a general rule, infections should show definite evidence of response within 72 h of initiation of treatment. The absence of response is not an automatic indication that the wrong antibiotic was selected. In most cases of apparent treatment failure the same antibiotic should be continued while a diligent search is made for a localized septic anatomic focus requiring drainage, debridement, or excision.

Selection of Antibacterial Agents

The combination of evidence from in vitro data, animal studies, and clinical trials has led to widespread acceptance of the need to provide empiric antimicrobial therapy directed against *E. coli* and other common members of the family Enterobacteriaceae and against *B. fragilis*. *B. fragilis* and *E. coli* are the most common isolates from intraabdominal infections and are most likely to cause bacteremia in abdominal sepsis, further attesting to their pathogenicity. The evidence in support of broadening therapy to cover organisms other than common facultative and obligate anaerobes such as *E. coli* and *B. fragilis* is more controversial. Initial empiric coverage of *P. aeruginosa* if these organisms are isolated from the site of infection is associated with a decreased likelihood of persistent or recurrent abdominal infection. Other clinical trials, however, do not demonstrate a high incidence of treatment failure associated with this organism, even if it is not treated.

There are a number of agents that are broadly active against the bacteria found in intraabdominal infections. These are best discussed as classes of drugs including the aminoglycosides, carbapenems, cephalosporins, penicillins plus b-lactamase inhibitors, and quinolones. Aztreonam will be considered as a cephalosporin-class agent. An example of what some consider appropriate antibiotics in peritonitis is provided in Table [32-8](#). Current sensitivities are listed in Table [32-9](#).

Aminoglycosides

The aminoglycosides have been the mainstay of therapy for serious gram-negative infections for the past thirty years. Because of their potential for nephrotoxicity and ototoxicity, aminoglycosides are less likely to be used as first-choice agents for community-acquired intraabdominal infections. The use of b-lactam antibiotics with clindamycin or metronidazole, b-lactam antibiotics combined with b-lactamase inhibitors, or single-agent imipenem/cilastatin in mixed-flora infections has resulted in clinical results equivalent to or better than those seen with aminoglycoside-based combinations. The aminoglycosides no longer represent a "gold standard" of comparison and need not be used for community-acquired intraabdominal infections.

Hypotensive patients with gram-negative bacteremias have higher survival rates if treated at least initially with aminoglycoside-based combination therapy. Approximately one-third of patients with nonappendiceal intraabdominal infections are bacteremic. While gram-negative organisms do not represent the same risk of endocarditis on normal valves or of metastatic abscess formation seen with *Staph. aureus*, combination bactericidal therapy may result in more rapid clearance of organisms and abbreviate host deterioration. Aminoglycosides should specifically be used in the initial treatment of patients with major intraabdominal infection and hypotension in combination with a b-lactam agent, which is effective against the anticipated gram-negative organisms.

Pharmacokinetics

In contrast to the b-lactam antibiotics, the efficacy of the aminoglycosides depends on the maximal concentration achieved. As the ratio of the concentration of aminoglycoside to the minimum inhibitory concentration (ratio of peak concentration to MIC) is increased, the rate and extent of killing is also increased. Moore and colleagues examined the relationship between plasma aminoglycoside concentrations, MICs for the infecting organisms, and therapeutic outcome. Higher and mean ratios of peak aminoglycoside concentration to MIC (6:1) were correlated with a positive clinical response. In addition to concentration-dependent bactericidal activity, aminoglycosides, like other antimicrobials that affect the synthesis of protein, DNA, or RNA, exhibit persistent suppression of bacterial growth after antimicrobial exposure, which is referred to as a post-antibiotic effect (PAE). This phenomenon allows killing to continue despite drug concentrations well below the MIC of the infecting organism. This translates into prolonged dosing intervals. Cell-wall-active agents such as penicillins and cephalosporins exert a PAE against gram-positive bacteria, but none against gram-negative bacteria. Imipenem exerts a modest PAE against gram-negative bacteria. In general, the greater the concentration-to-MIC ratio, the longer the PAE. The PAE allows aminoglycosides to be administered less frequently, and this in turn lowers the risk of aminoglycoside toxicity.

Nephrotoxicity

Aminoglycoside nephrotoxicity is believed to be caused by its accumulation in the proximal renal tubular cell by the process of adsorptive pinocytosis. Once in the cell, the antimicrobial is taken up by cellular lysosomes. This process is diminished when aminoglycoside tubular fluid concentrations fall below a critical value and becomes saturated at higher concentrations. Neither of these limiting values is known precisely, but 2 and 12 mg/dL, respectively, appear to be of the right order of magnitude. Once incorporated into lysosomes, aminoglycosides inhibit lysosomal phospholipase and sphingomyelinase activity, and if drug exposure is prolonged, proximal renal tubular cell phospholipidosis occurs. The overloaded lysosomes swell and then rupture, releasing aminoglycoside, lysosomal enzymes, and phospholipids into the cytosol. The functional integrity of the cell is disturbed, and the result is cell death.

This proposed mechanism may explain why allowing serum aminoglycoside concentrations to decline below 2 mg/dL ("trough level") before administering the next scheduled dose is effective, and why peak serum concentrations have such a minor influence on the nephrotoxic potential of aminoglycosides. In individuals with normal renal function, once-daily dosing of gentamicin and netilmicin routinely results in serum trough concentrations of less than 0.5 mg/dL. Multiple daily doses of aminoglycosides result in more frequent or more severe nephrotoxicity than do less frequent or once-daily doses of these agents.

Pharmacokinetics in Critically Ill Patients

Traditional aminoglycoside dosing methods such as manufacturer's recommendations or nomograms are based on the assumption that pharmacokinetic parameters such as volume of distribution (Vd) and clearance (Cl) are similar from patient to patient, regardless of severity of illness. In fact, however, there is not only a wide variability in these parameters among non-critically ill patients but also significant variability between non-critically ill and critically ill patients. Traditional dosing methods thus can result in inappropriate aminoglycoside regimens.

Septic and postsurgical patients in intensive care units often require large volumes of fluid for resuscitation with major fluid shifts. Alterations in the concentrations of hormones and vasoactive substances may affect blood flow and organ perfusion and function. Cardiac, respiratory, renal, and hepatic failure occur frequently, alone and in combination, and can significantly alter the absorption, distribution, metabolism, and elimination of many drugs. Critically ill patients with sepsis demonstrate altered aminoglycoside pharmacokinetic values as compared with those not critically ill. Critically ill patients typically are significantly older and have higher serum creatinine concentrations and lower elimination rate constants and total body clearances than non-critically ill patients. The major contributing factor, especially in surgical patients, appears to be volume replacement/resuscitation following operation. These differences require significant alterations in the dosing regimens of the aminoglycosides. In critically ill patients, a decrease in clearance and elimination rate constants means that the conventional 8-h dosing intervals may not be sufficient to maintain trough concentrations of less than 2 mg/dL. Additionally, the increase in distribution volume may require a larger dose of aminoglycoside to achieve therapeutic peak concentrations.

Failure to appreciate a decrease in clearance and an increase in distribution volume in critically ill patients can result in suboptimal aminoglycoside therapy. For critically ill patients, gentamicin and tobramycin doses of 3 to 5 mg/kg/day should be replaced with a higher dose of 7 mg/kg/day administered in two divided doses (i.e., every 12 h) or as a single daily dose.

In addition to the difference in pharmacokinetic parameters between critically ill and non-critically ill patients, there also appears to be a wide variation in aminoglycoside pharmacokinetics among critically ill patients. In addition, there is little correlation between changes in serum creatinine concentration and dosing interval. Serum creatinine levels should not be used as a measure of aminoglycoside toxicity or in order to estimate an appropriate dosing interval; rather, it is necessary in critically ill patients to obtain frequent aminoglycoside serum concentrations.

High-dose intermittent therapy with aminoglycosides is preferable. Patients with major infections have expanded volumes of distribution for aminoglycosides and commonly require 2.5 mg/kg gentamicin or tobramycin to achieve therapeutic levels. Regimens involving high-dose (10 mg/kg) gentamicin or tobramycin given every 24 h have been used. This is based on two key antimicrobial properties of aminoglycosides—dose- dependent killing and a post-antibiotic effect. Several trials have examined once-daily dosing, employing netilmicin, amikacin, tobramycin, and gentamicin. Infections treated have included cystic fibrosis, gangrenous and perforated appendicitis, other intraabdominal infections, gram-negative bacteremias, mixed serious infections, pelvic inflammatory disease, and urinary tract infections. Doses have ranged from 4.5 to 9 mg/kg/day for tobramycin, netilmicin, and gentamicin, and 15 mg/kg/day for amikacin. Peak serum concentrations with the once-daily regimen have been obtained with doses ranging from 8 to 52 mg/dL for tobramycin and netilmicin, 33 to 55 mg/dL for amikacin, and 35 mg/dL for gentamicin.

The patient's immunocompetence may affect the utility of this regimen. In an animal study, once-daily tobramycin was found to reduce the mean bacterial titer in the lungs significantly more than the every-4-h regimen in nonneutropenic animals. In neutropenic animals, however, the once-daily regimen proved to be less effective than the every-4-h regimen, as evidenced by a rapid regrowth of bacteria during the latter 12 to 16 h of the 24-h dosing interval.

b-lactam Antibiotics and b-lactam/b-lactamase Inhibitor Combinations

First-generation cephalosporins, including cefazolin, cephapirin, and cephalothin, have excellent gram-positive activity, moderate gram-negative activity, and no anaerobic activity. Cefonicid, cefamandole, and cefuroxime may be grouped with these agents because none has anaerobic activity. The second-generation cephalosporins cefoxitin, cefotetan, and cefmetazole all have some anaerobic activity, improved facultative gram- negative activity, and less gram-positive coverage. The anaerobic activity of these agents against *B. fragilis* is unimpressive, with about one-third to one-half of tested isolates being resistant. Because of the high incidence of *B. fragilis* and relatively large inoculum loads encountered in colon-derived infections, these agents are best used for prophylaxis and for treatment of low-inoculum infections such as appendicitis.

The third-generation cephalosporins, cefotaxime, ceftizoxime, cefoperazone, ceftriaxone, ceftazidime, and cefepime, have considerable facultative gram-negative activity, but no anaerobic and limited gram- positive coverage. Aztreonam, a monobactam, has activity against facultative gram-negative organisms equivalent to third-generation cephalosporins. It has no gram-positive or anaerobic activity. Metronidazole has remained highly effective against *Bacteroides* species, unlike clindamycin, and is now the preferred agent in combination therapy.

The choice of one third-generation cephalosporin or aztreonam versus another is not a major issue. As experience with these agents has widened, it has become apparent that the differences between agents do not affect outcome. Many hospitals have therefore taken the position that cephalosporins can be grouped into classes within each of which the agents are therapeutically interchangeable. Acquisition costs commonly determine which cephalosporin is used in each class. Ceftazidime is *not* recommended because broad usage is associated with decreasing susceptibility of nosocomial *P. aeruginosa*, and this agent is the only cephalosporin usable for this pathogen. Additionally, ceftazidime therapy is associated with an increased incidence of enterococcal superinfections.

Another, more speculative issue fueling interest in antibiotic regimens other than cephalosporins or penicillin derivatives is the possibility that such b- lactam therapy may exacerbate the degree of sepsis seen in the patient. There is clear evidence that agents such as cephalosporins and penicillin derivatives disrupt the outer membrane of gram-negative organisms, exposing the lipid structures ordinarily facing the interior of the cell and ordinarily hidden by long polysaccharides attached to the lipid core. Antibiotic treatment results in the formation of small bacterial fragments with considerable proinflammatory capabilities. Cultures of gram-negative bacteria treated with b-lactam antibiotics release endotoxin, and culture filtrates induce considerably more *TNF- a* from mononuclear cells than do filtrates from cultures treated with aminoglycosides, carbapenems, or quinolones.

b-lactamases in Bacterial Resistance

b-lactamases are ubiquitous, and therapeutic use of b-lactam antibiotics has been selected against bacterial strains with increased amounts of enzymes and/or selected enzymes with an extended spectrum of activity. The most important chromosomal enzymes are the Class I enzymes that occur in Enterobacteriaceae and *Pseudomonas* species; these enzymes have a high affinity for b-lactams but hydrolyze the drugs slowly. Some bacteria are able to synthesize large amounts of b-lactamases if challenged with a b-lactam. This induced expression is important for *P. aeruginosa*, *Enterobacter*, *Citrobacter*, *Serratia*, *Morganella*, and *Providencia* species, for *Proteus vulgaris*, and for bacteria in which enzyme activity is stably derepressed. Once selected, derepressed mutants remain sensitive to carbapenems, cefepime, ceftiprome, quinolones and aminoglycosides.

TEM-1 is now the commonest plasmid-mediated b-lactamase. This is followed in frequency by SHV-1, PSE-1 and -4, which confer resistance to ampicillin, amoxicillin, ticarcillin, piperacillin, mezlocillin, cefamandole, and cefoperazone. Mutants have been identified in clinical isolates; these mutants have even broader substrate profiles, including most penicillins, cephalosporins, and aztreonam.

A serine in the active site forms a noncovalent complex with the b-lactam, followed by acylation. During this reaction the b-lactam ring is opened, destroying the drug's activity. This reaction also occurs with the drug's true target, the penicillin-binding proteins (PBPs). In this case, drug binding to the PBP inactivates it and prevents cell replication. For b-lactamases, hydrolysis is much more rapid than for PBPs, and it liberates the enzyme and results in drug inactivation. In vivo susceptibility to hydrolysis depends on the pericytoplasmic concentration of the antibiotic, which often is quite low.

Clinically meaningful problems with b-lactamase induction have most commonly involved *Enterobacter* species. *Enterobacter* species have recently emerged as a major cause of nosocomial gram-negative bacteremia in the intensive care unit, ranking third to fifth in frequency. The organism's ability to rapidly develop resistance to third-generation cephalosporins, resulting in higher mortality rates, has been cause for concern.

Chow and associates reported that the administration of a third-generation cephalosporin within 14 days of *Enterobacter* bacteremia was more likely to be associated with a multiresistant *Enterobacter* species in the initial positive blood culture. Patients infected with multiresistant organisms had a higher mortality rate that was independent of other risk factors such as age and underlying disease. This supports the contention by some that highly resistant organisms may emerge more rapidly as a result of the routine use of

cephalosporins. Therefore, when *Enterobacter* organisms are isolated from blood, third-generation cephalosporins should be avoided regardless of in vitro susceptibility.

There is variability from one intensive care unit (ICU) to another in the incidence of *Enterobacter* and other organisms that can induce b- lactamases, and there are also unit-specific organisms with mutated b- lactamases. Therefore susceptibility data of ICU organisms should be reported separately from isolates from other hospital units, and this unit- specific data should be employed to make decisions regarding empiric therapy. Many units have deliberately continued the practice of initiating empiric treatment for suspected gram-negative infection with aminoglycoside/b-lactam combinations because of a relatively high incidence of resistant organisms. The aminoglycosides as well as the carbapenems and quinolones have thus far retained their in vitro activity against these organisms, although carbapenem-resistant *Enterobacter* have been reported.

An alternative strategy to the use of b-lactamase–resistant cephalosporins is to use b-lactams in combination with b-lactamase inhibitors that alone are not potent antibiotics. Sulbactam, clavulanic acid, and tazobactam are efficient inhibitors of b-lactamases from gram-positive and anaerobic organisms. They have lesser activity against the chromosomal b- lactamases seen in Enterobacteriaceae and do not completely compensate for the marginal gram-negative activity of the penicillin derivative. The primary concern has to do with organisms that hyperexpress b- lactamases. Organisms that typically do this include *P. aeruginosa*, and *Enterobacter*, *Citrobacter*, *Serratia*, and *Acinetobacter* species. These particular organisms are most commonly encountered in nosocomial infections but are also present in about 15 percent of community-acquired infections.

Carbapenems

Imipenem, a carbapenem derivative, has broad activity against facultative and obligate gram-negative anaerobes and excellent gram-positive activity, excepting methicillin-resistant staphylococci. This agent is formulated with cilastatin, a renal dehydropeptidase inhibitor that prevents renal tubular epithelial metabolism of the drug. In situations where plasma accumulation of the drug occurs, as in high-dose levels or renal failure, the drug may cause seizures. With lower-dose levels (500 mg) and appropriate adjustments for renal failure, this is not a problem.

Meropenem is a synthetic agent with a typical broad carbapenem antibacterial spectrum, but with no evidence of an increased risk of seizures compared with cephalosporin control agents. In several clinical trials in which serious intraabdominal infections were treated with meropenem dosed at 1 g every 8 h the drug has shown efficacy.

Fluoroquinolones

The quinolones are chemical modifications of the basic nalidixic acid structure. They have antibacterial potencies 1000 times greater than that of nalidixic acid and are active against gram-negative and gram-positive organisms. The intracellular target of the quinolones is DNA gyrase, an enzyme involved in DNA breakage and repair. It appears that quinolones actually bind to DNA breaks caused by DNA gyrase, with at least four drug molecules binding per site; the binding is stabilized by the enzyme itself. The quinolones are considered bactericidal and are not susceptible to inoculum effects.

Since quinolones are totally synthesized, a vast array of congeners are possible and it is probable that the number of available oral quinolones will eventually exceed the number of cephalosporins. The currently marketed fluoroquinolones are active against facultative and aerobic gram-negative bacteria. *Streptococcus* and *Enterococcus* species show in vitro susceptibility to quinolones, but other, more effective agents are available. *Staph. aureus* and coagulase-negative staphylococci have become progressively more resistant to fluoroquinolones. These drugs should be considered only as a last resort for treatment of infections caused by multiply antibiotic-resistant gram-positive strains, and only in combination with other antistaphylococcal drugs. Fluoroquinolones have excellent activity against gram-negative bacilli. Activity patterns for ciprofloxacin and ofloxacin are very similar. For *P. aeruginosa*, ciprofloxacin has consistently shown greater in vitro activity. Treatment of *P. aeruginosa* with ciprofloxacin risks development of resistance and is not recommended. Resistance to quinolones occurs primarily through spontaneous mutation of either the A or B subunits of the target DNA gyrase. Mutation rates in *Serratia marcescens* and *P. aeruginosa* appear higher than in other gram-negative organisms, and the mutation rate for *Staph. aureus* is also quite high. Resistance also may arise through changes in bacterial permeability. Norfloxacin, because of its poor absorption and low serum concentrations, should not be used to treat systemic infections.

A primary virtue of the quinolones is a very large volume of distribution. Because of their relatively small molecular size and the absence of localized electrical charges, these agents penetrate well into interstitial fluid and achieve high tissue levels, well in excess of those seen with other, larger molecules. Recognition of the wide volume of distribution and broad susceptibility profiles has led to clinical trials examining their efficacy in serious infections. These trials demonstrate an impressive level of activity in nosocomial gram-negative pneumonia and in serious intraabdominal infections. Ciprofloxacin has been found to show better results than imipenem in the treatment of pneumonia. Patients with *P. aeruginosa* did not fare well in either arm of the trial. Almost all quinolones are excreted through the kidneys, and both ciprofloxacin and norfloxacin have an enterohepatic circulation. Hence there is only a modest increase in half-life for these agents in the presence of renal failure.

One major benefit of currently available quinolones, as compared with b- lactam agents, is their continuing activity against *Enterobacter* and other species that have inducible b-lactamases. Another advantage of quinolone therapy is the potential for oral treatment of serious gram-negative infections, as there are a variety of clinical situations in which prolonged therapy is appropriate. These include hepatic abscesses and nosocomial pneumonia. Most significant intraabdominal infections are treated for 7 or more days.

The next generation of quinolones is represented by the agents trovafloxacin and clinafloxacin. These drugs have substantially expanded spectra of activity, including against many ciprofloxacin-resistant gram- negative and gram-positive organisms and against many anaerobic organisms, including *B. fragilis*. These agents also appear to have potential side effects of gene damage and photosensitization, and their futures remain unclear.

Quinolone antibiotics, which act by inhibiting DNA replication, appear to be potentially useful for serious systemic infections and have shown similar activity to imipenem in clinical trials for pneumonia and intraabdominal infection. Available quinolones have little anti- *B. fragilis* activity and should be combined with metronidazole if an anaerobic component of the infecting flora is suspected. Quinolones are applicable because serum levels following oral absorption parallel those seen in intravenous infusion. In patients with anatomically extensive infections such as diffuse peritonitis, prolonged therapy with oral quinolones is attractive.

Dosing of cephalosporin, carbapenem, and quinolone antibiotics should be optimized on the basis of their known pharmacodynamics. Cell-wall-active agents are effective at the minimum inhibitory concentration of the drug for the organism(s) being treated. Increasing the drug concentration substantially above about two to four times the *MIC* does not increase the rate of killing. Once the drug falls below the MIC, the organism begins regrowth immediately. Dosage regimens for cell-wall-active agents in critically ill patients should involve dosing intervals sufficiently short to maintain serum levels above the MIC. With these agents, relatively small doses should be given frequently to maintain the trough level above the MIC and to avoid the costs and toxicities seen with high doses. This is best accomplished by administering these drugs every four half-lives, with adjustments as needed for renal compromise. There has been interest in infrequent drug dosing with b-lactams; this is most likely to succeed in mild to moderate infections in otherwise intact hosts.

Combination Therapy

One result of the availability of multiple classes of effective antibiotics is the possibility of combination therapy with two effective agents. Combination therapy has several theoretical advantages. A broader spectrum of pathogens can be covered, which is of particular importance when treatment is initiated for suspected sepsis without an identified source. Agents with differing mechanisms of action also may delay or prevent the emergence of resistance and superinfection. Combinations of such agents may act synergistically to enhance killing of organisms.

The time-honored tradition of providing an aminoglycoside along with a b- lactam agent has gained support from two studies of gram-negative bacteremia, demonstrating improved survival in patients with severe acute illness treated with combination therapy as opposed to those treated with a single agent. In a similarly constructed study of *Klebsiella* bacteremia, a similar improvement in severely ill patients treated with combination therapy was noted. Survival rates in both studies were around 70 percent for the combination therapy group as compared to 50 percent for the monotherapy group.

Regimens employing two b-lactam agents have been attempted primarily in the setting of febrile neutropenia. No benefit has been found, and there is concern that coverage with such regimens is not as broad as that obtained with an aminoglycoside-based regimen since the b-lactam antibiotics have similar mechanisms of action.

Pharmacokinetics of Antibiotics

Pharmacokinetic parameters can be used to aid in drug selection. Unlike with the aminoglycosides, the maximum effect of the b-lactam-like agents, including cephalosporins, penicillins, aztreonam, imipenem, and quinolones, is not correlated with the peak/ *MIC* ratio seen with a single dose. Rather, the time over a 24-h period above the MIC is most important. Studies have shown that continuous-infusion cephalosporin therapy is maximally effective. Hence the general rule with these agents is to give relatively small doses frequently—every four half-lives, with any necessary adjustments for renal function—to maintain the trough level above the MIC and thus to avoid the costs and toxicities of higher doses.

Another pharmacokinetic variable of considerable significance is the route of elimination. Most agents are excreted by the urine and achieve high levels there. Imipenem is metabolized by renal tubular epithelial cells. To circumvent this, this agent is supplied with the dehydropeptidase inhibitor cilastatin, which provides high urine levels of intact drug and prolongs the circulating half-life. Agents that are excreted primarily by the liver require no dosage adjustment in renal failure but are associated with a high incidence of diarrhea. Such agents may result in acquisition of an enteric flora that is highly resistant to antimicrobial therapy and an increased risk of antibiotic-associated colitis. Since most nosocomial pneumonias are believed to arise from microaspiration of enteric contents, this greatly limits the utility of these agents in severely ill patients with extrabiliary infections.

This property is extremely helpful, however, in managing complex biliary infections such as those seen after enterobiliary anastomoses, biliary stent placement, and biliary stricture. Cefoperazone, piperacillin, and mezlocillin are eliminated at least in part in the bile, and consequently biliary antibiotic levels are high enough to kill bacteria that are resistant to these agents at levels normally achieved in serum. These antibiotics are effective in treating aminoglycoside-unresponsive gram-negative and *Enterococcus* infections in partially obstructed bile ducts.

(Bibliography omitted in Palm version)

CHAPTER 33 - Abdominal Wall, Omentum, Mesentery, and Retroperitoneum

John M. Daly
James T. Adams
Gary A. Fantini
Josef E. Fischer

ANTERIOR ABDOMINAL WALL

General Considerations

The abdominal wall functions in many capacities, serving to pad, protect, and surround the abdominal viscera. Physiologically, the muscles can act to assist bending and posture. Along with the intercostal and neck muscles, the abdominal wall muscles provide accessory respiratory function when needed. Urination and defecation can be initiated and prolonged by voluntarily increasing intraabdominal pressures. Topographically, the anterior abdominal wall is bounded by the flare of the costal margins and the xiphoid process of the sternum above and by the iliac crests, inguinal ligaments, and pubis below. The principal structures that comprise the anterior abdominal wall are the rectus, external and internal oblique, and transversus abdominis and lower intercostal muscles together with their enveloping fascial sheaths and aponeuroses. The linea alba, a tendinous raphe formed by a blending of the aponeuroses of the oblique and transversus muscles in the midline, divides the anterior abdominal wall into two parts and restricts the medial extension of pathologic processes that may arise within it. Deep to the muscles is the continuous transversalis fascia, considered to be the strongest layer of the abdominal wall, and peritoneum.

The blood supply of the anterior abdominal wall is furnished by the superior and inferior epigastric, lower intercostal, lumbar, and iliac circumflex arteries. The venous drainage corresponds to the arteries. Lymphatics in the upper half of the abdominal wall drain to the axillary nodes and those in the lower abdomen to the inguinal and then to the iliac nodes. Studies using radionuclides have shown a broad band in the midabdominal region that may drain in either or both directions. Lymph flow around the umbilicus also may ascend around the ligamentum teres (obliterated umbilical vein) to reach the porta hepatis. The nerve supply is via the intercostal and upper lumbar nerves.

Surgical diseases of the anterior abdominal wall include (1) hernia, (2) infection, (3) primary and metastatic tumors of soft tissue and muscle, (4) rectus sheath hematoma, and (5) desmoid tumor. With the exception of rectus sheath hematoma and desmoid tumor, these conditions are covered in other chapters of this book.

Rectus Sheath Hematoma

Bleeding into the rectus sheath may simulate the acute surgical abdomen. The bleeding usually is the result of rupture of the epigastric artery or veins rather than a primary tear of the rectus muscle fibers. It is often a self-limiting condition, but if it is not recognized, it can lead to an unnecessary emergency surgical procedure.

Anatomy (Fig. 33-1)

The rectus abdominis muscle is crossed by three transverse tendinous intersections on its anterior aspect. Its contractile force is divided into three parts. The lowermost part is the longest; therefore, its shortening with contraction is the greatest. There may be a difference of as much as 18cm in length between extreme contraction and relaxation.

A strong fascial sheath, made up of the aponeuroses of the oblique and transversus abdominis muscles and transversalis fascia, contains the muscle. Anteriorly, the sheath is complete throughout; however, midway between the umbilicus and the pubis, the posterior sheath ends, forming an arched border, the linea semicircularis (of Douglas). Cephalad to this level, the internal oblique aponeurosis splits into two leaves, one passing on either side of the rectus, while below it no such division takes place, and together with the aponeurosis of the transversus abdominis, it passes anteriorly. The rectus muscle below the linea semicircularis is therefore separated from the abdominal viscera only by transversalis fascia and peritoneum. The anterior leaf of the sheath is adherent to the transverse tendons and also to the lateral and medial margins of the muscle; posteriorly, the muscle is free. As a consequence of this anatomic arrangement, when there is bleeding within the rectus sheath below the umbilicus, the free blood may lie against the peritoneum, producing irritation and pain and suggesting an acute intraabdominal process.

The blood supply to the rectus muscle is from the superior and inferior epigastric arteries. The superior epigastric artery enters the rectus sheath from above as a terminal branch of the internal mammary artery and passes caudad behind the muscle to anastomose with the larger inferior epigastric artery coming from below. The inferior epigastric artery, a branch of the external iliac artery, enters the rectus sheath just above the inguinal canal and courses upward along the posterior surface of the rectus muscle. Both arteries give off numerous muscular branches. Two veins accompany each artery. As the rectus muscle contracts, the epigastric vessels must glide beneath it to avoid injury.

Etiology

Rectus sheath hematoma may follow direct trauma to the epigastric blood vessels or occur spontaneously in association with several diseases. It also has been noted after a convulsive seizure. Spontaneous bleeding from the smaller muscular arteries has been reported in (1) infectious diseases, notably typhoid fever, (2) debilitating diseases, (3) collagen diseases, (4) blood dyscrasias such as hemophilia and leukemia, and (5) patients on anticoagulation therapy.

Frequently, however, bleeding occurs without obvious trauma or disease. In these patients, the hematoma usually follows minor straining as in coughing or sneezing. Presumably, the underlying factor is an inelasticity of the artery or vein that prevents the vessel from accommodating itself to the sudden marked variation in length that the rectus muscle undergoes during contraction and relaxation. Spontaneous rectus hematoma also has been described in pregnancy and in the puerperium. It is not known whether the hematoma is from venous or arterial bleeding. Stretching of the epigastric vessels by a distended abdomen during pregnancy and sudden relaxation after delivery are probably factors causing vessel injury. In the elderly, atheroma of an epigastric artery may predispose the vessel to rupture after minor exertion.

Clinical Manifestations

Rectus sheath hematoma is three times more frequent in women than in men. The condition is rare in children and has a peak age incidence in the fifth decade. A history of previous trauma, sudden muscular exertion, generalized vascular disease, or anticoagulation suggests the diagnosis. Hematomas related to anticoagulation therapy usually become apparent 4 to 14 days after treatment is instituted. The first symptom is pain. This is sudden in onset, sharp, and progressively severe. The pain is felt in the side of the abdomen where the bleeding occurs and remains localized, because the hematoma in the rectus muscle is limited by the confines of its sheath. Usually this is manifest in the lower abdomen and more often on the right side. Anorexia, nausea, but rarely vomiting, tachycardia, low-grade fever, and a moderate leukocytosis are frequent findings. With severe bleeding, signs of peripheral vascular collapse may develop. This is more apt to occur with bleeding below the linea semicircularis, where the peritoneum is only loosely adherent to the rectus muscle and cannot tampon the ruptured epigastric vessel. Tenderness and spasm are frequently present over the site of the hemorrhage. The bowel sounds usually are not altered. There may or may not be a palpable mass, depending on the extent of the bleeding. If present, the mass is tender, does not usually cross the midline, and remains palpable when the patient tenses the rectus muscle (Fothergill's sign). A bluish discoloration of the overlying skin is virtually diagnostic; however, this finding usually does not occur until 3 or 4 days after the patient is first seen.

Rectus sheath hematoma has been mistaken for almost every acute disease of the abdomen. Before the advent of ultrasound and computed tomography (CT), a correct preoperative diagnosis was made in less than 30 percent of patients. Using these scanning techniques, the condition can be diagnosed in most patients by the demonstration of a cystic or complex mass lesion within the confines of the rectus sheath (Fig. 33-2). Above the linea semicircularis, the hematoma is limited medially by the linea alba and confined to one side; below this level, the mass may project across the midline.

Treatment

If the diagnosis is made and the rectus hematoma is not causing severe symptoms, the condition may be managed nonoperatively with bed rest and analgesics. Anticoagulants should be discontinued. It is rarely necessary to abruptly reverse a coagulation deficit when operation is not undertaken. Surgical intervention occasionally is necessary to relieve symptoms of the hematoma or, if the diagnosis is in doubt, to rule out other more serious diseases. With a paramedian incision, the diagnosis becomes obvious when the rectus sheath is opened and free blood is found. The hematoma may be diffuse throughout the rectus muscle or may be a localized clot. Bleeding arteries may be present. Ideally, the hematoma is evacuated without entering the peritoneal cavity. Bleeding points are ligated, and the wound is closed without drainage. The prognosis depends on the underlying or concurrent disease but is generally good, and a full recovery can be anticipated.

Desmoid Tumors

The term *desmoid* is derived from the Greek and means “bandlike appearance.” Desmoid tumors in the past have been characterized as a variant of fibromatosis or fibroma, namely, fibrous tumors of slow growth. More recently, their locally aggressive behavior, their tendency to recur, and the high ultimate mortality rate of patients with fibroma have led them to be reclassified as low-grade malignant lesions, specifically low-grade fibrosarcomas. They are nonencapsulated and rise from the musculoskeletal aponeurotic tissues. Their course is variable, but they proliferate by local invasion and may produce obstruction with invasion of adjacent structures.

Historical Background

The first description of desmoid tumors was in 1832 by John Macfarlane of Glasgow. Muller introduced the term *desmoid* in 1838. Nichols reported the association of desmoids and gastrointestinal polyposis in 1923. The classic paper of Gardner, published in 1953, concerning “kindred 109” (kindred with familial polyposis) suggested the association between desmoid tumors and familial polyposis.

Etiology

Desmoid tumors occur spontaneously in individuals and in patients with familial adenomatous polyposis. The incidence in the general population is thought to be between 2 and 5 cases per 1 million population. Sporadic desmoid tumors occur primarily in the abdominal wall and in extraabdominal sites. They are juvenile or are associated with women of childbearing age, thereby raising the question of estrogen involvement in causation. This hypothesis is supported by in vitro laboratory studies and scattered reports noting the effectiveness of tamoxifen and toremifene in the treatment of desmoids.

Spontaneously occurring desmoid tumors are more common in females, with a ratio of 4.4 to 5:1. There are four major age categories, each demonstrating different biologic properties:

(1) *Juvenile desmoid tumors* — predominantly extraabdominal and in females

(2) *Fertile* — almost exclusively abdominal and in fertile females

(3) *Menopausal* — frequently abdominal with an almost equal sex ratio

(4) *Senescent* — occurring in males and females equally and in abdominal and extraabdominal sites.

In patients with *familial adenomatous polyposis (FAP)*, the incidence has been estimated to be between 6 and 31 percent, with an absolute risk of desmoid in the FAP patient at 2.56 per 1000 person-years, a risk 850 times that of the general population. The female-to-male ratio in this group is 1.4:1. The current molecular evidence is consistent with the Knudson antioncogene model. The familial polyposis gene has been mapped to chromosome 5q, and these patients frequently show allele loss on chromosome 5q. It is possible that the polyposis gene behaves as a tumor suppressor gene or antioncogene. For patients with familial polyposis, the first event is the loss of the allele on 5q. The APC gene has been cloned, and approximately 250 mutations have been identified. Studies have revealed a deletion in the (5)(q14q31) region that involves the region q21@22 where the FAP gene is localized in Gardner's syndrome. These abnormalities of the APC gene on the long arm of chromosome 5 frequently result in a truncated protein in which suppression may be lost.

There are numerous extraabdominal abnormalities in patients with FAP. Epidermoid cysts, osteomas, gastric polyps, benign and malignant lesions of the duodenum (specifically around the ampulla of Vater), papillary carcinomas of the thyroid, bilateral fibromatosis of the breast, adrenal adenomas, and pigmented ocular fundus lesions have been reported. Mutations in codons 1445 to 1478 can result in the development of desmoid tumors, but congenital hypertrophy of the retinal pigment epithelium is lacking. Mutations in codons 463 to 1387 regularly manifest congenital hypertrophy of retinal pigment epithelium. The genetic defects in these diseases are being rapidly elucidated.

Pathology

Desmoid tumors usually are solitary and reasonably well defined. They have a hard, rubbery consistency and on the cut surface exhibit a walled grayish white appearance, with white being predominant. The size varies from a few centimeters to a lesion that may weigh several kilograms (Fig. 33-3). There is no true capsule, but a pseudocapsule with multiple projections into the surrounding tissue. This feature (extension into the surrounding tissue) is critical to their surgical management.

Well-differentiated fibroblasts and fibrocytes make up the bulk of the tumor. The degree of cellularity can vary from moderate to almost acellular, but there are few or no mitoses. The cells infiltrate the surrounding muscle, giving rise occasionally to sarcolemmal giant cells.

Local invasion is very common. Symptoms usually are produced by compression or invasion of adjacent muscles, tendons, nerves, periosteum, bones, lymph nodes, and ureters. The tumor does not metastasize to lymph nodes and usually does not invade the skin. Thick-walled hamartoma vascular channels exist within the tumor. Desmoids have been reported to metastasize, but this is extremely rare and has never been observed by us. The differentiation between a desmoid and a low-grade fibrosarcoma is difficult and depends on local behavior and long-term prognosis.

Diagnosis

Because the tumor often interrupts intramuscular and soft tissue planes, there may be some clue to the diagnosis. Invasion of the adjacent muscle and bone may give rise to a periosteal reaction.

On plain abdominal radiographs, visceral displacement or obstruction can occur. When contrast-enhanced studies are performed, thinning, spiculation, and even an occasional extravasation of intraluminal material into the desmoid, occasionally presenting as an abscess, may occur. On excretory urography, displacement of the ureter and compression may be seen.

On ultrasound, the desmoid is a nonspecific, hypoechogenic mass with poor enhancement. It is indistinguishable from rhabdomyosarcoma, hematoma, abscess, or metastatic disease unless there is gas in the abscess. Angiography reveals the expected arterial displacement, occasional neovascularity, and occasional tumor staining. Arterial stretching is the most likely finding. A fibrotic reaction along mesenteric arteries is the most common early manifestation of desmoids.

The most useful radiologic examinations are computed tomography (CT) and magnetic resonance imaging (MRI). These studies usually show a homogeneous soft tissue mass (bright on a T2-weighted image) in which the adjacent viscera are displaced. This can be confused with other soft tissue tumors and lymphoma of the mesentery, which usually has associated retroperitoneal adenopathy.

Management

In a patient with FAP without cancer who develops a soft tissue mass within the first 2 years after colectomy that is adherent to the scar or in the mesentery, biopsy is not indicated. In patients with a spontaneous soft tissue mass in whom there is no reason to suspect a desmoid or in patients with extremity desmoids, adequate tissue biopsy is essential. A longitudinal, incisional biopsy for an extremity mass, which will be excised with the definitive tumor resection, should be performed.

Operative Management

Although ancillary pharmacologic and radiation management has been advocated, complete surgical extirpation of the tumor is the only therapy resulting in long-term remission. The principle is complete excision with a 1- or 2-cm margin of all tissue. Desmoid tumors often invade the surrounding tissue with various pseudopods of microscopic extension; therefore, it is essential to get an adequate tissue margin. Desmoids represent a low-grade fibrosarcoma with a propensity for local recurrence after excision. Every effort must be made at the time of the first operation to eradicate all known disease, which will not always be technically feasible due to the size or location of the tumor.

The following features are contributory to local failure: (1) age between the late teens and 30 years, (2) recurrent disease at presentation, (3) inadequate excision at the time of the first operation, and (4) radiation therapy not administered for gross residual disease.

The extent of the excision must be radical. Thoracoabdominal wall excisions and large abdominal or flank tumor excisions frequently necessitate the use of musculocutaneous flaps to close the defect. Desmoids of the chest wall should be managed with excision of the skin, underlying muscle, and the chest wall, including rib resections. A plastic surgeon should be available to perform the myocutaneous flap and closure. Where possible, artificial mesh should be avoided for reconstruction, especially when there are small bowel anastomoses. If necessary for reconstruction, the underlying viscera should be protected from the foreign mesh material.

Intraabdominal desmoids are difficult to treat because they frequently arise along the superior mesenteric artery and its primary branches, sometimes necessitating massive small bowel resection. An arteriogram can be obtained to track the relation of the vessels to the mass. Radial incisions in the mesentery are performed, identifying the vasculature along the superior mesenteric artery and vein and taking vascular branches in turn. If the situation is favorable, fairly extensive small bowel resection can be undertaken, leaving the blood supply to the proximal and distal small bowel. Desmoids may occur in the mesentery of patients who previously underwent an ileal pouch/anal anastomosis procedure, creating a challenging situation.

For pelvic primary desmoids or recurrences, pelvic exenteration and hemipelvectomy have been performed with discouraging results. Adjunctive therapy with radiation for microscopic or gross disease is essential to prevent local recurrence.

Radiation Therapy

In patients in whom desmoid tumors are unresectable or in patients with gross disease left at the margins, radiation therapy is recommended. At least 55 Gy is necessary, with tumor doses ranging from 4.8 to 61.2 Gy, given in doses of 1.6 to 1.8 Gy per fraction, 5 fractions per week. The port should be designed to cover the tumor bed with a generous overlap margin.

Responses to radiation therapy are varied. Some authors have described the benefits of intraabdominal radiation therapy to the ablation of ovarian function secondary to radiation with the corresponding loss of estrogen being responsible for tumor shrinkage.

Chemotherapy

The behavior of desmoid tumors in response to chemotherapy is erratic. Most authorities claim that it takes up to 27 months for desmoid tumors to respond to chemotherapy. A number of drugs have been used. Current therapy includes a combination regimen of vincristine, actinomycin D, and cyclophosphamide. Most reports are anecdotal, and many patients do not respond. Steroids may or may not be used as an adjunct. Their effect is thought to be secondary to inhibition of prostaglandin synthesis, which is essential to the growth of desmoids.

Pharmacologic Treatment

A number of drugs have been used in the treatment of desmoids, including polynucleotides and theophylline, which are effective in decreasing intradermal growth of tumor cells in mice. Indomethacin with high-dose ascorbic acid also has been used to treat desmoid tumors because of their effect on cyclic AMP and inhibition of ornithine decarboxylase.

Sulindac and other nonsteroidal anti-inflammatory drugs, in conjunction with coumadin, also have been reported to cause regression of desmoid tumors. In our experience, these are ineffective in controlling desmoid tumors. Because estrogens are thought to be involved in causing desmoid tumors, tamoxifen and some of the newer derivatives, such as tomorifene, also are used.

Other combinations of therapy include dicarbazine and doxorubicin, which have been used in the management of unresectable tumors. Iridium-192 implantation for locally advanced and recurrent tumors is advocated.

Some authors advocate no therapy, but this is untenable because of the reclassification of desmoid tumors as low-grade, invasive fibrosarcomas. Patients with FAP should be screened with CT scans after total colectomy because up to 30 percent (depending on the kindred) may manifest desmoid tumors in the surgical scar and in the mesentery.

Prognosis

Approximately 85 to 90 percent of patients will be alive at 5 years after the initial desmoid resection, but many of these patients are less than 30 years of age and may die of their disease over the next 10 years because of inexorable growth of tumor recurrences.

Desmoid tumors of the head and neck pose specific problems. They are found most commonly in the neck area, and radical neck dissection has been advocated at the first appearance. Radical en bloc resection is required with a margin of surrounding normal tissue. Desmoids in the orbit require exenteration, plastic reconstruction, and prosthetic materials. The incidence of recurrence is higher in head and neck lesions.

DISEASES OF THE OMENTUM

General Considerations

The greater omentum consists of a double sheet of flattened endothelium. Between the folds, the epiploic vessels, lymphatics, and nerves pass in areolar tissue enmeshed with a variable amount of fat. The structure hangs in a double fold, or sling, between the greater curvature of the stomach and the transverse colon. At birth, an agglutination of the two layers occurs, creating an apron-like shield overlying the intestinal coils. The right border attaches to the pylorus or first portion of the duodenum, while the left border forms the gastrosplenic ligament. The right side usually is longer and heavier and may possess tongue-like processes extending into the pelvis. Occasionally, accessory omenta exist attached to the main portion. The size of the greater omentum is related to the amount of fat it contains. It is often huge in obese individuals and very thin and small in emaciated persons. The omentum in infants usually is underdeveloped and may be almost nonexistent. With growth of the individual, there is elongation and thickening of the organ from the deposition of fat within its layers.

As a peritoneal fold, the omentum assumes the mechanical function of a mesentery, i.e., the fixation of viscera and the transmission of a vascular supply. It is not a vital organ and can be removed without disturbance to the individual.

The omentum possesses an inherent motility that allows it to seek out and arrest trouble that may arise within the peritoneal cavity. It has been referred to as the “police officer of the abdomen.” The omentum often is found at the site of an intraabdominal pathologic condition, but evidence shows that it has no spontaneous or ameboid activity and that displacement occurs as a result of intestinal peristalsis, diaphragmatic excursions, and postural changes of the individual. The areolar tissue is rich in macrophages that rapidly remove injected bacteria or foreign particles. The usefulness of the omentum in inflammatory processes is related to its bactericidal and absorptive properties and its ability to form adhesions.

Torsion

Torsion of the omentum is a condition in which the organ twists on its long axis, causing vascular compromise. This may vary from mild vascular constriction, producing edema, to complete strangulation, leading to infarction and frank gangrene. For torsion to occur, two situations must exist: (1) a redundant and mobile segment and (2) a fixed point around which the segment can twist.

Etiology

Omental torsion has been classified as primary or secondary. Primary, or idiopathic, omental torsion is relatively rare, and the cause is obscure. Leitner and associates group the causes of primary torsion into predisposing factors and precipitating factors. Among the suggested predisposing factors are a variety of anatomic variations including tongue- like projections from the free edge of the omentum, bifid omentum, accessory omentum, a large and bulky omentum with a narrow pedicle, and obesity associated with irregular distribution of fat within the organ. Venous redundancy relative to the omental arterial blood supply also has been cited as a predisposing factor. The omental veins are larger and more tortuous than the arteries, allowing venous kinking and offering a point of fixation around which twisting can occur. The higher incidence of right-sided omental torsion is related to the greater size and mobility of the right omentum.

Precipitating factors are those which cause displacement of the omentum. These include heavy exertion, sudden change in body position, coughing, straining, and hyperperistalsis with overeating. Primary omental torsion is always unipolar; there is only one locus of fixation.

Secondary omental torsion is associated with adhesions of the free end of the omentum to cysts, tumors, foci of intraabdominal inflammation, postoperative wounds, or scarring, or internal or external hernias. It is more common than the primary type and usually is bipolar; torsion of the central portion occurs between two fixed points. About two-thirds are found in patients with hernias, usually of the inguinal variety. The precipitating factors that cause secondary torsion are the same as those for primary torsion.

Pathology

The omentum in the primary and secondary varieties twists a variable numbers of turns around a pivotal point, usually in a clockwise direction (Fig. 33-4). The whole omentum or, more often, a small portion may undergo torsion. The right side is involved more frequently. Venous return is restricted, and the distal omentum becomes congested and edematous. Hemorrhagic extravasation results in a characteristic serosanguineous effusion into the peritoneal cavity, which, if of sufficient duration, causes acute hemorrhagic infarction and eventual necrosis. If not excised, the mass becomes atrophied and fibrotic and on rare occasions is autoamputated.

Clinical Manifestations

The clinical features of primary and secondary omental torsion are similar. The condition usually occurs in the fourth or fifth decade of life. Males are affected twice as frequently as females. Pain is the initial and predominant symptom. The onset of pain usually is sudden and constant with a gradual increase in severity. Occasionally, the pain is experienced first in the periumbilical region or is generalized. Invariably it becomes localized to the right side of the abdomen, usually the right lower quadrant. This is in keeping with the more frequent involvement of the right side of the omentum. Movement intensifies the pain. Nausea and vomiting occur in less than half the patients. There is moderate leukocytosis and fever that rarely exceeds a rise of 1°C. Tenderness is present, and rebound tenderness and voluntary spasm are frequent findings. A mass may be palpable if the involved omentum is sufficiently large.

The manifestations usually are not sufficient to allow an accurate diagnosis. CT scan has been used in diagnosing omental pathology. This can be a sensitive modality for showing an omental mass but is not usually specific for making a diagnosis of omental torsion. Secondary torsion of the hernial type can be suspected if a tender mass is palpable in the groin. Impressions, in order of frequency, are acute appendicitis, acute cholecystitis, and twisted ovarian cyst. Manifestations usually justify abdominal exploration. The finding of free serosanguineous fluid at the time of laparotomy in the absence of a pathologic condition in the appendix, gallbladder, or pelvic organs should suggest the possibility of omental torsion.

Treatment

Treatment consists of resection of the involved omentum. In patients with secondary torsion, the underlying condition (hernia, cysts, adhesions, etc.) also should be corrected. Postoperative mortality and morbidity are rare.

Idiopathic Segmental Infarction

Idiopathic segmental infarction is an acute vascular disturbance of the omentum. The criteria for diagnosis of this condition are that it is not accompanied by omental torsion, that there is no associated cardiovascular disease or local intraabdominal pathologic condition, and that there is no history of external abdominal trauma, situations that produce secondary omental infarction.

Etiology and Pathology

The condition is precipitated by thrombosis of omental veins secondary to endothelial injury. Halligan and Rabiah summarized the several proposed causes of endothelial damage and thrombosis. These include (1) stretching or primary rupture of the omental veins by a sudden increase in intraabdominal pressure such as that caused by coughing, sneezing, or lifting, especially after ingesting a heavy meal, (2) gravitational pull of an extremely fatty omentum on the omental veins, causing their rupture, and (3) an anatomic peculiarity of the venous drainage of the omentum that predisposes to thrombosis.

The right lower segment of the omentum, which is the most mobile and richest in fat, is the portion usually involved. The area of infarction may vary from 2 to 20 cm at its greatest diameter. Grossly, the involved segment is well demarcated, edematous, and hemorrhagic or gangrenous. It usually is closely adherent to the parietal peritoneum or adjacent abdominal viscera. Microscopically, there is a hemorrhagic infarction with thrombosis of the omental veins and filtration of the omentum with inflammatory cells.

Clinical Manifestations

Most patients are young or middle-aged adults, and there is a 3:1 predilection for males. The manifestations are nonspecific. Patients usually present with a gradual onset of abdominal pain that is steady and virtually always on the right side of the abdomen. Anorexia and nausea are frequent, but vomiting is rare. Diarrhea or constipation is unusual. There is always tenderness and often rebound tenderness over the region of the infarction. Voluntary guarding and, occasionally, spasm also are common. The infarcted segment, if large enough, may be palpable. A slight fever (rarely >38.5°C) and a moderate leukocytosis are usual. The diagnosis can be suspected by finely infiltrated fat giving a “smudged” appearance to the omentum on CT scan.

Treatment

Treatment of this condition is resection of the infarcted area to prevent the possible complications of gangrene and adhesions. A correct preoperative diagnosis is unusual, and most patients are explored for acute appendicitis or acute cholecystitis. The finding of serosanguineous fluid in the abdomen and a normal appendix or gallbladder should indicate disease in the omentum. Postoperative complications are rare.

Cysts

Pathology

Cysts of the omentum are rare. The pathogenesis of these lesions is unclear, but presumably most true cysts are caused by obstruction of lymphatic channels or by growth of congenitally misplaced lymphatic tissue that does not communicate with the vascular system. They contain serous fluid and may be unilocular or multilocular. The cysts have an endothelial lining similar to cystic lymphangiomas. Their size may vary from a few centimeters to over 30 cm in diameter. Dermoid cysts, which are rare, are lined with squamous epithelium and may contain hair, teeth, and sebaceous material.

Pseudocysts of the omentum result from fat necrosis, trauma with hematoma, or foreign-body reaction. These have a fibrous and inflammatory lining and usually contain cloudy or blood-tinged fluid.

Clinical Manifestations

True omental cysts are discovered most frequently in children or young adults but have been reported in the aged. Small cysts generally are asymptomatic and discovered

incidentally at laparotomy or at autopsy. Large cysts present as a palpable abdominal mass or produce diffuse abdominal swelling. These may cause symptoms of heaviness or pain or manifestations of possible complications of omental cysts such as torsion, infection, rupture, or intestinal obstruction. Complications are more frequent in children and often produce a clinical picture of an acute surgical condition of the abdomen. The uncomplicated omental cyst usually lies in the lower midabdomen and is freely movable, smooth, and nontender.

Plain radiographs sometimes show a circumscribed soft tissue haziness in the abdomen, or after a barium meal, there may be displacement of intestinal loops with pressure on adjacent bowel. The presence of bone or teeth is diagnostic of dermoid cyst. The ultrasound or CT scan shows a fluid-filled mass that often contains internal septations.

Differential diagnosis includes cysts and solid tumors of the mesentery, peritoneum, and retroperitoneal region. An absolute diagnosis can be made only at the time of exploratory surgical procedures. Treatment consists of local excision.

Solid Tumors

The most common solid tumor of the omentum is metastatic carcinoma, which generally involves the omentum by tumor implantation. The primary source is usually the colon, stomach, pancreas, or ovaries. There often is ascites, presumably from “weeping” of serous or blood-tinged fluid from the metastatic implants. Diffuse neoplastic infiltration of the greater omentum produces a distinctive CT scan of a soft tissue mass (“omental cake”) separating the colon or small intestine from the anterior abdominal wall (Fig. [33-5](#)).

Primary solid tumors of the omentum are exceedingly rare. They may be benign or malignant. Stout and associates recorded only 24 seen over a 55-year period at a major cancer center. Most are of mesenchymal origin, and about one-half are malignant. Benign tumors consist of lipomas, leiomyomas, fibromas, and neurofibromas. Leiomyosarcoma and hemangiopericytoma are the most common malignant tumors. The malignant tumors spread by direct extension or tumor implants and kill by involvement of vital abdominal organs. The mean age of patients with primary omental tumors is in the fifth decade. Although some patients are symptom-free, about one-half complain of vague abdominal pain. A palpable abdominal mass is present in one-third of patients.

The only treatment is surgical excision. Primary malignant tumors are highly invasive and often require resection of adjacent organs and total omentectomy. The prognosis is poor. Resection of benign tumors is curative, and recurrences have not been reported. Palliative omentectomy for metastatic tumor implants in the omentum has been suggested to control any associated ascites.

MESENTERY AND MESENTERIC CIRCULATION

Mesentery

The mesentery is a reflection of the posterior parietal peritoneum onto the surface of the intestine, where it becomes visceral peritoneum. It connects the intestine to the posterior abdominal wall and transmits blood vessels and nerves. In addition, the mesentery serves as a suspensory ligament of the jejunum and ileum. It is fan-shaped, with the root extending downward and obliquely from the ligament of Treitz (duodenojejunal flexure) at the level of L₂ to the right sacroiliac articulation (ileocecal junction) (Fig. [33-6](#)). The entire root is only about 15 cm in length and allows free motion of the small intestine in any direction, limited only by the length of the mesentery. Within its two fused layers of peritoneum run the intestinal branches of the superior mesenteric artery and accompanying veins, along with lymphatics, mesenteric lymph nodes, visceral nerve fibers, and a variable amount of adipose tissue.

After the embryonic formation of a distinct intestinal loop, torsion of the loop takes place about the superior mesenteric artery. About the third or fourth fetal month, posterior peritoneal fixation of the colon takes place. The leaves of the mesentery to the ascending colon fuse with the right parietal peritoneum, and those to the descending colon fuse with the left parietal peritoneum. The posterolateral parietal peritoneum then passes directly from the abdominal wall over the ascending and descending colon, respectively, toward the midline and the root of the mesentery proper. For this reason, the mesentery to these portions of the large intestine usually is short or nonexistent. These fusions, however, form surgical cleavage planes allowing bloodless mobilization of the colon with its vascular supply. Within the embryonic mesentery to the ascending colon are the colonic arteries and veins from the superior mesenteric vessels, while those to the descending colon are derived from the inferior mesenteric vessels. On occasion, posterior fusion of the ascending or descending colon is incomplete or does not occur, leaving a well-developed mesentery and allowing free mobility of segments of the large bowel. This anomaly is more frequent with the right colon, predisposing to torsion with resulting intestinal obstruction.

The mesenteries of the transverse colon and sigmoid colon, in contrast to those of the ascending and descending colon, do not fuse with the posterior parietal peritoneum. These remain well developed and are referred to as the *transverse mesocolon* and *sigmoid mesocolon*, respectively. The segment to the transverse colon extends obliquely across the posterior abdominal wall just below the pancreas, remaining fixed at the hepatic and splenic flexures of the bowel (see Fig. [33-6](#)). The fixation of the splenic flexure is higher than that of the hepatic flexure because of the presence of the liver on the right side. The mesocolon allows the transverse colon to hang over the small intestine. This sagging may be so marked that the transverse colon occasionally reaches the symphysis pubis. Within the transverse mesocolon run branches of the middle colic artery and accompanying veins. Fusion between the mesocolon and the undersurface of the greater omentum from the stomach offers stability that prevents the transverse colon from undergoing torsion.

The sigmoid mesocolon originates at the end of the descending colon in the left iliac fossa and has an inverted V-shape course. It runs diagonally upward along the left iliac artery toward the aortic bifurcation and then bends directly downward into the pelvic fossa, where it is reflected off the rectum. It contains sigmoid vessels and branches of the superior hemorrhoidal vessels from the inferior mesenteric artery and vein. The length of the sigmoid mesocolon determines the location and mobility of the pelvic colon. If the sigmoid mesocolon is long, the bowel may cross the midline. Such a mobile pelvic colon may twist on itself. The sigmoid colon is the most frequent site of torsion producing volvulus of the intestinal tract.

The lateral fixations of the ascending and descending colon and the superior origin of the transverse mesocolon serve to confine the small intestine within the midabdomen. The transverse mesocolon and greater omentum also restrict the small bowel from entering the upper abdomen to become adherent to inflammatory lesions of the stomach, duodenum, gallbladder, or liver.

Defects in the mesenteries are potential sites for internal hernia. Most defects are created inadvertently by the surgeon during the course of intraabdominal operations. On rare occasions, congenital defects occur in areas of the mesentery that are thin and avascular. These usually are found in the mesenteries of the lower ileum, the sigmoid mesocolon, and the transverse mesocolon, the last through a wide avascular space just to the left of the middle colic artery (space of Riolan).

The mesenteries, with the omenta, share bactericidal and absorptive properties and an ability to form adhesions. In this regard, they function to localize and combat intraperitoneal infection and to seal intestinal perforations.

Mesenteric Circulation

The mesenteric vascular bed is of major importance in the maintenance of bodily homeostasis. Under resting conditions, the splanchnic (visceral) vascular bed receives 25 to 30 percent of cardiac output and contains as much as one-third of total blood volume. It has been suggested that this reservoir of blood produces a mechanism for “autotransfusion” during periods of hypovolemia when a relatively large volume of blood can be released rapidly into the circulation by active constriction of the splanchnic vessels. Control of the mesenteric vascular bed is primarily neural via sympathetic autonomic elements carried by the splanchnic nerves. These nerves accompany the celiac, superior mesenteric, and inferior mesenteric arteries and contain a- and b-adrenergic receptors. Stimulation of the splanchnic nerves produces vasoconstriction, with an increase in regional resistance. The mesenteric vasculature also is responsive to a number of pharmacologic agents. Norepinephrine, an a-adrenergic stimulator, produces vasoconstriction, while epinephrine elicits a classic dose-dependent b- or a-adrenergic response, low concentrations producing vasodilatation and higher concentrations producing vasoconstriction. Isoproterenol, a b-adrenergic stimulator, effects a dilator response that can be blocked by propranolol. Nitroglycerine, tolazoline hydrochloride, and papaverine hydrochloride elicit a direct vasodilatory effect; the direct infusion of these agents can be useful in the treatment of nonocclusive mesenteric ischemia. In contrast, the digitalis glycosides produce mesenteric vasoconstriction and may contribute to the pathogenesis of nonocclusive mesenteric ischemia.

Arteries

With the exception of the stomach and duodenum, which are supplied by the celiac artery, and the distal rectum, which is supplied by the inferior hemorrhoidal vessels arising from the internal iliac arteries, the arterial supply to the entire intestinal tract is derived from the superior and inferior mesenteric arteries. The superior mesenteric artery arises from the anterior aspect of the aorta just below the celiac artery at the level of the second lumbar vertebrae. Owing to their anterior origin, lateral aortography is necessary to visualize the origins of the celiac and superior mesenteric arteries. Accurate knowledge of the anatomy of the visceral vessels is of critical importance in the evaluation and treatment of visceral ischemia in the clinical setting. The superior mesenteric artery passes behind the neck of the pancreas but in front of the uncinate process and crosses in front of the third portion of the duodenum to enter the root of the mesentery. The acute angle at which the superior mesenteric artery arises from the aorta may compress the transverse portion of the duodenum between it and the aorta, causing partial intestinal obstruction, a condition referred to as the *superior mesenteric artery compression syndrome*. In addition, this acute angle makes it the most susceptible of the visceral vessels to cardioarterial embolism.

As the superior mesenteric artery continues downward between the two leaves of the mesentery, it gives off 12 or more major branches from its left side to supply the jejunum and ileum (Fig. [33-7](#)). These jejunal and ileal arteries divide and then reunite within the mesentery to form groups, or arcades. Two to five such anastomotic arches are formed and allow collateral pathways for blood to reach the intestinal wall should occlusion of short arterial segments occur. The arcades become more numerous as the terminal ileum is reached. From the terminal arcades, straight branches (vasa recta) alternately pass to opposite sides of the jejunum and ileum. Within the intestinal wall, the vessels run parallel to the circular muscle coat and perpendicular to the direction of the lumen, traversing successively the serous, muscular, and submucosal layers. Each of these terminal arteries supplies only 1 or 2 cm of bowel length. For this reason, they must be preserved as close to the cut margins of the intestine as is technically possible when performing a bowel resection to avoid subsequent necrosis and breakdown of the anastomosis. The terminal straight arteries do not anastomose until reaching the submucous plexuses, where their ramifications anastomose freely. This situation predisposes to serious compromise of the blood supply to the antimesenteric border of the intestine after segmental small bowel resection. To ensure adequate circulation to the antimesenteric portion, it is customary to transect the small intestine obliquely rather than at a right angle (Fig. [33-8](#)).

Arising from the right side of the superior mesenteric artery is the inferior pancreaticoduodenal artery and then, successively, the middle colic, the right colic, and the ileocolic arteries (see Fig. [33-7](#)). Except for the ileocolic artery, these vessels do not form anastomotic arcades until nearly reaching the bowel wall.

The middle colic artery arises below the pancreas, enters the transverse mesocolon, and passing to the right, divides into a right and left branch. The right branch connects with the superior branch of the right colic artery and the left branch with the ascending branch of the left colic artery from the inferior mesenteric artery. It supplies the transverse colon. The location of the main arterial trunk to the right of the midline allows the left side of the transverse mesocolon to be opened through a relatively avascular area (space of Riolan) when performing a retrocolic gastrojejunal anastomosis.

The right colic artery arises just below the middle colic artery and passes to the right just behind the peritoneum. On reaching the midascending colon, it divides into superior and inferior branches, which anastomose, close to the bowel wall, with branches from the middle colic and ileocolic arteries, respectively. The right colic artery supplies the ascending colon.

The ileocolic artery is the terminal branch of the superior mesenteric artery. It supplies several centimeters of ileum, the cecum, the appendix, and the lower portion of the ascending colon. It terminates by dividing into ascending and descending branches. The ascending branch anastomoses with the inferior branch of the right colic artery, while the descending branch forms secondary and tertiary arcades by anastomosing with terminal branches of the superior mesenteric artery within the mesentery proper. From these arcades arise the appendicular artery to the appendix and cecal and ileal branches.

The superior mesenteric artery supplies the intestinal tract (midgut) from the third portion of the duodenum to the midtransverse colon. Collaterals between the inferior pancreaticoduodenal artery and the superior pancreaticoduodenal artery from the gastroduodenal, a secondary branch of the celiac artery, enable the third part of the duodenum and proximal 10 to 12 cm of jejunum to survive when the superior mesenteric artery is occluded.

The inferior mesenteric artery supplies the left transverse colon, descending colon, sigmoid colon, and proximal part of the rectum. It arises from the anterior aspect of the aorta at the level of the third lumbar vertebrae and passes downward and to the left, entering the pelvis as the superior hemorrhoidal artery (Fig. 33-9). As it descends, it gives off the left colic and sigmoidal arteries. The left colic artery is the principal branch. It divides into ascending and descending limbs that anastomose with branches from the middle colic and sigmoid arteries, respectively. The sigmoid artery passes into the sigmoid mesocolon and divides into branches that anastomose with one another, forming several arcades. The lowest sigmoid arcade joins with arcades from the superior hemorrhoidal artery. The superior hemorrhoidal artery continues downward behind the rectum, where it communicates with branches from the middle and inferior hemorrhoidal arteries from the internal iliac artery, giving the rectum a dual source of arterial supply.

The anastomoses between primary branches of the superior and inferior mesenteric arteries form an arcade that passes along the margin of the colon and is referred to as the *marginal artery of Drummond* (see Fig. 33-9). It is situated about 1 cm from the margin of the bowel and extends from the end of the ileum to the end of the sigmoid colon. Through its anastomoses, it generally is capable of supplying the bowel even though one of the major arteries is ligated.

Veins

The venous drainage of the small intestine and colon is through tributaries of the inferior and superior mesenteric veins, which in turn ultimately terminate in the portal vein (Fig. 33-10). The portal circulation begins within the mucosa of the intestine. Small venules coalesce, and the confluent veins pass through the wall of the intestine, emerging alternately in a similar manner to that of the straight arteries entering the bowel wall. These then converge to form a system of venous arcades within the mesentery from which blood enters the main tributaries to the superior and inferior mesenteric veins.

The inferior mesenteric vein is a continuation of the superior hemorrhoidal vein. It passes upward to the left side of the inferior mesenteric artery, receiving tributaries that correspond in name and location to the branches of the artery. The main trunk of the vein does not accompany the artery but rather courses over the duodenojejunal flexure just lateral to the ligament of Treitz and, passing over the body of the pancreas, joins with the splenic vein (see Figs. 33-9 and 33-10). It drains the left side of the large intestine from the upper rectum to the left midtransverse colon. A plexus of anastomoses around the midrectum between the superior hemorrhoidal vein and the middle and inferior hemorrhoidal veins to the internal iliac veins forms a collateral pathway between the portal and systemic circulations. The inferior mesenteric vein generally can be ligated with impunity.

The superior mesenteric vein runs within the mesentery proper lateral to the superior mesenteric artery. It receives tributaries that accompany corresponding branches of the superior mesenteric artery and that drain the entire small intestine and right half of the colon. As it passes over the third portion of the duodenum and behind the neck of the pancreas, it receives the confluence of the inferior mesenteric and splenic veins to become the portal vein.

The venous drainage from the entire gastrointestinal tract passes through the liver via the portal circulation before returning to the heart. Together with the mesenteric lymphatics, it represents the sole means by which ingested food products find their way into the circulation. The normal portal venous pressure is between 12 and 15 cmH₂O; that within the inferior vena cava (systemic pressure) varies between a positive pressure of about 3 cmH₂O during the expiratory phase of respiration to a negative pressure of 1 to 3 cmH₂O during inspiration. Like the vena cava, the portal system does not contain valves, and therefore, the blood flows in the direction of reduced venous pressure.

Lymphatics and Lymph Nodes

The lymph drainage of the small intestine and colon follows the course of the main blood vessels. Those accompanying the inferior mesenteric artery drain to periaortic nodes and then to the superior mesenteric nodes before entering the cisterna chyli of the thoracic duct. Those accompanying branches of the superior mesenteric artery drain into the mesenteric glands within the mesentery, where they are closely related to the vascular arcades. The mesenteric nodes are distributed in three locations: (1) juxtaintestinal, at the last anastomotic branch of the mesenteric arteries before they enter the intestines, (2) intermediate, in the region of the larger anastomosing branches, and (3) central, at the root of the mesentery near the origin of the main mesenteric artery. The nodes are more numerous in the right half of the mesentery, and they increase in size and number as they approach its root. These nodes are the usual site for mesenteric adenitis, tuberculosis, and other inflammatory and neoplastic conditions. From the mesenteric nodes, lymph drains into the superior mesenteric and celiac nodes and then to the thoracic duct.

MESENTERIC VASCULAR DISEASE

Mesenteric vascular disease is not a single entity but rather a syndrome that includes (1) complete occlusion or stenosis of mesenteric arteries by embolism, thrombosis, or obliterative disease, (2) thrombosis of mesenteric (portal) veins, (3) extraluminal obstruction of mesenteric arteries by aortic aneurysm, dissecting aneurysm, fibrous and ligamentous bands, or tumors, (4) aneurysms of the splanchnic arteries, and (5) traumatic injury to visceral vessels. These conditions produce vascular insufficiency or infarction of the affected intestine. Intestinal disease because of impaired circulation is relatively uncommon when compared with the more frequently occurring mechanical obstructions of the mesenteric vessels by adhesive bands, strangulated hernia, and intestinal volvulus.

Occlusions of the mesenteric arteries may be acute and complete (those resulting from emboli or thrombosis), gradual and partial (those resulting from obliterative arterial disease), or acute and complete superimposed on a previously narrowed or stenotic vessel. Collateral vessels permit gradual occlusion of either the celiac or the superior mesenteric artery to be tolerated. Acute occlusion of the celiac or inferior mesenteric artery generally is asymptomatic in an otherwise normal person; acute occlusion of the superior mesenteric artery, if untreated, results in intestinal infarction and death. The superior mesenteric artery at its origin or close to the takeoff of its middle colic branch is the usual site of acute and chronic mesenteric arterial occlusions. Complete occlusion of the inferior mesenteric artery produces symptoms only if there is compromise of collateral blood flow from the superior mesenteric or internal iliac (hypogastric) artery. Clinically apparent venous occlusions are sudden and complete and invariably a consequence of thrombosis. Partial mesenteric venous occlusion usually is the result of external compression and is asymptomatic.

The relative incidence of mesenteric arterial as opposed to venous occlusions is not known. When intestinal infarction occurs, it is not unusual to find thrombosis of both sides of the splanchnic circulation at laparotomy or autopsy because initial occlusion of one eventuates in clot formation in the other. The clinical distinction between the two is often difficult. It has been variously estimated that 15 to 20 percent of all significant mesenteric vascular accidents are a result of primary venous thrombosis and that approximately 50 percent are a result of primary arterial occlusion. In the remaining 30 to 35 percent of cases, intestinal infarction occurs in the absence of major arterial or venous occlusion.

Acute Occlusive Visceral Ischemia

Etiology

Acute occlusion of the superior mesenteric artery may be the result of cardioarterial embolus or in situ thrombosis, with the incidence of each being approximately equal. The steep angle in which the superior mesenteric artery arises from the aorta makes it the most likely target among the visceral vessels to receive embolic material from the heart. Typically, emboli lodge at the origin of the superior mesenteric artery or at the origin of the major distal branch vessels, such as the middle colic, right colic and ileocolic arteries (Fig. 33-11). Thrombotic occlusion of the superior mesenteric artery typically occurs at the vessel origin, as a result of progression of a preexisting stenotic atherosclerotic plaque to complete occlusion. The atherosclerotic plaque generally arises in the visceral bearing segment of the aorta and “overhangs” the origin of the visceral vessels. For this reason, atherosclerosis of the visceral arteries is localized to the vessel origin. While this is nearly always true of the celiac artery, it is not uncommon for the atherosclerotic process in the superior mesenteric artery to extend distally over a length of 4 to 5 cm. Less common causes of stenosis and occlusion of the visceral arteries include Takayasu's arteritis, periarteritis nodosa, and thromboangiitis obliterans. Extrinsic compression of the celiac artery by diaphragmatic fibers (median arcuate ligament syndrome) may produce a chronic visceral pain syndrome.

Whatever the cause, the initial effect of proximal occlusion of the superior mesenteric artery is to cause intense spasm of its distal branches, exacerbating the ischemic insult. Acute occlusion of the origin of the superior mesenteric artery produces ischemia of the small intestine from the level of the ligament of Treitz to the ileocecal valve and of the ascending colon and proximal two-thirds of the transverse colon. Patency of the celiac artery can result in viability of the proximal 10 to 12 cm of jejunum owing to anastomoses between the superior pancreaticoduodenal artery (celiac-based) and the inferior pancreaticoduodenal artery (first branch of the superior mesenteric artery). Acute occlusion of distal branch vessels, such as the middle colic, right colic, and ileocolic arteries, results in segmental intestinal ischemia that might infarct depending on the status of the collateral circulation.

Pathology

High-grade partial or complete occlusion of the proximal superior mesenteric artery results in intense vasospasm throughout the mesenteric vascular bed, even in areas that would otherwise continue to be directly perfused either by branch vessels originating proximal to the occlusion or by the inferior mesenteric distribution. This observation constitutes the basis for specific catheter-directed pharmacologic therapy.

Because the mucosa is the layer of the intestinal wall most sensitive to ischemia, mucosal sloughing and ulceration, often manifest as gastrointestinal bleeding, are early pathophysiologic events that may be recognized endoscopically. Acutely ischemic intestine will appear pale, and pulsation in the mesentery, which is generally visible and palpable, will be absent. Peristalsis may persist for several hours after the onset of ischemia; the presence of bowel contractility should not dissuade from the diagnosis of acute intestinal ischemia when the findings of pallor and pulselessness are present. As the ischemic process progresses to infarction, over a period of approximately 6 h in the setting of profound complete ischemia, the bowel wall becomes dusky, then cyanotic, and ultimately frankly gangrenous and perforated. Intestinal ischemia and infarction from mesenteric venous occlusion are easily differentiated from those which are purely arterial in origin because the former results in a bowel wall that is intensely hemorrhagic and edematous, along with engorgement of the mesentery. In general, mesenteric venous thrombosis pursues a more virulent course and is less amenable to successful intervention than is mesenteric arterial occlusion.

Clinical Manifestations

Acute intestinal ischemia, regardless of cause, is characterized by pain out of proportion to physical findings. The abdominal pain is acute in onset, intense and diffuse, may be accompanied by vomiting, and is unresponsive to narcotic administration. Occult blood in the stool or frankly bloody diarrhea is a later finding, representing mucosal sloughing. Early abdominal examination is remarkable for a paucity or absence of findings. Localization of abdominal pain and development of peritoneal signs mark the onset of intestinal necrosis.

A history of atrial fibrillation or of a previous cardioarterial embolic event should raise suspicion of embolic occlusion of the superior mesenteric artery. Alternatively, a history of weight loss and food avoidance, along with the stigmata of diffuse atherosclerotic occlusive disease, is suggestive of an acute thrombotic occlusion of a previously stenotic superior mesenteric artery.

Diagnostic Studies

Laboratory investigation is too nonspecific to rule out the diagnosis of acute intestinal ischemia; the diagnosis must be made on clinical grounds and requires a high index of suspicion. Although the leukocyte count may be normal early on, it often increases to greater than 20,000/mm³ as the ischemic process progresses. Hemoconcentration, manifest as a high hematocrit, may be present secondary to fluid accumulation in the extravascular compartment and vomiting. Metabolic acidosis, manifest as a base deficit, may be present, the degree of which is related to the extent and duration of the intestinal ischemic process. Elevation of the serum lactate level similarly reflects the degree and duration of intestinal hypoperfusion. Persistence of a base deficit despite adequate resuscitative measures is indicative of advanced ischemia and most likely intestinal infarction. Elevation of the serum amylase level is a relatively late finding and is often indicative of bowel necrosis.

Plain film findings occur late, if at all, in the course of acute intestinal ischemia and cannot be relied on for diagnosis, although they may be useful in excluding other conditions. A notable exception is that of thumbprinting of the bowel wall because of submucosal edema, which is indicative of intestinal ischemia. More commonly, thumbprinting is seen on double-contrast barium enema examination in the setting of colon ischemia. Plain films of the abdomen in the setting of acute intestinal ischemia reveal only nonspecific air-filled loops of small intestine. Air in the mesenteric or portal venous system is an unusual but ominous diagnostic finding that occurs late in the disease process and signals bowel necrosis.

Ultrasound, *computed transaxial tomographic (CTT)* scanning, and MRI in the diagnosis of acute mesenteric ischemia are applicable. These modalities may be time-consuming, require radiologic expertise to interpret, offer no possibility of therapeutic intervention, and their diagnostic accuracy at the present time must be considered unproved. Contrast arteriography, including lateral aortography and selective injection of the superior mesenteric artery, remains the single most important diagnostic maneuver in evaluation of the patient with suspected acute mesenteric ischemia.

Prompt arteriography and catheter-based therapies form the basis of an aggressive approach toward the diagnosis and treatment of acute mesenteric ischemia, popularized by Boley and colleagues in the early 1980s, and represent a major advance in the treatment of this highly lethal condition. Arteriography almost always distinguishes between embolic and atherosclerotic occlusive causes of acute mesenteric ischemia. This is a critical determination because the operative approach to mesenteric revascularization is vastly different in each instance. Arteriography demonstrates instances of nonocclusive mesenteric ischemia, which often can be managed successfully nonoperatively with catheter-based pharmacologic therapies.

In the case of cardioarterial embolus, arteriography reveals the level of embolic occlusion, which is typically at the origin of a branch vessel, such as the middle colic or right colic artery, and directs the operative approach to embolectomy. In the case of atherosclerotic occlusion, arteriography reveals the site and length of the occlusion, typically at the origin of the superior mesenteric or celiac artery, and allows planning of operation type, e.g., endarterectomy or bypass. Critical information also is provided regarding suitability of inflow source (usually the supraceliac aorta).

Treatment

Once the diagnosis of acute occlusive mesenteric ischemia has been established at arteriography, continuous infusion of a vasodilator, such as papaverine or nitroglycerine, may be begun directly into the superior mesenteric artery if the origin of this vessel is patent and can be cannulated (Fig. [33-12](#)). This maneuver can be especially useful if an embolus has lodged at the takeoff of the middle colic or right colic artery because it may break the accompanying vasospasm occurring outside the area of ischemia. The catheter may be left in place for ongoing infusion of vasodilator therapy during and after relief of arterial obstruction by surgical embolectomy. When acute mesenteric arterial occlusion has been identified and vasodilator therapy begun, time is of the essence in proceeding to the operating room. Resuscitation of the patient with an intravenous balanced salt solution, nasogastric decompression, establishment of invasive hemodynamic monitoring, optimization of cardiac performance with pharmacologic intervention as needed, and administration of broad-spectrum antibiotics aimed at enteric organisms are supportive aspects of care. These should be instituted on initial assessment of the patient, performed in an ongoing fashion during work-up, and continued in the perioperative period.

At operation, through a long midline incision, initial decision regarding treatment is based on the extent and severity of intestinal ischemia. Methods of surgical revascularization will be vastly different depending on the cause of the ischemic insult. With embolic or atherosclerotic occlusion, only intestine that is frankly necrotic should be resected initially, followed by an attempt at revascularization. If only a short segment of intestine appears ischemic from an embolus to a distal branch of the ileocolic artery, resection of the affected intestine with primary anastomosis and anticoagulation may be sufficient.

An embolus lodged at the origin of the superior mesenteric artery produces patchy cyanosis of the intestine from the level of the ligament of Treitz to the mid-transverse colon, with many intervening areas of questionable viability. Revascularization of the ischemic intestine by embolectomy of the superior mesenteric artery should be performed. The superior mesenteric artery can be isolated in its midportion as it crosses in front of the fourth part of the duodenum, after the transverse colon has been reflected cephalad and the ligament of Treitz taken down. After anticoagulation with heparin and control of the superior mesenteric artery with atraumatic vascular clamps or doubly looped vessel tapes, a transverse arteriotomy is made, and a small Fogarty balloon embolectomy catheter is passed proximally. The balloon is gently inflated and the embolus extracted. After pulsatile inflow is obtained, an additional pass of the catheter into the aorta is made and any residual thrombus withdrawn. If adequate backbleeding is present and no thrombus is visualized, the Fogarty catheter need not be passed distally, and the arteriotomy can be closed primarily. If thrombus or embolus is encountered at the arteriotomy site, it may be gently teased out, followed by passage of a Fogarty catheter if adequate inflow or backbleeding cannot be readily established. Once embolectomy has been achieved and intestinal circulation reestablished, intestinal viability should be reassessed. Only those areas of intestine which are frankly necrotic should be resected at this time. The routine use of fluorescein injection for intraoperative assessment of intestinal viability cannot be recommended. Intestine that is of questionable viability should be left in situ and reassessed at a planned “second look” laparotomy 24 to 48 h later. The decision to proceed with “second look” laparotomy is made at the time of the initial operation and is not based on the subsequent postoperative course.

At the “second look” laparotomy after successful mesenteric revascularization, it is often found that bowel that appeared to be of questionable viability initially is now clearly viable. This avoids the consequences of massive intestinal resection, such as short-gut syndrome and lifelong parenteral nutrition. All necrotic bowel must be resected at this time, and depending on the extent of intestinal loss, total parenteral nutrition is frequently an important adjunct in the postoperative period.

Necrosis of the midgut from the level of the ligament of Treitz to the mid-transverse colon might be encountered at the time of initial laparotomy. Consequently, there is no point in attempting revascularization because all the end-organ is nonviable. The extent of intestinal resection required is often incompatible with life in the elderly patient, with or without total parenteral nutrition, and the surgeon has a difficult ethical decision. In this circumstance, the most humane course may be to close the abdomen and provide comfort care until death from sepsis inevitably occurs.

In the case of acute mesenteric ischemia resulting from atherosclerotic occlusion of the superior mesenteric artery, with or without associated occlusion or stenosis of the celiac artery, after resection of frankly necrotic bowel, intraoperative efforts are focused on complete visceral revascularization, including revascularization of the celiac artery, if necessary, in addition to the superior mesenteric artery. Techniques of visceral revascularization include transaortic endarterectomy of the origins of the celiac and superior mesenteric arteries via a “trapdoor” aortotomy or bypass. Methods of retrograde bypass from the infrarenal aorta or iliac arteries have been described, but none has proved as durable as antegrade bypass from the supraceliac aorta using prosthetic graft material. Although technically more demanding, transaortic endarterectomy performed via medial visceral rotation is advisable for intestinal necrosis because the need for prosthetic graft material is obviated. Antegrade bypass and transaortic endarterectomy have proved durable, providing symptomatic relief in approximately 95 percent of patients at 1 year and 85 percent at 5 years. After successful revascularization, “second look” laparotomy may be necessary.

The mortality rate after sudden onset of mesenteric ischemia is as high as 85 percent. This can be attributed to delay in diagnosis and the reluctance of many physicians unfamiliar with this entity to proceed with arteriography in a critically ill patient. Using the multimodal approach, Boley and associates reported a decrease in mortality rate to 45 percent. Mortality is higher after acute occlusion by thrombosis compared with embolism.

Reoxygenation of ischemic tissues exacerbates preexisting ischemic injury, especially in an organ with a relatively high cell turnover rate, such as small intestine. This reperfusion injury is mediated by highly reactive free-radical metabolites of oxygen and is characterized by cell membrane injury, phospholipid peroxidation, cell swelling, and in many instances cell death. Bacterial translocation from the intestinal lumen into the circulation has been documented. Injury occurring in organ systems remote from the focus of ischemia-reperfusion, most notably the lung, also has been noted. Mannitol, a tertiary alcohol and scavenger of the deleterious hydroxyl radical, has some efficacy and may be administered in a dose of 25 g before reestablishment of perfusion to ischemic intestine.

Nonocclusive Visceral Ischemia

In up to one-third of patients presenting with the manifestations of acute mesenteric ischemia, lateral aortography demonstrates patency of the main superior mesenteric artery. Selective injection of the superior mesenteric artery reveals multiple segmental areas of narrowing of small and intermediate branch vessels and diminution or absence of a mural intestinal circulation (Fig. [33-13](#)). Nonocclusive mesenteric ischemia occurs in the patient experiencing a prolonged low-flow state, such as that accompanying congestive heart failure complicating acute myocardial infarction or on a chronic basis. The patient usually is hypovolemic from aggressive diuretic therapy aimed at relieving cardiac failure and is receiving cardiac inotropic support and α -adrenergic agents to maintain blood pressure.

Splanchnic vasoconstriction may occur initially on a compensatory basis with profoundly depressed cardiac output. Persistent and intractable splanchnic vasoconstriction results in severe intestinal ischemia. The distribution of ischemia is diffuse from the level of the ligament of Treitz to the mid-transverse colon, and the process often has progressed to bowel necrosis at diagnosis. Nonocclusive mesenteric ischemia can represent the terminal event in a patient with a persistent low-flow state and multiple organ system failure.

Clinical Manifestations

The clinical picture may be identical to that of patients with acute arterial or venous mesenteric occlusions. The patients, however, usually are older, and the infarction develops

slowly over a period of several days, during which time there may be prodromal symptoms of malaise and vague abdominal discomfort. Associated congestive heart failure, with or without arrhythmia, is frequent. An unusually large number of patients have been found to be overdigitalized. Infarction of the intestine is preceded by the sudden onset of severe abdominal pain and vomiting. The patient usually becomes acutely hypotensive and develops a rapid pulse. Watery diarrhea is frequent, and the stools may be grossly bloody. The abdomen becomes diffusely tender and rigid. Bowel sounds are diminished and later absent. Fever and leukocytosis are usual, and frequently there is a thrombocytopenia related to intravascular thrombosis. A characteristic early laboratory finding is a markedly elevated hematocrit, which occurs because of the “trapping” of serum in the bowel wall and seepage into the peritoneal cavity. Metabolic acidosis with a significant base deficit is present almost universally.

In disorders in which the splanchnic circulation is thought to be diminished, unexplained abdominal signs and symptoms should be viewed with the possibility of mesenteric vascular insufficiency and intestinal necrosis in mind.

Contrast arteriography, including lateral aortography and selective injection of the superior mesenteric artery, is the most important diagnostic procedure for evaluating patients believed to have acute mesenteric ischemia. In addition, the ability to administer catheter-based pharmacologic interventions is available immediately. Optimization of cardiac function, discontinuation of splanchnic vasoconstricting agents, and infusion of vasodilators such as papaverine or nitroglycerin directly into the superior mesenteric artery are the recommended therapy. Such infusions can be maintained safely for several days. Serial arteriography over a period of 48 to 72 h might be necessary in order to monitor therapeutic response. Administration of broad-spectrum antibiotics aimed at enteric organisms can diminish bacterial translocation from the gut to the bloodstream. Once peritoneal signs have developed, exploration is necessary.

At initial exploration, only that intestine which is necrotic should be resected, with anticipation of performing a “second look” laparotomy over a period of 24 to 48 h, during which time ongoing vasodilator therapy is administered directly into the superior mesenteric artery. The mortality from nonocclusive mesenteric artery ischemia is in the range of 80 percent in most series, resulting from the presence of bowel necrosis at the time of diagnosis and from the high association of multiorgan dysfunction and preterminal illness in the target population. Improved survivorship has been reported by identifying groups of patients at high risk for this problem, using a policy of aggressive and multiple reoperations, and delaying intestinal anastomosis.

Chronic Visceral Ischemia

Etiology and Pathology

Atherosclerosis of a visceral artery with plaque overhanging or encroaching on the origins of the celiac and superior mesenteric arteries is the cause of chronic visceral ischemia in more than 95 percent of patients. The atheroma may extend into the celiac artery over a length of approximately 1 cm, but it is common for the atherosclerotic process in the superior mesenteric artery to extend distally over a length of 4 to 5 cm. Two of the three main visceral arteries (celiac, superior mesenteric, and inferior mesenteric) must be stenotic or occluded to produce significant gut ischemia because of the rich collateral network.

Other causes of stenosis and occlusion of the visceral arteries include Takayasu's arteritis, periarteritis nodosa, thromboangiitis obliterans, and fibromuscular dysplasia. Obstructing aortic atheroma in the suprarenal segment, termed “*coral reef*” *atheroma*, can produce visceral ischemia without narrowing the origin of the visceral vessels by restricting aortic flow or embolization. A proposed nonatheromatous cause of chronic visceral ischemic pain is extrinsic compression of the celiac artery by diaphragmatic fibers, known as *median arcuate ligament syndrome*. Extrinsic compression of the celiac artery, however, can be demonstrated by lateral aortography in many patients without symptoms. The existence of this syndrome has been questioned by many and may be insignificant.

Clinical Manifestations

Chronic visceral ischemia is more common in women, with a female-to-male ratio as high as 4:1 in many series. Onset of symptoms usually is in the sixth or seventh decade, and the usual risk factors and stigmata of atherosclerosis frequently are present. The pain of chronic visceral ischemia is dull and persistent, occurring most commonly in the epigastric and periumbilical areas. Often referred to as *intestinal angina*, the pain usually occurs approximately 20 min after eating and may persist for several hours. The pain is exacerbated by ingesting large meals and eating solid foods as opposed to liquids. If the pain is severe, nausea and vomiting occur. Initially the patient complains of constipation and later of diarrhea. There is usually steady progression in the frequency and duration of symptoms. The food-pain relationship leads to a reluctance on the part of the patient to eat. The subsequent rapid and severe weight loss characterizes the syndrome. As the intestinal ischemia progresses, a form of malabsorption syndrome occurs that contributes to the weight loss and is manifest by bulky, foamy stools high in fat and protein content. Symptoms of intestinal angina can exist for months or years before the visceral circulation becomes critically curtailed. As the mesenteric occlusive process progresses, patients typically avoid solid foods (“food fear”) and drink small amounts of liquid many times per day. This often leads to weight loss and malnutrition, which is contributed to by the decreased nutrient absorptive capacity of the chronically ischemic intestine. Although chronic visceral ischemia produces a well-defined clinical syndrome, it is not always promptly recognized by the clinician, and up to one-third of patients presenting with intestinal infarction have prior symptoms of chronic visceral ischemia that went unrecognized.

Examination usually reveals a thin patient. Auscultation of the abdomen often reveals a bruit in the epigastrium. Stigmata of atherosclerosis are common but not always present. Chronic visceral ischemia may be the initial and only manifestation of atherosclerosis, especially in women under 60 years of age.

Diagnostic Studies

Blood work is not helpful in the diagnosis of chronic visceral ischemia. Duplex ultrasound, combining B-mode ultrasound imaging with spectral analysis of Doppler signals, is useful for identifying high-grade stenoses of the celiac and superior mesenteric arteries. The detection of a stenotic lesion is based on acceleration of the peak systolic velocity of flowing blood across the area of stenosis. In addition to evaluation in the fasting state, repeat evaluation in the postprandial state may increase diagnostic accuracy. Because these patients are generally thin, it usually is possible to identify the visceral vessels with B-mode ultrasound.

Contrast arteriography is the definitive method for the diagnosis of chronic visceral ischemia and is the only imaging modality that provides the detailed information necessary to plant treatment. The typical atherosclerotic process is shown in Figs. [33-14](#) and [33-15](#). A markedly dilated and elongated inferior mesenteric artery filling the superior mesenteric artery through collaterals is indicative of a superior mesenteric artery stenosis or occlusion (Fig. [33-16](#)). When this finding is noted on a flush aortogram taken in the anteroposterior projection, lateral aortography of the visceral bearing segment of the aorta should be obtained.

When obtaining contrast arteriography, the administration of a dye load to a dehydrated patient with critical stenosis of the visceral arteries can be disastrous because it can precipitate thrombotic occlusion at the site of stenosis with resulting intestinal infarction. Attention to volume status before arteriography is essential, as is hydration before and after arteriography.

The arteriographic appearance of median arcuate ligament syndrome is that of extrinsic compression of the celiac artery along its superior border with caudal displacement of the artery. This compression occurs approximately 1 cm beyond the vessel origin, and there is no evidence of atherosclerosis. Poststenotic dilatation of the artery may be present.

Reports of other imaging modalities for the detection of visceral artery stenosis include MRI and MR angiography, conventional CTT scanning and CTT angiography, and various radioisotope techniques.

Treatment

Once the diagnosis of chronic visceral ischemia has been established, hemodynamic status should be optimized. The cardiac function should be closely monitored intraoperatively via flow-directed pulmonary artery catheter or transesophageal echocardiography in anticipation of supraceliac aortic clamping. Afterload reduction to relieve left ventricular wall stress is required during the period of supraceliac aortic occlusion. The goal of the operation is to provide complete revascularization of the ischemic gut. The two principal approaches to mesenteric revascularization are antegrade aortovisceral bypass and transaortic visceral endarterectomy. Both procedures have proved durable, providing symptomatic relief in approximately 95 percent of patients at 1 year and 85 percent at 5 years.

Antegrade bypass from the supraceliac aorta to the celiac or superior mesenteric artery is technically less demanding than transaortic endarterectomy. The supraceliac aorta usually is free from significant atherosclerotic occlusive disease. The gastrohepatic omentum is divided and the stomach and pancreas retracted inferiorly. The left lobe of the liver is mobilized and retracted to the right. The nasogastric tube is palpated easily within the esophagus, which is retracted to the left. The diaphragmatic crura and celiac ganglionic tissue are divided, exposing the supraceliac aorta. Prosthetic graft material, Dacron or expanded polytetrafluoroethylene, is the conduit of choice. The short straight segment of a bifurcated graft will be sewn in end-to-side fashion to the supraceliac aorta, with the limbs of the graft sewn in end-to-end fashion to the celiac and superior mesenteric arteries. The graft limbs should be in the anatomic position, with the limb to the superior mesenteric artery placed in a retropancreatic tunnel (Fig. [33-17](#)). In this way, the prosthesis will resist kinking regardless of the position of the viscera. If only one vessel is to be revascularized, a flanged graft can be used, with attachment of the flanged portion of the graft to the supraceliac aorta.

The anatomic exposure necessary to perform transaortic visceral endarterectomy can be achieved via long midline incision and medial visceral rotation. Medial visceral rotation is accomplished by reflecting the abdominal viscera from left to right, including the stomach, spleen, and pancreas. Only the left kidney is left in anatomic position (Fig. [33-18](#)). After supraceliac aortic clamping, a “trapdoor” aortotomy is created and endarterectomy of the paravisceral aorta and visceral (and, if necessary, renal) artery origins performed (Fig. [33-19](#)). Closure of the aortotomy provides restoration of aortic continuity. Intraoperative duplex scanning provides an excellent method of assessment of technical result.

Mesenteric revascularization using reversed saphenous vein grafts originating from the iliac arteries or infrarenal aorta has been described. Such grafts placed in retrograde fashion are prone to kinking, and long-term durability is not comparable with that of antegrade prosthetic aortovisceral bypass or transaortic visceral endarterectomy.

There have been isolated reports of percutaneous transluminal angioplasty of the visceral arteries in the treatment of chronic visceral ischemia. The number of patients treated has been small, and in almost 20 percent of patients, the planned procedure has proved technically impossible because of anatomic considerations peculiar to the visceral arteries. Early follow-up reveals a failure rate of up to 30 percent, necessitating reintervention. Limited applicability and high early failure rate render this a compromise procedure at best, which should be used for patients with prohibitive operative risks.

If operation for median arcuate ligament syndrome is undertaken, narrowing of the celiac artery can be relieved by transecting the constricting fibers of the median arcuate ligament. Concomitant arterial reconstruction may be necessary if the stenosis is persistent or if poststenotic dilatation is present.

Colonic Ischemia

Etiology and Pathology

Colonic ischemia is usually manifest in the elderly population. Causes include atherosclerotic stenosis or occlusion of the inferior mesenteric artery, cholesterol emboli, and the inflammatory arteriopathies. In most instances, no identifiable cause can be found. Iatrogenic causes include operative ligation occurring during extracolonic operative procedures, such as abdominal aortic aneurysm repair. In the setting of chronic occlusion of the inferior mesenteric artery, perfusion of the descending and sigmoid colon is through the collateral circulation emanating from the superior mesenteric artery via the middle colic artery proximally and the internal iliac arteries via the inferior hemorrhoidal vessels distally.

Clinical Manifestations

The presentation is nonspecific, characterized by cramping abdominal pain localized to the left lower quadrant. Compared with the pain of acute ischemia of the small intestine, it is mild in nature. The urge to defecate may be present, followed over a period of 24 h by melena or hematochezia. Physical examination usually reveals tenderness over the area of the involved segment of colon. The descending and sigmoid colon is the most common site of involvement, but any part of the colon can be affected. Colonic ischemia may be reversible, with results ranging from complete healing to stricture formation with obstruction and colonic infarction and perforation. The outcome of colonic ischemia cannot be predicted from the presenting symptoms unless colonic infarction is obviously present.

Diagnostic Studies

Mild elevation of the white blood cell count is present, but the results of blood work are generally nonspecific. Plain film examination of the abdomen usually is not helpful, although it may demonstrate thumbprinting in up to 20 percent of patients. Thumbprinting consists of multiple round, smooth, soft tissue densities projecting into the air-filled colonic lumen, representative of submucosal edema or hemorrhage. Double-contrast barium enema is the most useful study in detecting colonic ischemia but should not be performed in any patient with signs of peritonitis or when intestinal infarction or perforation is suspected. Findings on barium enema also include transverse ridging and ulceration. CTT imaging may demonstrate thickening of the colonic wall. Flexible sigmoidoscopic or colonoscopic examination often is diagnostic and should be performed with caution in selected patients to avoid converting an ischemic but intact colon to a perforated colon as a result of insufflation and manipulation. Whatever diagnostic modality is used, sequential examination may be required to confirm the diagnosis. Contrast arteriography has little value in the diagnosis of colonic ischemia.

Treatment

When the diagnosis of colonic ischemia is suspected or established, initial treatment consists of parenteral fluid administration with expansion of circulation volume, bowel rest, and administration of broad-spectrum antibiotics aimed at coverage of enteric organisms, including anaerobic bacteria. According to Boley and colleagues, colonic ischemia is reversible in approximately 50 percent of patients, requiring only supportive measures. Approximately 10 percent develop stricture formation, 20 percent persistent colitis, and 20 percent colonic infarction and perforation. Serial abdominal examination and monitoring of urine output and white blood cell count are important in identifying those patients progressing to colonic infarction. Segmental resection of the involved area may be sufficient. Because the bowel is not prepared before operation, resection with end- colostomy and mucous fistula or Hartmann's procedure should be performed. No attempt should be made to prepare the bowel because preparation might precipitate perforation. Late manifestations of colonic ischemia include stricture and symptoms of chronic colitis, either of which can necessitate elective resection.

Mesenteric Venous Occlusion

Etiology

Thrombosis of the mesenteric venous circulation occurs less frequently than that of the mesenteric arterial circulation, accounting for less than 20 percent of all cases of acute mesenteric ischemia. Predisposing conditions include previous deep venous thrombosis, intraabdominal infectious or inflammatory processes, portal hypertension, malignancy, smoking, oral contraception, and the postoperative state. Hypercoagulability because of congenital absence of naturally occurring antithrombotic proteins also has been recognized as a predisposing factor and explains the pathogenesis of cases once thought to be idiopathic. Inherited deficiency or absence of protein S, protein C, and antithrombin III is transmitted in an autosomal dominant pattern and results in recurrent deep venous thrombosis if left untreated. Other hypercoagulable disorders associated with mesenteric venous thrombosis include polycythemia vera, thrombocytosis, hyperfibrinogenemia, and myeloproliferative disorders. Up to 50 percent of patients were noted to have previous episodes of mesenteric venous thrombosis.

Pathology

The jejunum and ileum are involved in greater than 80 percent of the commonly affected sites. Colonic and duodenal involvement are unusual, occurring in only 10 percent of patients. A similarly low incidence of duodenal involvement has been noted. Sudden occlusion of the main stem superior mesenteric vein in the dog leads to rapid sequestration of splanchnic venous circulation, stagnation shock, and hemorrhagic infarction of the bowel progressing to necrosis and gangrene. In human beings, ligation of the portal vein or superior mesenteric vein does not produce infarction unless secondary thrombosis extends to the bowel wall and involves the venous arcades and vasa recta. Primary thrombotic occlusion of the visceral veins usually begins in the smaller tributaries. Depending on the extent and location of the propagating clot, the bowel lesion may be represented by small localized areas or extensive segments of infarction. With extensive venous occlusion, thrombosis of the arterial side of the splanchnic circulation often follows, so it becomes impossible to determine accurately whether the occlusion was initially arterial or venous.

Phlebitis secondary to inflammatory disease of the bowel may extend to involve the entire portal system (pyelophlebitis) or give rise to septic emboli that lodge within the liver, causing intrahepatic abscess. This complication of mesenteric venous thrombosis has become less frequent with the advent of antibiotics, and the bowel symptoms usually are overshadowed by those resulting from infection.

Mesenteric venous thrombosis usually results in a segmental and more limited pattern of intestinal ischemia. At operation, the typical appearance is that of a massively edematous bowel wall with bluish discoloration and intramural hemorrhage. The mesentery is edematous, and arterial pulsations are present. Thrombosis of the small-caliber veins adjacent to the bowel wall may be visualized.

Clinical Manifestations

An indication of acute mesenteric venous thrombosis is abdominal pain disproportionate to physical findings. The abdominal pain may be colicky in nature, poorly localized, and accompanied by anorexia and diarrhea, with occult blood or frank blood in the stool. The symptoms accompanying mesenteric venous thrombosis can be insidious and may be present on a subacute basis for a period of 1 to 4 weeks before recognition of the diagnosis or before presentation as an intraabdominal catastrophe. A chronic form of mesenteric venous thrombosis is characterized by persistence of mild symptoms without progression to bowel infarction. The development of persistent, severe abdominal pain with peritoneal signs is indicative of intestinal infarction, which has a mortality rate in excess of 50 percent.

Narcotics usually do not relieve the pain. Bloody diarrhea is more frequent than with arterial occlusions. The bowel sounds are hypoactive or absent. Generalized abdominal tenderness, guarding, and distention are usual; however, true rigidity is not present unless gangrene and perforation of the bowel have occurred. A marked leukocytosis and elevated hematocrit are characteristic findings in venous thrombosis. The latter reflects the trapping of plasma in the occluded bowel segment as arterial blood continues to flow into a splanchnic bed without adequate venous drainage. As with acute mesenteric arterial occlusion, elevations of semen concentrations of phosphate, amylase, and creatinine phosphokinase may be seen. Metabolic acidosis with a base deficit occurs frequently. Abdominal paracentesis yields serosanguineous fluid, which, if foul smelling, makes immediate laparotomy mandatory.

Plain films of the abdomen are generally not helpful. Contrast-enhanced CTT scanning is the most useful diagnostic modality, often demonstrating thrombus in the superior mesenteric vein. Thickening of the bowel wall, as well as air in the bowel wall or portal vein, may be noted as well. Arteriography is not necessary to establish the diagnosis. However, in instances where arteriography is obtained in the acute setting to rule out an arterial etiology, filling defects or nonfilling of the superior mesenteric vein is typically noted and is diagnostic.

Treatment

The cornerstone of therapy rests on early recognition of the disease process and immediate anticoagulation with intravenous heparin. Improved survival correlates with early administration of heparin in several series. In the absence of peritoneal signs, intravenous heparin and supportive therapy may be sufficient. Lifelong anticoagulation with coumadin should be instituted in order to decrease the likelihood of recurrent thrombosis, which is significant even with ongoing anticoagulation. Investigation of the patient and family for hypercoagulability should be carried out.

Those patients in whom peritoneal signs have developed should undergo prompt exploration with resection of infarcted bowel. Intravenous heparin therapy should be started when the diagnosis is established, either preoperatively or intraoperatively. "Second look" celiotomy may be necessary, the need for which should be established at the conclusion of the initial operation. Even with improved methods of diagnosis, reported mortality rates are approximately 30 percent.

The majority of patients with acute mesenteric venous thrombosis have a diffuse thrombotic process with distal extension of clot into the smaller veins, precluding successful venous thrombectomy. Thrombolysis with urokinase as adjunct therapy also has been reported. After successful anticoagulant treatment of mesenteric venous thrombosis, recanalization of the superior mesenteric vein has been demonstrated.

Aneurysms

Visceral artery aneurysms are uncommon. The splenic artery is the most common site, accounting for 60 percent of such aneurysms. The hepatic artery is the second most common site (20 percent), and the remaining 20 percent are aneurysms of the superior mesenteric, celiac, gastroduodenal, left gastric, pancreaticoduodenal, jejunoileal, and inferior mesenteric arteries.

Etiology and Pathology

Visceral artery aneurysms most commonly occur spontaneously; the cause is unknown. Though found in conjunction with atherosclerosis, the atherosclerotic process is not believed to be causative. Recognized causes include fibromuscular dysplasia, connective tissue disorders, infection, pancreatitis, and trauma.

Splenic artery aneurysms are unique, occurring more commonly in women with a ratio of 4:1. Age distribution is nearly equally divided between women of childbearing age in their third decade and women in their sixth or seventh decade of life. Men with hepatic artery aneurysms outnumber women 9:1. Visceral artery aneurysms typically occur in isolation, but they may be associated with other aneurysms of the visceral, aortic, and peripheral arteries.

A relatively large number of aneurysms of the superior mesenteric artery result from mycotic involvement or necrotizing agents. Catheter sepsis, particularly from total parenteral nutrition catheters, may lead to superior mesenteric artery aneurysm when the infecting organism is *Candida* spp.

The size of splanchnic artery aneurysms varies. The majority are less than 2 cm in diameter, but aneurysms of the splenic artery measuring up to 10 cm have been reported.

Aneurysms of the superior mesenteric artery and its branches are typically very small, and when rupture occurs, frequently no site of bleeding can be found at laparotomy or at autopsy despite the loss of several liters of blood. In contrast to aneurysms of the abdominal aorta, size does not predict the risk of rupture. It has been estimated that only about 2 to 10 percent of splenic artery aneurysms rupture, whereas the risk of rupture of hepatic, celiac, and superior mesenteric artery aneurysms is high, approximating 30 percent of patients.

Clinical Manifestations and Diagnostic Studies

The majority of visceral artery aneurysms are asymptomatic. Splenic artery aneurysms are most commonly detected on plain film of the abdomen, where concentric calcification of the aneurysm is present in up to 80 percent of patients. Splenic artery aneurysms may present with left upper quadrant pain and pain referred to the region of the left shoulder. Hepatic artery aneurysms may present with right upper quadrant pain, hemobilia, and jaundice. The initial presentation of visceral artery aneurysm in 25 to 40 percent of patients is that of free intraperitoneal rupture with hypotension. The “double rupture” phenomenon has been reported to occur in up to 20 percent of patients with ruptured splenic artery aneurysm. As the initial rupture is contained within the lesser sac, hemodynamic stability is temporarily maintained, only to be followed by free intraperitoneal rupture precipitating hemorrhagic shock.

The incidence of rupture of visceral artery aneurysms is difficult to determine. The rupture rate of hepatic artery aneurysm is estimated at 20 percent and may occur with bleeding into the biliary tract, gastrointestinal tract, or freely into the peritoneal cavity. Mortality rates from rupture also are difficult to determine because many patients bleed to death before diagnosis. Mortality rates from rupture of visceral artery aneurysm are between 35 to 100 percent. Splenic artery aneurysm represents a unique situation because the hormonal and hyperdynamic circulatory changes associated with pregnancy predispose aneurysms of even small diameter to rupture. In the elective setting, selective visceral arteriography should be performed.

Treatment

Resection of the aneurysm, with or without arterial reconstruction, is the accepted treatment. Experience has accumulated with the technique of percutaneous transcatheter embolization, and this has emerged as a viable treatment in selected patients. Indications for intervention in the elective setting are not precise. Since the risk of rupture of splenic artery aneurysm during pregnancy is increased, a splenic artery aneurysm of any size should be repaired before planned pregnancy. The mortality rate associated with rupture for mother and fetus is almost 70 percent. In women of childbearing age, a splenic artery aneurysm measuring 2 cm or greater in diameter or an aneurysm diameter twice that of the caliber of the proximal splenic artery is an indication for intervention. Splenic artery aneurysm occurring in the sixth or seventh decade of life usually should be treated. In the presence of significant comorbidity, splenic artery aneurysms of up to 3 cm can be monitored safely. The rupture rate of hepatic artery aneurysm is approximately 20 percent, with a mortality of at least 35 percent. Therefore, an aggressive approach to intervention is indicated. Indications for repair of hepatic and other visceral artery aneurysms, based on aneurysm diameter, are the same as those for splenic artery aneurysm.

Splenic artery aneurysm is treated by resection of the aneurysm. The spleen can be preserved after resection of an aneurysm of the proximal splenic artery, while aneurysm of the distal splenic artery often necessitates splenectomy. Restoration of splenic artery continuity in order to preserve splenic viability is not necessary because usually the spleen can be perfused adequately by the short gastric circulation. Laparoscopic ligation of the splenic artery for aneurysm also has been described. Percutaneous transcatheter embolization of the aneurysm or occlusion of the main splenic artery on either side of the aneurysm also may be successful. Splenic preservation may be complicated by splenic abscess formation after successful surgical or transcatheter therapy.

Transcatheter embolization is particularly useful for intrahepatic arterial aneurysms and for other visceral artery aneurysms of noncritical vessels that may be difficult to expose at operation. Aneurysm of the proximal hepatic artery (e.g., common hepatic artery) may be treated by simple resection or exclusion because collateral circulation to the liver should be sufficient. Aneurysms of the more distal hepatic artery (e.g., right or left) may necessitate arterial reconstruction with interpositional grafting. Treatment of aneurysms of the celiac and superior mesenteric arteries demands restoration of arterial continuity.

Patients who present with ruptured aneurysms and intraabdominal or gastrointestinal blood loss require immediate resuscitation with fluid and blood and surgical intervention as soon as their condition allows. The bleeding may be so brisk that it will be necessary to take the patient to the operating room while in shock. The chances of finding the open vessel are better if the blood pressure can be elevated. If time allows, preoperative localization by arteriography should be carrier out. When the patient's condition is critical and preoperative studies are not advisable, an operative aortogram may aid in localization of the site of bleeding if it is not obvious at the time of exploration. In the case of a ruptured splenic artery aneurysm, the urgency of the situation usually necessitates a splenectomy to facilitate exposure for ligation of the ruptured artery. Rupture of a celiac artery aneurysm can be treated by ligation, and although most hepatic artery aneurysms also can be treated safely by ligation, an attempt should be made to reconstruct the hepatic artery with a graft. When it is necessary to ligate the hepatic artery, protection should be afforded the liver with large doses of broad-spectrum antibiotics postoperatively. In patients who are not candidates for an operation, catheter embolization of the artery offers a reasonable alternative. Rupture of a main stem superior mesenteric artery aneurysm requires replacement with a graft, preferably of autogenous vein, after excision. Most ruptured aneurysms of the smaller visceral arteries can be treated by ligation. Adequate circulation to the bowel must be ensured and obviously ischemic intestine resected. If there is a question of viability of a long segment of intestine, a “second look” operation in 24 to 36 h may avoid an extensive resection.

The operative mortality for elective treatment of nonruptured splanchnic artery aneurysms is less than 5 percent. The mortality rate is considerably higher once the aneurysm has ruptured and is related to the artery involved. Overall mortality after rupture of a splenic artery aneurysm is about 25 percent, with a 65 percent maternal and 95 percent fetal mortality if rupture occurs during pregnancy. The mortality subsequent to rupture of a celiac or superior mesenteric artery aneurysm is between 40 and 60 percent, whereas rupture of a hepatic artery aneurysm is associated with a 70 percent mortality rate. Death is the result of massive blood loss complicated by the effects of systemic arteriosclerosis, which many patients also have.

Trauma

Etiology

Mesenteric arteries and veins may be injured by penetrating or nonpenetrating abdominal trauma or accidentally during abdominal operations. Often the vessels are lacerated; less frequently, contusion of an artery by blunt trauma results in thrombosis or subsequent aneurysmal formation with rupture.

Most penetrating injuries of the mesenteric vessels are from stab or gunshot wounds, and associated injury to other organs is frequent. Isolated injury to mesenteric vessels after blunt abdominal trauma is rare and usually involves vessels in the mesentery proper or porta hepatis.

Clinical Manifestations

Depending on the size of the vessel lacerated, the rapidity of the bleeding, and associated organ injury, the patient will present with varying degrees of shock, abdominal pain, tenderness, distention, and spasm. Pain referred to the left shoulder is a valuable diagnostic symptom. Plain films of the abdomen are not helpful unless associated visceral rupture has occurred, in which case free air may be seen.

Treatment

Gunshot wounds of the abdomen and most stab wounds should be explored early regardless of physical findings. In nonpenetrating injury, a paracentesis yielding nonclotting blood prompts early operative intervention. Treatment consists of controlling the bleeding vessel. Lacerations of the inferior mesenteric artery or smaller mesenteric arteries and most mesenteric veins usually can be treated successfully by ligation. In the past, few patients with injury to the main trunk of the superior mesenteric artery survived; most died from infarction of the intestine or complications arising from associated organ injury. Currently, if the patient can be resuscitated and reaches the operating room, arterial repair may be possible. This can be accomplished by primary suture or interposition of a vascular graft between the severed ends of the vessel. Every attempt should be made to repair lacerations of the hepatic artery or portal vein. Whether treatment is by ligation or reconstruction, adequate circulation to the intestine should be established and any obviously ischemic bowel resected before closing the abdomen. Long segments of small intestine that are of questionable viability are left in place and reexamined at a second operation 24 h later.

Nonspecific Mesenteric Lymphadenitis

Nonspecific mesenteric lymphadenitis is a common cause of acute abdominal pain in children and young adults. Its existence as a distinct clinicopathologic entity is now well accepted, although the condition received little attention in early medical texts following its initial description by Wilensky and Hahn in 1926. An extensive review of the disease has been reported by McDonald. Because the condition is self-limiting and can be accurately diagnosed only at celiotomy, its true incidence is unknown. It is the most common cause of inflammatory enlargement of abdominal lymph glands, surpassing that resulting from tuberculosis, with which it has been confused in the past. Consideration of the disease is important because of its clinical similarity to several abdominal conditions requiring operative intervention, notably acute appendicitis.

Pathology and Pathogenesis

The lymph nodes primarily involved in nonspecific mesenteric lymphadenitis are those which drain the ileocecal region. The stasis of intestinal contents in the terminal ileum favors absorption of toxic or bacterial products from the bowel lumen, agents that may have a bearing on the pathogenesis of the disease.

The nodes are enlarged, discrete, soft, and pink at first; later they may become firm and white. It is uncertain whether calcification ever occurs in nontuberculous adenitis, and suppuration is rare unless specific bacterial infection is present. Histologically, the involved nodes present a pattern of reactive hyperplasia similar to that found in inflammatory and allergic affections of lymph nodes in other parts of the body. The nodes, with rare exception, prove to be sterile on culture or on animal inoculation. *Yersinia pseudotuberculosis* or *enterocolitica*, two gram-negative coccoid or ovoid organisms, have been insolated from the stool in some cases.

Clinical Manifestations

The disease most commonly occurs in patients under 18 years of age. There is no sex predilection. The signs and symptoms are not particularly characteristic. Very often there has been a recent sore throat or upper respiratory tract infection. Pain usually is the first symptom and varies in intensity from an ache to a severe colic. The mechanism responsible for producing pain is not completely understood. Lymph nodes do not have sensory innervation, and therefore, enlargement of the node by itself should not produce pain. It is probable that the pain is referred from the mesentery, which has an abundance of sensory end organs that are stimulated when the mesentery is stretched during peristalsis. The initial pain usually is in the upper abdomen, but it also may begin in the lower right quadrant or be generalized. Eventually the pain localizes to the right side; an important point in differentiating the disease from acute appendicitis is that the patient is unable to indicate the exact site of the most intense pain. Between spasms of colic, the patient feels well and moves about without difficulty. Nausea and vomiting occur in about one- third of patients, while malaise and anorexia are inconstant symptoms.

The patient often appears flushed, and an associated rhinorrhea or acute pharyngitis is not unusual. Approximately 20 percent of patients have lymphadenopathy elsewhere, most

often in the cervical region. The usual finding on examination of the abdomen is tenderness in the lower aspect of the right side, which is somewhat higher and more medial and considerably less severe than in acute appendicitis. The point of maximal tenderness often varies from one examination to the next. An appreciable number of patients also have diffuse or periumbilical tenderness. Rebound tenderness may or may not be demonstrated. Voluntary guarding is sometimes present; however, true muscular rigidity is rare. Early in the attack, the temperature is moderately elevated, 38 to 38.5°C, and at least half the patients have leukocyte counts over 10,000/mm³.

Differential Diagnosis

The disease is most often confused with acute appendicitis but also must be differentiated from regional enteritis, intussusception, specific bacterial and granulomatous adenitis, and other forms of mesenteric glandular enlargement, such as infectious mononucleosis or lymphoma. The clinical similarity to acute appendicitis is such that in several large series, as many as 20 percent of patients undergoing appendectomy were found to have nonspecific mesenteric adenitis and a normal appendix. Important differentiating factors are the more localized and constant location of the pain and tenderness, the presence of muscle rigidity, and the frequent occurrence of nausea and vomiting in children with appendicitis.

Differentiation from *acute regional ileitis* is at times difficult. Mesenteric adenitis is an almost constant feature of regional ileitis, and indeed, the adenitis is often nonspecific in this disease. An ultrasound that demonstrates mural thickening of the terminal ileum and thickening of the mesentery is characteristic of regional enteritis.

The low incidence of lymphadenopathy in other parts of the body and the brief course of nonspecific mesenteric adenitis are factors in excluding lymphomas and infectious mononucleosis. A peripheral blood smear and a Paul-Bunnell test for sheep red cell agglutinins are also helpful.

Treatment

The prognosis of nonspecific mesenteric adenitis is excellent, and complete recovery from an individual attack can be expected without specific treatment. Death from the disease is extremely rare and occurs only when secondary specific bacterial infection, usually caused by hemolytic streptococci, causes suppuration of the nodes with rupture leading to abscess or peritonitis.

If the condition is mistaken for acute appendicitis, as it frequently is, laparotomy should be undertaken. It is safer to find a normal appendix than to run the risk of allowing acute appendicitis to go on to rupture. The diagnosis is readily established at the time of operation with the finding of enlarged mesenteric nodes in the absence of disease in the appendix or elsewhere in the intestinal tract or abdomen. In view of the tendency for recurrence and the difficulty of differentiating it from appendicitis, appendectomy should be performed.

Mesenteric Panniculitis

Mesenteric panniculitis is a term applied by Ogden and associates in 1965 to describe a process of extensive thickening of the mesentery by a nonspecific inflammatory process. It also has been variously designated *retractile mesenteritis*, *mesenteric lipodystrophy*, *lipogranuloma of the mesentery*, and *mesenteric manifestations of Weber-Christian disease*. Many consider it a variant of retroperitoneal fibrosis.

Etiology and Pathology

The cause of the condition is unknown; however, the process apparently results from an insult to the fatty tissue of the mesentery. Trauma, allergy, and subacute infection have all been implicated. The possibility that it is an autoimmune disease also has been suggested. The process usually involves the mesenteric root of the small bowel. Grossly, the normal fat lobulations of the markedly thickened and firm mesentery are lost. Scattered throughout are irregular areas of discoloration, which vary from reddish brown plaques to pale yellow foci resembling fat necrosis. The superior mesenteric vessels, though surrounded by the tumor-like mass of tissue, pass through it unaltered. Histologic sections show inflammatory involvement of the fibroadipose tissue with round cells, foam cells, and giant cells and various degrees of necrosis, fibrosis, and calcification.

Clinical Manifestations

Men are affected more often than women. It is rarely described in children, in whom mesenteric fat usually is scant. The clinical features are nonspecific; they include recurrent episodes of moderate to severe abdominal pain, nausea, vomiting, and malaise, usually in the right side of the abdomen. Radiographs are helpful only if the mass displaces or compresses viscera. CT scan demonstrates mesenteric panniculitis as a localized fat-density mass containing areas of increased density representing fibrosis.

Treatment

Laparotomy is necessary to establish the diagnosis and to rule out other intraabdominal diseases of the abdomen. The widespread involvement of the mesentery precludes doing more than obtaining a biopsy. Because neoplasms of the mesenteric lymph nodes may present a similar gross appearance, several biopsies from different sites should be obtained. Rarely, colostomy of bypass will be necessary to relieve symptoms of obstruction. Treatment of the disease with steroids, immunosuppressive agents, and irradiation has been suggested. The benefits from these are difficult to evaluate because the inflammatory process is self-limiting and seldom causes any serious complications.

Tumors of the Mesentery

Tumors originating between the leaves of the mesentery are rare. In contrast, malignant implants from intraabdominal or pelvic tumors or metastases to mesenteric lymph nodes are relatively common. Tumors arising from mesenteric lymph nodes occur; however, these are not generally included in a discussion of primary mesenteric tumors.

Pathology

Primary tumors of the mesentery may be cystic or solid. Of these, cystic growths occur more frequently than solid ones in a ratio of 2:1. A number of tissues, including lymphatic, vascular, nervous, and connective tissues, are the source of these tumors. In addition, cystic tumors may arise from embryonic rests (enteroceles or dermoids), from developmental defects (chylous or serous retention cysts), or after trauma (hemorrhagic cysts). A classification of these tumors is shown in Table [33-1](#).

The majority of cystic mesenteric tumors are benign. Rare exceptions are lymphangiosarcomas, which are true neoplasms arising from lymph channels, and malignant teratomas arising from multipotential embryonic rests. Chylous or lymphatic cysts are the most frequently encountered benign mesenteric masses. These are thought to arise from developmental defects in mesenteric lymphatics creating closed spaces within which fluid accumulates. They may be unilocular or multilocular, have an endothelial lining, contain a grossly cloudy fluid resembling chyle, and often grow to an extremely large size. A similar cause has been ascribed to serous cysts, which are differentiated from chylous cysts in that they contain clear fluid, are invariably unilocular, and may or may not have an endothelial lining. Lymphangioma of the mesentery is apparently a true neoplasm of lymphatics similar to those found in other parts of the body (cystic hygroma). Grossly and histologically, it is often difficult to differentiate this tumor from a chylous cyst, and in many series the two are grouped in a single category. Traumatic cysts follow external or surgical injury to the mesentery. They are lined with fibrous tissue and usually contain bloody fluid. It is probable that many serous retention cysts are in reality traumatic cysts that have evolved from disruption of mesenteric lymph channels. Enteric cysts are lined with intestinal mucosa and represent duplications of the intestinal tract that do not communicate with the bowel lumen. These and dermoid cysts of the mesentery are exceedingly rare.

Benign solid tumors of the mesentery are more common than malignant ones, and of these, lipomas and fibromas predominate. Recurrence after incomplete excision of histologically benign mesenteric tumors has been reported, and malignant degeneration also has been suggested. Histologically benign tumors can kill by local invasion with mechanical compression of adjacent viscera. The benign tumors of nerve elements and smooth muscles are uncommon. Vascular tumors of the mesentery are rare, and of these, hemangiopericytomas dominate. Ackerman states that liposarcoma is the most frequently encountered malignant tumor of this area, whereas Yannopoulos and Stout report leiomyosarcoma to be the most frequent. Few malignant mesenteric tumors have embolic metastases until very late. They spread by local extension or by peritoneal implants, which occur most often with leiomyosarcomas. As a rule, the malignant solid tumors arise near the foot of the mesentery, whereas solid benign tumors have a greater tendency to develop peripherally near the intestine.

Approximately two-thirds of mesenteric tumors, cystic or solid, are located in the mesentery of the small intestine, usually that of the ileum. Occasionally, they arise in the transverse or sigmoid mesocolon or in the gastrohepatic ligament. In the greater number of cases, the tumor is located peripherally in the mesentery, where it is often adherent to the adjacent intestine. The mobility of the mesentery permits benign and malignant tumors to grow to very large sizes before causing symptoms.

Clinical Manifestations

Early clinical features usually do not differentiate benign tumors from malignant ones. There is an equal sex incidence, although benign cystic tumors are somewhat more common in women and malignant tumors occur more frequently in men. These tumors have been described in children and also the very aged; however, the average age of patients with benign tumors is 45 years, while those with malignant tumors average 55 years old.

The manifestations of mesenteric tumors depend on the size, location, and mobility of the growth. In the majority of patients, symptoms are few or nonexistent, and the tumor is detected during a routine examination. Symptoms appear sooner when the tumor is situated in the periphery of the mesentery near the intestine than when it is located at its root. The patient may experience a sensation of fullness or pressure in the abdomen, particularly after eating. Less frequently, there are frank abdominal complaints. About one-half of patients with malignant tumors complain of abdominal pain, weakness, and weight loss, and one-third have diarrhea, cramps, anorexia, and nausea. Only rarely will the patient present with symptoms of complete intestinal obstruction or symptoms resulting from complications of the tumor per se, such as torsion, hemorrhage, or infarction of the tumor mass. In the absence of intestinal obstruction or these complications, the sole finding will be the presence of a nontender intraabdominal mass, usually in the lower right part of the abdomen. The mass varies in size from a few inches in diameter to one that may fill the entire abdomen. The extremely large masses usually are cystic, in which case they are tense and fluctuant. Cystic and solid tumors of the mesentery are mobile; they can be easily moved from side to side but only slightly in an upward and downward direction.

The differential diagnosis includes a variety of intraabdominal processes. Contrast-enhanced radiographs are helpful only when the mesenteric mass is sufficiently large to cause compression and displacement of the bowel or ureters (Fig. [33-20](#)). They do not differentiate a benign tumor from a malignant one, and in most instances the x-ray studies are not helpful. Calcification in the mass is suggestive of a dermoid or teratoma.

Imaging techniques are the most useful means for diagnosing cystic and solid mesenteric tumors. On ultrasound, a mesenteric cyst appears as a well-outlined, sonolucent transonic abdominal mass. CT demonstrates a simple mesenteric cyst as a nonenhancing near-water-density mass with a thin wall (Fig. [33-21](#)) and a solid tumor as an irregular sheetlike mass of greater density.

Treatment

Surgical excision is the only treatment for benign and malignant lesions. All mesenteric cysts of a size sufficient to be palpated should be removed if at all possible because even benign lesions eventually cause pain and compression of neighboring structures. Benign cystic tumors can be removed by enucleation or local excision, although resection in

continuity with the adjacent intestine is often necessary because of possible compromise to the vascularity of the bowel or difficulty in separating the tumor from the intestine. Wide excision together with resection of adjacent intestine is recommended for benign solid tumors, since these have a tendency toward local recurrence and malignant degeneration. Prognosis after adequate excision of cystic and solid benign tumors of the mesentery is excellent.

The outlook for malignant mesenteric tumors depends on whether complete removal is possible and is generally poor. Since malignant growths tend to occur in the root of the mesentery and often involve the great vessels and vasculature to most of the small intestine and colon, curative resection is often prohibitive. Resectable lesions invariably require removal of a portion of bowel; however, fewer than one-third of the malignant tumors are totally resectable. Since these growths may enlarge slowly and embolic metastases occur late, it is worthwhile to partially remove them to relieve obstructions, and prolonged survival has been recorded in a few instances. Irradiation therapy offers little, if any, benefit because the tumors are invariably radioresistant. Few patients with malignant primary mesenteric tumors are alive after 5 years. Death results from invasion with obstruction of the gastrointestinal tract leading to perforation and hemorrhage or from metastases to liver and lung.

RETROPERITONEUM

General Considerations

The retroperitoneum consists of that portion of the body which is bounded anteriorly by the peritoneum, posteriorly by the spine and psoas and quadratus lumborum muscles, superiorly by the twelfth ribs and attachments of the diaphragm, and inferiorly by the brim of the pelvis. The lateral margins of the space correspond to the lateral borders of the quadratus lumborum muscles. These limits define both an actual and a potential space, the actual space containing solid organs and major blood vessels and the potential space including soft tissues, nerve elements, and small blood vessels. Since there are no anatomic barriers in this area, pathologic processes may extend easily throughout it and are often bilateral.

Contained within the retroperitoneum are the kidneys, ureters, adrenal glands, portions of the autonomic and peripheral nervous systems, pancreas, abdominal aorta, inferior vena cava, spermatic or ovarian vessels, lymphatics and lymph nodes, and certain portions of the intestinal track, notably the duodenum. The space also contains fatty and areolar tissue and fibrous connective tissue.

The diagnosis of diseases involving the retroperitoneum has been enhanced by the application of such radiographic studies as pyelography, venography, arteriography, ultrasonography, and MRI. Nevertheless, it remains an obscure area of the body, enabling pathologic processes to become advanced before producing symptoms.

The multiplicity of structures within the retroperitoneum gives rise to a variety of pathologic conditions such as (1) idiopathic retroperitoneal fibrosis and (2) primary tumors of the retroperitoneal space.

Idiopathic Retroperitoneal Fibrosis

Idiopathic retroperitoneal fibrosis is a nonspecific, nonsuppurative inflammation of fibroadipose tissue of unknown cause that produces symptoms by the gradual compression of the tubular structures in the retroperitoneal space. It is believed that the disease represents one of the manifestations of a widespread entity termed *systemic idiopathic fibrosis*.

Idiopathic mediastinal fibrosis, Riedel's struma, sclerosing cholangitis, mesenteric panniculitis, Peyronie's disease, pseudotumor of the orbit, and perhaps desmoid tumor are other fibromatoses that are considered to be localized forms of systemic idiopathic fibrosis. A factor common to all these diseases is an inflammatory fibrotic process involving areolar and adipose tissue.

Retroperitoneal fibrosis was first described in 1905 by Albarran, a French urologist, who performed ureterolysis for ureteral compression produced by the disease. The first report in English is credited to Ormond in 1948. It has since been referred to as *Ormond's syndrome* but also has been labeled *idiopathic fibrous retroperitonitis*, *periureteritis plastica*, and *sclerosing retroperitonitis*. Retroperitoneal lipogranulomatosis (xanthogranulomatosis), which can produce a similar clinical picture, may be a granulomatous, prefibrotic stage of retroperitoneal fibrosis.

Etiology

Attesting to the obscure etiology of retroperitoneal fibrosis are the many theories that have been advanced to explain its origin. The possibility that extravasated urine might cause a fibrotic reaction in the retroperitoneum was mentioned by Ormond, and both he and Hackett also suggested that an abortive infection elsewhere might later start an inflammatory reaction in the lymphatic and perivascular tissues of the retroperitoneum. Hache and associates proposed the concept that the fibrosis was the end result of an ascending lymphangitis, adenitis, or periadenitis in the retroperitoneum with the infection arising from chronic or recurrent genitourinary infections or inflammatory diseases of the gastrointestinal tract or pelvic organs. Reports of retroperitoneal fibrosis occurring in patients taking the antiserotonin drug methysergide for headache with reversal of the fibrotic process after discontinuing the medication prompted the theory that the disease may be from a hypersensitivity reaction to the drug. Indirect evidence favoring an immune basis as the cause of the disease is confirmed by the impressive response of some patients to immunosuppressive therapy. It is probable that no single factor is responsible for causing the disease in all patients and that multiple factors may be implicated in any one patient.

Pathology

The gross appearance of retroperitoneal fibrosis usually is that of a plaque of woody, white fibrous tissue that is distributed along the course of the periaortic lymphatics. In about one-third of patients, it is bilateral. The diseased tissue, which may be 2 to 12 cm thick, extends from the sacral promontory to the renal pedicles and laterally to cover the iliopsoas muscle. It is sharply demarcated but not encapsulated. The mass surrounds and constricts but does not invade the regional structures in the retroperitoneum, primarily the blood vessels, nerves, and ureters, to which it becomes adherent.

A localized form has been observed as a circumscribed fibrous reaction surrounding only the ureters, and extensive involvement of the entire retroperitoneum with compression of the duodenum, common bile duct, and pancreas also has been seen. A similar fibrotic process has been described penetrating the diaphragm along the great vessels into the mediastinum, causing superior vena cava obstruction, and also extending into the root of the mesentery, causing intestinal obstruction.

Microscopically, the pattern varies from a subacute cellular process with polymorphonuclear cells, lymphocytes, fibroblasts, and fat cells to a completely hyalinized, relatively acellular sclerosis. Eosinophils, foreign-body giant cells, and small areas of calcification also may be present. Suppuration with abscess formation does not occur. The amount of fat and cellular infiltration and the degree of fibrosis vary between patients and in different biopsy specimens from the same patient. The more cellular picture usually is seen in the early stages of the disease, when there may be systemic signs of inflammation, whereas the dense fibrotic process is found late.

Clinical Manifestations

Retroperitoneal fibrosis is two to three times more common among men. It may occur in children and the aged; however, about two-thirds of the patients are between 40 to 60 years of age.

The protean manifestations of the disease are related to the phase and extent of the process and the structures secondarily involved. Ormond divided the natural history of the disease into three periods: (1) the period of incidence and development, (2) the period of activity, i.e., spread of the cellular and fibrotic process to envelopment of the retroperitoneal structures, and (3) the period of contraction of the fibrotic mass with compression of the involved structures. The disease is self-limiting once the fibrotic stage is reached, a factor of major importance in considering types of therapy.

Early symptoms are vague and nonspecific, but the first complaint is invariably pain. This is dull, noncolicky, and insidious in onset. It usually originates in the flank or low back and often radiates to the lower abdomen, groin, genitalia, or anteromedial aspect of the thigh. The pain is unilateral at first but may become bilateral as the fibrotic process spreads. Anorexia, nausea, diarrhea, generalized malaise, and weight loss variably occur in the early and late phases of the disease. Features of a subacute inflammation such as lower abdominal or costovertebral tenderness, moderate fever, and leukocytosis are often present early. Invariably, the erythrocyte sedimentation rate is elevated. A transabdominal or pelvic mass is palpable in about one-third of patients during some phase of the disease.

Symptoms resulting from compression of the tubular retroperitoneal structures may follow the initial complaints by 1 month to 2 years and reflect the late fibrotic phase of the disease with sclerotic contraction. The major structures involved are the ureters, aorta, and inferior vena cava. These all lie within the same fascial compartment. The aorta is resistant to compression and the inferior vena cava has abundant collaterals, so the symptoms are generally related to ureteral involvement. Partial or complete ureteral obstruction occurs in 75 to 85 percent of patients. The usual site of obstruction is in the lower third of the ureter. The ureteral obstruction usually is functional, rather than organic, as a consequence of cessation of peristalsis in the incarcerated ureteral segment. In the majority of patients, a ureteral catheter can be passed in a retrograde manner.

Dysuria, frequency of urination, and chills and fever occur with secondary infection of a hydronephrotic kidney. These symptoms may be intermittent for years, or a single attack may culminate in sudden anuria from bilateral obstruction. As many as 40 percent of patients have oliguria or anuria with laboratory evidence of azotemia. Clinically, the enlarged kidneys may be palpable, and the urine, if infected, contains white blood cells and bacteria. Hematuria, in the absence of infection, is rare.

Lower extremity edema, presumably from lymphatic and venous obstruction, occasionally occurs and may be unilateral. The level of obstruction in most cases corresponds to that of the ureteral obstruction and can be demonstrated on phlebograms.

Arterial insufficiency resulting from fibrous constriction of the aorta or iliac arteries is uncommon but can occur and may constitute the major problem. The intermittent claudication, rest pain, and limb ischemia are indistinguishable clinically and radiographically from those of atherosclerotic occlusion.

Rarely, the fibrotic process involves the retroperitoneal duodenum and common bile ducts, causing duodenal and biliary obstruction. Mechanical or functional intestinal obstruction resulting from extension into the root of the mesentery or sigmoid mesocolon also is a rare manifestation.

The diagnosis of retroperitoneal fibrosis usually is suggested by the contrast-enhanced radiographs of the urinary tract. Gray-scale ultrasonography and CT scans also have been used for the diagnosis of the entity and for follow-up management. A characteristic CT scan shows a homogeneous soft tissue mass enveloping the ureters, aorta, and inferior vena cava. In contrast to malignant retroperitoneal adenopathy, with which retroperitoneal fibrosis is often confused, there is no anterior displacement of the great vessels. Intravenous pyelography is the most definitive noninvasive diagnostic test. Suggestive of retroperitoneal fibrosis on the pyelogram are (1) hydronephrosis with a dilated, tortuous upper ureter, (2) medial deviation of the ureter, and (3) extrinsic ureteral compression. The medial deviation of the ureter is in contrast to lateral displacement that is characteristically associated with retroperitoneal tumors. Since a variety of conditions can produce a similar picture, final confirmation of the diagnosis can be made only after exploratory operative procedures and biopsy of the fibrotic mass. The differential diagnosis includes the primary retroperitoneal tumors, notably the malignant lymphomas, and metastatic tumor from the kidneys,

pancreas, or pelvic organs. Inflammatory conditions to be excluded include tuberculosis, pancreatitis, and intraabdominal inflammation of the intestinal tract.

Treatment

Once the diagnosis is established, the patient should be carefully followed and operative intervention timed properly. Improvement may be anticipated in some patients with supportive measures alone. The onset of urinary tract infection or depression of renal function necessitates operative intervention.

The discontinuance of methysergide is sometimes followed by a reversal of the fibrotic process with an improvement in symptoms. Steroids, antibiotics, and x-ray therapy have been used with inconsistent results. Tamoxifen occasionally produces a beneficial response. The self-limiting nature of the disease and the reports of spontaneous resolution in untreated patients make it difficult to evaluate the results of these therapeutic modalities. Steroids are of theoretical use in the early inflammatory stages to diminish the generation of fibrosis or control a hyperimmune reaction, if one exists. For patients in the prefibrotic stage of the disease with renal insufficiency and prominent constitutional symptoms, steroid-induced regression of the inflammatory edema may reestablish urinary patency and facilitate elective, rather than emergency, surgery. Usually, an advanced stage of fibrosis has been reached before diagnosis.

Surgical treatment is directed toward relief of the tubular obstructions, which are usually urinary, less often vascular, and rarely intestinal. Since the disease is fundamentally a midline process that is often bilateral, a midtransabdominal approach offers the best exposure. Several deep biopsies of the mass should be obtained to eliminate the possibility of an underlying neoplasm, because neoplasms may produce a similar picture, particularly tumors of lymphatic origin. The aorta, inferior vena cava, small bowel mesentery, sigmoid mesocolon, and ureters should be examined for possible involvement.

Ureterolysis with intraperitoneal transplantation is the most effective means of relieving obstruction of the involved ureter. This consists of freeing the ureter from the enveloping mass of fibrous tissue and transferring it into the peritoneal cavity, closing the posterior peritoneum behind it. Lateral reposition of the ureter within the retroperitoneal space has been reported to yield equally good results. A preliminary nephrostomy may be indicated if bilateral ureteral obstruction has resulted in severe renal impairment with uremia. On rare occasion, it will be necessary to reimplant the mobilized ureter into the bladder or, if this is not technically possible, to perform renal autotransplantation. Aortic or iliac artery obstruction is best treated by arteriolysis or bypass with a synthetic vascular graft.

Symptoms resulting from venous obstruction are best treated with elevation and elastic support to the lower limbs until a sufficient collateral venous system develops. The extent of any permanent venous insufficiency depends on the availability of collateral pathways and the competency of deep vein valves. Release of the obstructed vein from its fibrous encasement may be difficult and hazardous, and bypass procedures for obstruction of the inferior vena cava have been uniformly unsuccessful. The prognosis of the disease is generally good, provided that appropriate treatment has been instituted before the development of irreversible renal damage.

Retroperitoneal Tumors

Primary tumors of the retroperitoneum develop independently from cells distinct from the major retroperitoneal organs such as the pancreas, kidney, adrenal gland, and major blood vessels. The most common malignancies in this area are sarcomas, though lymphomas, extragonadal germ cell tumors, and carcinomas also occur. Approximately 6000 soft tissue sarcomas are noted annually in the United States, and nearly 1000 of these occur in the retroperitoneum. The cause of most soft tissue sarcomas is unclear. Rarely, these tumors are associated with a history of therapeutic radiation usually more than 10 years before onset. The incidence of reported sarcomas after therapeutic radiation is low, ranging from 0.03 to 0.3 percent, with the most common type being malignant fibrous histiocytoma. Exposure to vinyl chloride, thorium dioxide, and other agents is associated with sarcomas. Several familial disorders such as Gardner's syndrome, familial retinoblastoma, neurofibromatosis, and Li-Fraumeni syndrome are associated with benign and malignant soft tissue tumors. Germ-line mutations of the *p53* tumor suppressor gene on chromosome 17 is present in some of these familial disorders. Retroperitoneal sarcomas are relatively rare mesenchymal and neurogenic tumors often occurring in the fifth or sixth decade of life. Despite their rarity, they should be considered in the differential diagnosis of an unknown abdominal mass.

Pathology

Retroperitoneal tumors generally arise from mesodermal or neuroectodermal tissues or remnants of the embryonal urogenital apparatus. Tumors may arise from fat, aerolar tissue, vascular or nervous tissue, muscle, fascia, lymphatics, and nodal tissues. Tumors of smooth muscle origin, germ cell tumors, teratomas, and other complex lesions also may occur. Table 33-2 provides a classification of benign and malignant characteristics.

Lymphomas and retroperitoneal sarcomas are the most common malignant lesions of the retroperitoneum. Among the later, liposarcomas and leiomyosarcomas are the most frequent, but while it is of pathologic interest to determine the cell of origin of these retroperitoneal sarcomas, it is of greater importance to define their margins, size, and histologic grade to obtain prognosis (Table 33-3). These tumors develop pseudocapsules of collagen, vessels, inflammatory tissues, and tumor cells that extend within the pseudocapsule. Local tumor recurrence is common when these tumors are simply enucleated. Because of their extension within loose retroperitoneal tissue planes and their relationship to vital structures, it is difficult to obtain negative surgical margins during tumor removal. The histologic grade is the most important characteristic to determine prognosis. Grade is defined based on degree of tumor differentiation, nuclear atypia, amount of fibrous/myoid stroma, mitotic activity, cellular pleomorphism, presence of necrosis, and cellularity. The American Joint Commission on Cancer criteria distinguish three grades and tumor sizes 5 cm and less or more than 5 cm to provide prognostic criteria, but other systems use only two grades. A low grade is assigned to myxoid and well-differentiated liposarcoma and high grade to rhabdomyosarcoma, synovial sarcoma, and alveolar soft part sarcoma (Table 33-4).

If metastasis occur, it typically spreads hematogenously to liver and lung. Lymph node metastases are rare, occurring in less than 5 percent of patients, except in those with embryonal rhabdomyosarcoma, lymphangiosarcoma, and epitheloid sarcoma, which may be somewhat more frequent. Complete nodal dissections are unnecessary, but gross tumor removal with adequate margins should be attained whenever possible.

The tumors may be solid, cystic, or a combination of both. Their color varies from white (fibroma), to yellow (lipoma), to pink or red (sarcoma) depending on the predominant tissue. They may be single or multiple and vary in size from small outgrowths to tumors weighing as much as 18 kg (40 lb). The predominantly cystic tumors usually are benign, whereas the solid tumors are most often malignant.

Clinical Manifestations and Diagnosis

Because of the great dimension, loose boundaries, and flimsy areolar tissues of the retroperitoneal space, tumors in this location can attain a large size before any symptoms occur. Sarcomas grow along fascial plains and envelop rather than directly invade nearby organs. Symptoms are produced by compression or obstruction of adjacent tissues.

Patients may present initially with a history of an enlarging mass in the abdomen, a vague abdominal discomfort, or a sense of fullness or heaviness. Pain may become severe if compression of adjacent nerves or nerve plexuses occur. Gastrointestinal stromal cell tumors (leiomyosarcomas) may result in nausea, vomiting, intestinal obstruction, and gastrointestinal bleeding. Occasionally, fever, weakness, weight loss, and genitourinary symptoms can occur depending on the size and location of the neoplasm. Pelvic tumors give rise to urinary frequency or rarely anuria. They also may compress pelvic veins resulting in lower extremity swelling and varicosities.

Physical examination reveals the presence of an abdominal mass that is usually nontender, firm, and rubbery. It should be defined as fixed or movable and may involve a flank, the midline, or the entire abdomen. Pelvic tumors can be felt on vaginal and rectal examinations. Examination of the regional nodal areas for metastases should be performed. The scrotum should be examined for the possibility of a testicular neoplasm causing retroperitoneal adenopathy. If there is a question of a testicular mass, ultrasound should be done. The diagnosis of a retroperitoneal tumor often is by exclusion of other abdominal masses such as those involving the kidneys, adrenals, pancreas, aorta, ovaries, or spleen. Laboratory studies such as serum α -fetoprotein and b-human chorionic gonadotropin measurements should be obtained to exclude retroperitoneal germ cell tumors.

The primary roentgenographic diagnostic modality is the CT scan (Fig. 33-22). Use of fast, helical CT scanners allows excellent definition of the extent of the mass, the potential involvement of vital organs or major vascular structures with the primary tumor, and the involvement of the liver, lungs, or peritoneal cavity with metastases. MRI is an excellent diagnostic tool permitting sagittal and coronal views and providing a three-dimensional image to define tumor extent and resectability. Use of contrast enhancement with gallium may be helpful to improve specificity and sensitivity of the examination. These studies can determine tumor size and whether the lesion is solid or cystic. They also are useful for planning the operative approach using abdominal, flank, or thoracoabdominal incisions. Ultrasound, gastrointestinal barium studies, and angiography (aortography) are rarely indicated except when gastrointestinal stromal tumors are suspected.

If lymphoma is suspected as the cause of a retroperitoneal mass, a CT-directed needle biopsy often can be done (Fig. 33-23). Fine-needle aspirates are used, but most pathologists require analysis of cells and stroma for diagnostic accuracy, requiring a percutaneous needle biopsy. The exact location of the retroperitoneal mass, its appearance, and the presence or absence of adenopathy and lymphoma (B) symptoms of fever, weight loss, and night sweats dictate whether a biopsy should be performed. Preoperative tissue diagnosis is useful only if a lesion treated nonsurgically could be determined.

Treatment

Preoperative preparation of the patient should include a mechanical and antibiotic bowel preparation when a bowel resection may be necessary for complete tumor removal. Adjacent involved organs may need to be resected in order to obtain complete gross tumor removal. These organs include the kidney, colon, pancreas, and spleen. Adjacent organ resection may be necessary in 50 to 80 percent of patients in order to obtain complete tumor removal. Tumors of the retroperitoneum should not be enucleated because local tumor recurrence is inevitable. Radical resection of multiple organs is inappropriate when the tumor's margin is the aorta or vena cava. Kidney removal should be contemplated only when the opposite kidney is functioning properly. Knowledge of the exact tumor location, adjacent structures, and the surgeon's ability to obtain complete tumor removal are critical to operative planning.

Benign retroperitoneal masses can be removed and cured by simple excision. Malignant lesions, such as sarcomas, require more extensive resection. Low-grade tumors have a propensity for local recurrence and often become more biologically aggressive with each local recurrence. Distant metastases are less common. High-grade sarcomas have a greater propensity for local and distant metastases. Operative treatment is the most effective form of therapy, with the initial operation providing the greatest opportunity for cure of retroperitoneal sarcomas. Complete surgical resection is the primary determinant of patient outcome, with tumor grade being the next most important factor. Operative mortality should be under 5 percent, with morbidity related to hemorrhage, infection, and delayed intestinal function.

After complete surgical resection of retroperitoneal sarcomas, the 5-year survival ranges from 50 to 70 percent, with the most important determinant being the histologic grade (high grade, 20 to 30 percent and low grade, 70 to 80 percent 5-year survival) of the tumor. Partial tumor resection lowers the 5-year survival outcome to approximately 10 to 30 percent. After resection, the most common pattern of recurrence is at the local tumor resection site. Reoperation and resection of an isolated locally recurrent tumor may be of substantial benefit by improving long-term survival. Strategies to reduce the incidence of locoregional recurrence are important.

Radiotherapy combined with operative resection may be beneficial in reducing the incidence of local tumor recurrence after resection. The dose of radiotherapy must be large, approximately 6000 cGy or greater, to effect substantial benefit. At these doses, the incidence of complications such as intestinal enteritis, fibrosis, and obstruction is prohibitive. A strategy to reduce the radiation dose to normal adjacent organs is to use intraoperative radiotherapy (IORT). One method uses plexiglas cones to move normal tissues aside from the tumor bed, and external-beam treatment is given. Another method uses brachytherapy with radioactive iodine seeds or irridium wires to deliver high doses to the local tumor bed while sparing adjacent tissues. In the former approach, investigators used IORT and external-beam radiotherapy and noted a significant improvement in local control, although painful neuritis was a major side effect. Radiotherapy alone is not curative for treatment of soft tissue sarcomas. Patients with unresectable disease may achieve substantial palliation. Radiation treatment is reserved for patients having pain with unresectable tumors, those who develop tumor recurrence, or those who have positive tumor margins after primary resection.

Chemotherapy does not improve outcome when used in an adjunct setting for adults with retroperitoneal soft tissue sarcomas. For patients with metastatic disease, multidrug chemotherapy uses Adriamycin as the key component. Tumor response can occur in up to 30 percent of patients, but survival outcome benefit has not been demonstrated.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 34 - Abdominal Wall Hernias

George E. Wantz

Hernias of the abdominal wall are the most common condition requiring major surgery. Despite the frequency of surgical repair, perfect results continue to elude surgeons, and the rate of surgical failure (recurrence) is humbling. The outcome of hernia surgery is highly surgeon-dependent, and Astley P. Cooper's oft-quoted statement of 1804 is still pertinent and an appropriate introduction to this chapter: "No disease of the human body, belonging to the province of surgeons requires in its treatment a greater combination of accurate anatomical knowledge with surgical skill than hernia in all its varieties."

GENERAL CONSIDERATIONS

Definitions

A hernia (Latin, rupture; Greek, bud) is defined as a protrusion of a viscus through an opening in the wall of the cavity in which it is contained. Clinically the important part of the definition is the protrusion, because without the protruding viscus a diagnosis of hernia is essentially impossible. Anatomically the important features of a hernia are the hernial orifice and the hernial sac. The hernial orifice is the defect in the innermost aponeurotic layer of the abdomen, and the hernial sac is an outpouch of peritoneum. The neck of a hernial sac corresponds to the hernial orifice. The size of a hernia is determined by the dimension of the neck and the volume of the distended sac. A hernia is external if the sac protrudes completely through the abdominal wall, interparietal if the sac is contained within the abdominal wall, and internal if the sac is within the visceral cavity. The hernia is reducible when the protruded viscus can be returned to the abdomen and irreducible when it cannot. The manual manipulation required to reduce viscera entrapped in a hernial sac is known as *taxis*.

A strangulated hernia is one in which the vascularity of the protruded viscus is compromised, usually at the neck. Strangulation is most likely to occur in hernias that have small orifices and relatively voluminous sacs. Strangulation is always serious and can be fatal, and the potential for strangulation is the main reason for the repair of hernia. The word "incarceration" is frequently used in connection with a hernia. An incarcerated hernia is nothing more than an irreducible hernia, and the term does not refer, as some surgeons contend, to an irreducible hernia that is on the verge of becoming strangulated. A Richter's hernia is a hernia in which the contents of the sac consist of only one side of the wall of the intestine (always antimesenteric). In a Richter's hernia, strangulation of the bowel wall occurs without concomitant intestinal obstruction.

Sites of Herniation

Hernias of the abdominal wall occur only in areas where aponeurosis and fascia are devoid of the protecting support of striated muscle. Many such sites are normally present, but some may be acquired through muscular atrophy or surgery. Without a counteracting force, the bare aponeurotic areas are subject to the ravages of intraabdominal pressure and give way if they deteriorate or contain anatomic irregularities. Predictably, the common sites of herniation are thus the groin, the umbilicus, the linea alba, the semilunar line of Spieghele, the diaphragm, and surgical incisions. Other similar but rare sites of herniation are the perineum, the superior lumbar triangle of Grynfeltt, the inferior lumbar triangle of Petit, and the obturator and sciatic foramina of the pelvis.

Symptoms and Diagnosis

Some patients have hernias but are unaware of them until the hernia is pointed out to them. The natural history of all hernias of the abdominal wall is slow enlargement to the point of irreducibility and disfigurement, with the risk of strangulation ever present. Symptomatic hernias produce a wide variety of nonspecific discomforts related to the contents of the sac and the pressure by the sac on adjacent tissue. The discomforts produced by hernias are always worse at the end of the day and are relieved at night when the patient reclines and the hernia reduces. Groin hernias are not a cause of testicular pain; groin pain without a demonstrable hernia usually does not indicate or herald the onset of a hernia. Most hernias develop insidiously, but some are precipitated by a single forceful muscular event.

Hernias are easy to diagnose on physical examination. Typically a hernial sac with its contents enlarges and transmits a palpable impulse when the patient strains or coughs. Usually the patient must stand during the examination because it is impossible to palpate a reduced groin hernia with certainty when the patient is supine. Hydroceles may resemble an irreducible groin hernia, but they transilluminate, whereas hernias do not. Hernias undetectable by physical examination can be demonstrated by the radiologist using ultrasonography, computerized tomography, magnetic resonance imaging, and herniography; in the latter a nonirritating contrast agent is injected into the peritoneal cavity. Strangulation produces intense pain in the hernia followed quickly by tenderness, intestinal obstruction, and signs and symptoms of sepsis. A strangulated hernia, in contrast to an irreducible hernia, does not enlarge or transmit an impulse when the patient coughs. Taxis of a strangulated hernia is contraindicated if there is sepsis or if the contents of the sac are thought to be gangrenous.

Indications for Surgery

In general, all hernias should be repaired unless local or systemic conditions in the patient preclude a safe outcome. The possible exception to this generalization is a hernia with a wide neck and shallow sac that is expected to enlarge slowly. Trusses and surgical belts are helpful in the management of small hernias when operation is contraindicated. However, trusses are contraindicated for patients with femoral hernias.

HERNIAS OF THE GROIN

The groin is one of the natural weak areas in the abdominal wall and is the most common site for abdominal herniation. Both sexes of all ages are afflicted, but men are 25 times more likely to have a groin hernia than women.

Hernias arising above the abdominocrural crease are inguinal (Latin, groin), and those arising below the crease are femoral (Latin, thigh) or crural (Latin, leg). Inguinal hernias may be direct or indirect. The sac of an indirect inguinal hernia passes obliquely or indirectly toward and ultimately into the scrotum, in which case it is known as a scrotal hernia. The sac of a direct inguinal hernia protrudes directly outward and forward. Clinically distinguishing an indirect from a direct inguinal hernia is often impossible and is of little importance since the operation to repair them is the same. In men, indirect hernias outnumber direct hernias at a ratio of 2:1, whereas in women direct hernias are a rarity. Strangulated indirect hernias in men may produce concomitant strangulation of the spermatic cord and testicle.

Femoral hernias are usually considered a separate entity but are actually a form of direct inguinal herniation. Femoral hernias are uncommon, comprising about 2.5 percent of all groin hernias. They occur occasionally in women, especially in multiparous elderly women, but not as frequently as inguinal hernias; in men they are rare. Ten percent of women and 50 percent of men with femoral hernias have or will develop an inguinal hernia. Femoral hernias almost always appear as an irreducible mass about the size of a walnut at the medial base of Scarpa's femoral triangle. A femoral hernia appears irreducible, even though the sac may be empty, because fat and lymph nodes from the femoral canal surround the sac. A solitary enlarged lymph node or a synovial cyst may mimic a femoral hernia exactly. The infrequent reducible femoral hernias may be confused with a saphenous varix.

Strangulated Richter's type femoral hernias occur relatively frequently and carry a significant morbidity and mortality. The diagnosis of such strangulated femoral hernias is invariably delayed because they develop without intestinal obstruction and with minimal local manifestation until the entrapped knuckle of small bowel is gangrenous. A bruit over the femoral vein is an indication that the adjacent femoral hernia is incarcerated or strangulated because the unforgiving hernia compresses the vein.

Both indirect inguinal and femoral hernias are twice as common on the right as on the left. In indirect inguinal hernias this is attributed to a delay in the atrophy of the processus vaginalis that follows the normally slower descent to the scrotum of the right testis. In femoral hernias the right side's predominance is credited to the sigmoid colon's tamponading the left femoral canal.

Epidemiology

Hernias are a common health problem. Exactly how prevalent they are is not known, but the accepted estimated incidence is 3 to 4 percent of the male population. Aging increases the incidence of groin hernias, the likelihood of strangulation, and the need for hospitalization. Strangulation, the most common and serious complication of a groin hernia, increases mortality and morbidity, the need for general or regional anesthesia, and the need for hospitalization. Strangulation occurs in 1.3 to 3.0 percent of groin hernias. Strangulation occurs at the extremes of life and in the elderly is a grave surgical problem. Most strangulated hernias are indirect inguinal hernias, but the femoral hernia has the highest rate of strangulation (5 to 20 percent) of all hernias.

The probability of a groin hernia's strangulating varies with location and duration. For an inguinal hernia the probability of strangulation is 2.8 percent after 3 months from development and increases to 4.5 percent after 2 years; for femoral hernias it is 22 percent at 3 months and 45 percent at 21 months. Femoral hernias require urgent surgery. The probability of strangulation, in both inguinal hernia and femoral hernia, is greatest in the first 3 months, suggesting that patients with a short history of herniation should be referred urgently to a surgeon.

Anatomy of the Groin

Many surgeons have contributed to the study of the anatomy of the groin and groin hernia, and the names of some are eponymically memorialized. The most important contributions were made by Cooper, Anson and McVay, and Fruchaud. Cooper, in a two-volume masterpiece (1804–1807), was the first to accurately describe and depict the groin. Among other things he named the endopelvic fascia and the transversalis fascia and described the iliopectineal ligament, now called Cooper's ligament. Anson and McVay, in a series of monumental publications beginning in 1937, defined, described, and statistically analyzed the anatomic structures of the groin. Fruchaud, in a spectacularly illustrated two-volume work in 1956 on the anatomy and treatment of groin hernias, unified the hernias of the groin by redefining the weak area of the groin to include the area of occurrence of all the hernias of the groin.

The groin is the area of junction of the lower abdomen and the thigh. The anatomy of the various myoaponeurotic layers and other important structures is illustrated in Fig. 34-1. The only structurally important layer of the groin of concern to hernia surgeons is the innermost aponeuroticofascial layer of the abdomen. It consists of the transverse abdominal muscle, the transverse aponeurosis, and the transversalis fascia. To visualize this layer from the anterior approach, the superficial structures consisting of the external oblique abdominal muscle, the inguinal ligament, and the internal oblique abdominal muscle must be divided, reflected, or partially removed. From the posterior exposure it is necessary to reflect only the peritoneum.

The transverse aponeuroticofascia at the upper border of the femoral sheath is known by North American surgeons as the iliopubic tract, by French surgeons as the bandelette of Thomson, and by English surgeons as the deep crural arch. This portion of the transverse aponeuroticofascia becomes the inferior crus of the deep ring. The superior crus of the deep ring is formed by the portion of the transverse aponeurosis known as the transverse aponeurotic arch. The transverse aponeurosis lies on the posterior aspect of the transverse abdominal muscle and cannot be seen from the anterior. From the posterior, however, it forms a very obvious broad aponeurotic arch that inserts on the pectineal line of the pubis (pecten pubis). The portion of the tendon of the rectus abdominis muscle that recurves laterally onto the pecten pubis is known as Henle's ligament.

The suspended portion of the inguinal ligament is mobile because it is attached to the femoral sheath by the investing fascia of the external oblique abdominal muscle only.

The bulk of the cremaster muscle arising from the lowermost fibers of the internal oblique abdominal muscle embraces the inferior aspect of the spermatic cord in the inguinal canal. The cremaster vessels arise from the inferior epigastric vessels and pass through the posterior wall of the inguinal canal by their own foramen. They are accompanied by the genital nerve and supply the cremaster muscle and the tunica of the testes. They are not part of the spermatic cord and may be divided without damaging the testicle to expose the floor of the inguinal canal during hernioplasty.

A fascial condensation in the region of the inferior epigastric vessels is known as Hesselbach's interfoveolar ligament and is not always obvious. Muscular contraction retracts the ligament and complements the shutter action of the internal oblique abdominal muscle by pulling the cord at the deep ring laterally and superiorly.

The path of the inguinal canal through the abdominal wall is oblique (Figs. 34-1 and 34-2). The anterior wall consists of the aponeurosis of the external oblique abdominal muscle. The posterior wall is the aponeurosis of the transverse abdominal muscle and the transversalis fascia. The angle of entrance of the deep ring is acute medially and obtuse laterally. The transverse aponeurosis and transversalis fascia comprise the medial border of the deep ring. It is fibrous, definable, and palpable and is the margin surgeons repair during hernioplasty. The lateral border of the deep ring is formed by the transverse abdominal muscle. It is soft, elastic, muscular, and indistinct. The internal oblique abdominal muscle covers the deep ring and forms the shutter mechanism. This muscle obscures the deep ring from view and prevents its palpation from the anterior approach. The lateral border of the deep ring is so blunt as to be indistinguishable and unrecognizable even from the posterior view.

The spermatic cord begins at the deep ring and contains the vas deferens and its accompanying artery, the testicular artery and veins, lymphatics, autonomic nerves, and a variable quantity of fat. Collateral circulation for the testicle is abundant and comes from the vessels supplying nearby structures. Distally in the cord and mainly beyond the pubic tubercle are the delicate veins of the pampiniform plexus from which the testicular veins arise.

The elements of the spermatic cord split proximal to the deep ring, and the vas deferens and its artery descend to the seminiferous tubules, while the spermatic vessels, lymphatics, and lymph nodes incline superiorly to the region of the kidneys. At the level of the deep ring there is one testicular artery and two or three testicular veins.

The iliohypogastric and ilioinguinal nerves and the genital branch of the genitofemoral nerve are the nerves in the groin that are of concern to the surgeon. The iliohypogastric and ilioinguinal nerves, which intermingle, are sensory to the skin of the groin, the base of the penis, and the medial upper thigh. The position of these nerves, arising from T12 and L1, are variable. The ilioinguinal nerve is found in its typical anatomic position on top of the spermatic cord in only 60 percent of patients. In others its course and position are behind or within the cremaster muscle. Often it cannot be found or it exists in fibers too small to be seen.

The genital nerve is both motor and sensory and innervates the cremaster muscle and the skin of the side of the scrotum and labia. It arises from L1 and L2 and may substitute for the ilioinguinal nerve when this nerve is deficient. In the inguinal canal, the genital nerve lies on the iliopubic tract, and it accompanies the cremaster vessels to form a neurovascular bundle that is usually divided during anterior inguinal hernioplasty.

Anatomy of Groin Hernia

An indirect hernial sac is actually a dilated persistent processus vaginalis (Fig. 34-3). It passes through the deep ring, lies within the spermatic cord, and follows the indirect course of the cord to the scrotum. At the deep ring the sac occupies the anterior lateral side of the cord. Properitoneal fat is often associated with an indirect hernial sac and is known as a lipoma of the cord, although the fat is not a tumor. Lipomas of the spermatic cord can mimic an indirect hernial sac exactly.

An indirect hernial sac is said to be complete if the sac has descended to the testes and filled the side of the scrotum, and incomplete when it has not. When the processus vaginalis remains completely open, the testicle will be within the sac. This type of hernia is known as a congenital hernia or communicating hydrocele. They are common in infants but rare in adult men.

Retroperitoneal organs such as the sigmoid colon, cecum, and ureters may slide into an indirect sac. They thereby become a part of the wall of the sac and are susceptible to injury during hernioplasty. These sliding hernias often are large and partially irreducible.

Direct inguinal hernial sacs originate through the floor of the inguinal canal, that is, Hesselbach's triangle; they protrude directly; and they are contained by the aponeurosis of the external oblique muscle (see Fig. 34-3). Only rarely do they so enlarge as to force their way through the superficial ring and descend into the scrotum. Direct hernias usually are diffuse and involve the entire floor of the inguinal canal. The less common discrete direct hernias have small orifices and diverticulum-shaped sacs. Direct inguinal hernias also originate lateral to the inferior epigastric vessels and present through the deep ring or interstitially through slips of fatty muscular atrophy in the shutter muscles of the deep ring (see Fig. 34-3). These types of direct inguinal hernia are rare and commonly referred to erroneously as extrafunicular indirect and interstitial indirect hernias, respectively. They do not follow the spermatic cord and enlarge interparietally. The inferior epigastric vessels are not a proper anatomic boundary distinguishing a direct hernia from an indirect hernia in all instances, as some surgeons believe. The bladder is a common sliding component of a direct hernial sac.

Femoral hernial sacs originate from the femoral canal through a defect on the medial side of the femoral sheath (see Fig. 34-3). The femoral canal contains one or two lymph nodes, the largest of which is named Cloquet by the French and Rosenmüller by the Germans. These structures are expelled from the femoral canal by a peritoneal protrusion and frequently create a palpable mass. Femoral hernias also occur through the anterior femoral sheath. Prevascular femoral hernias are rare and are most often encountered as a recurrence after inguinal hernioplasty in which the femoral sheath or iliopubic tract has been used in the repair.

Fruchaud's Myopectineal Orifice

Traditionally the hernias of the groin have been defined as separate entities, which creates confusion. Fruchaud's (1956) concept of the anatomy of hernias of the groin is important. Rather than viewing hernias solely by their varied clinical presentation (i.e., indirect, direct, femoral, prevascular, interstitial), Fruchaud emphasized their common origin by noting that all the hernias of the groin begin within a single weak area that he called the myopectineal orifice (Fig. 34-4).

The myopectineal orifice is the area in the groin bounded superiorly by the internal oblique muscle and the transverse abdominal muscle, laterally by the iliopsoas muscle, medially by the rectus muscle and sheath, and inferiorly by the pecten pubis. This bony-muscular framework is bridged and bisected by the inguinal ligament, traversed by the spermatic cord and femoral vessels, and sealed like a drum on its inner surface by the transversalis (endopelvic) fascia only. Therefore the integrity of the myopectineal orifice is dependent on the transversalis fascia. A groin hernia is defined as protrusion of a peritoneal sac through the transversalis fascia spanning the myopectineal orifice. Failure of the transversalis fascia to retain the peritoneum then becomes the fundamental cause of all hernias of the groin.

The inguinal ligament and its broad, recurved, lacunar medial insertion on the pecten pubis are only loosely attached to the adjacent iliopubic tract and transversalis fascia (see Fig. 34-1 C and Fig. 34-3). Nevertheless they strongly brace the myopectineal orifice, both separating inguinal herniation from femoral herniation and defining the medial border of the orifice of the femoral canal. The passage of the testicle through the abdominal wall during the embryonic stage weakens and enlarges the myopectineal orifice above the inguinal ligament, predisposing men to indirect and direct inguinal hernias. In women the increased diameter of the true pelvis as compared to men proportionally widens the femoral canal, probably predisposing women to femoral herniation.

Classification

Attempts to classify hernias have been made. These classifications assign numbers and letters to the various types and sizes of hernias of the groin and are intended to promote a common surgical understanding. Thus far, however, the classifications are incomplete and lack unanimity. Until accord is reached, hernias should be described by their type and location, volume of the sac, and diameter of the neck.

Etiology

Inguinal hernias can be congenital or acquired, and in both a family history of groin hernias is usually strongly positive. Most hernias of the groin therefore may be said to be transmitted genetically. All indirect inguinal hernias are congenital and result from a patent processus vaginalis, with which the patient is born. A patent processus vaginalis is found in 80 percent of newborns and in 50 percent of 1-year-olds. Closure continues until the age of 2 years. The incidence of a patent processus vaginalis in adults is 20 percent. Having the potential for a hernia does not mean that a hernia will develop. Other factors must be present to cause failure of the transversalis fascia to retain the visceral sac in the myopectineal orifice.

The erect stance of human beings, in contrast to that of four-legged animals, promotes herniation by stretching and exposing the groin and, when a hernia is present, permitting the dependent intestines to drop into the hernial sac.

Muscle deficiency contributes to herniation. Congenital or acquired insufficiencies of the internal oblique abdominal muscles in the groin expose the deep ring and the floor of the inguinal canal to the destructive effects of intraabdominal pressure (Fig. 34-5).

Destruction of connective tissue resulting from the physical stress of intraabdominal pressure, smoking, aging, connective tissue disease, and systemic illnesses reduces the strength of the transverse aponeurosis and fascia. Fracture of the elastic fibers and alterations in the structure, quantity, and metabolism of collagen have been demonstrated in the connective tissue structures of the groin in patients with hernias.

Other factors may also have an effect in some cases. Abdominal distention and chronic increase in intraabdominal pressure from ascites and peritoneal dialysis may damage the myopectineal orifice and cause a patent processus vaginalis to dilate. Fracture deformities of the pelvis and denervation of the shutter mechanism following a low cosmetic appendectomy incision are well-known but uncommon causes of inguinal herniation.

Inguinal hernias of all types occur equally in sedentary and in physically active men. Vigorous physical activity per se is not a cause of inguinal herniation, although strenuous effort may aggravate predisposing factors and precipitate herniation. The notion that carcinoma of the colon is a cause of inguinal hernia is incorrect.

Basics of Groin Hernioplasty

The object of groin hernioplasty is to prevent peritoneal protrusion through the myopectineal orifice. The integrity of the myopectineal orifice is restored in two fundamentally different ways, which are based on Fruchaud's concept of groin hernia, namely, (1) aponeurotic closure of the myopectineal orifice to the extent necessary, and (2) replacement of the defective transversalis fascia with a large synthetic prosthesis. The two methods are sometimes combined.

Hernias are repaired either anteriorly through a groin incision, in which case the structures in and around the inguinal canal must be divided in order to reach the innermost aponeuroticofascial layer, or posteriorly through an abdominal incision, in which case the hernial orifices are exposed directly on entering the properitoneal space. The anterior approach is used in all the classical repairs and for more than a century has been the most popular incision for inguinal hernioplasty. Posterior hernia repairs are called properitoneal hernioplasty. Until recently the distinct advantages of a posterior approach were largely ignored.

Tension is the principal cause of failure of all the hernioplasties that close the myopectineal orifice by aponeurotic approximation. Assiduous efforts to prevent suture-line tension are essential, and sutures must never be drawn up or tied so tightly as to cause necrosis. Permanent monofilament synthetic sutures are preferable.

Synthetic mesh prostheses currently play a major role in the management of all hernias of the groin. Synthetic mesh prostheses are used to patch or plug the myopectineal orifice, to reinforce a classical repair, and to replace the transversalis fascia. Prosthetic techniques have surfaced in bewildering proliferation, each having enthusiastic proponents.

Anterior Classical Groin Hernioplasty

Only three anterior classical hernioplasties have withstood the test of time and are still used today: the Marcy simple ring closure, the Bassini operation (either in its original form or as performed at the Shouldice Hospital in Toronto), and the McVay-Lotheissen Cooper ligament repair. All produce equally satisfactory results in primary hernias when correctly indicated and are easily performed with local anesthesia in adults. Formerly, recurrent inguinal hernias were fixed by one of the classical repairs, but nowadays prosthetic techniques are preferred, as the results are distinctly better. Classical hernioplasty has three parts: *dissection of the inguinal canal, repair of the myopectineal orifice,* and *closure of the inguinal canal*. The dissection and closure of the inguinal canal are essentially identical in all variants of classical repair, and the repairs vary to the extent to which the myopectineal orifice is repaired.

A *complete dissection* consists of opening the inguinal canal, preservation (if possible) of the ilioinguinal nerve, division of the cremaster muscle and its neurovascular bundle to expose the deep ring, mobilization of the spermatic cord, division and excision of the posterior wall of the inguinal canal in areas that it is weak, assessment of the transverse aponeurosis, elimination of the peritoneal sac, removal of cord lipomas, and relaxing incisions when necessary (Fig. 34-6). For a successful hernioplasty, the dissection is as important as the repair. Division of the cremaster muscle and of the posterior wall of the inguinal canal are important steps that many surgeons omit, although Bassini routinely used them and surgeons at the forefront of inguinal hernia repair have repeatedly emphasized their importance.

Gentleness and precise, sharp dissection are essential to minimize surgical trauma to the spermatic cord. An important point is never to dissect beyond the pubic tubercle unless absolutely necessary. Hydroceles, cysts of the cord, and other conditions in the scrotum should never be dealt with at the time of inguinal hernia repair because of the risk of damaging the spermatic cord and causing testicular atrophy. Also, the distal portion of indirect hernial sacs should never be dissected from the cord unless absolutely necessary. They should be divided at the neck, with the distal part allowed to remain in situ while the proximal end is dissected beyond the neck and amputated with or without ligation. The retained distal sac will not become a hydrocele, although aspiration of serosanguinous fluid may be necessary in the early postoperative period. Sliding indirect hernial sacs are an exception and must be carefully dissected from the cord because they contain a retroperitoneal organ. Sliding sacs do not require other special treatment and are merely returned to the abdomen. Direct hernial sacs of all sizes are merely inverted and not opened.

Repair of the myopectineal orifice in men includes reconstruction of the deep ring. Contrary to the belief of some surgeons, the anatomy of the deep ring is such that strangulation of the spermatic cord by reconstruction of the posterior wall of the inguinal canal is virtually impossible. Indeed, insufficient repair of the deep ring is the principal cause of indirect recurrence.

The Marcy repair of the myopectineal orifice consists of tightening an enlarged deep ring only. This procedure, commonly called simple ring closure, is indicated in men and women who have indirect hernias with only minimal damage to the deep ring (Fig. 34-7). This procedure was clearly illustrated in Marcy's remarkable 1892 book on hernias. Whether or not this is the hernioplasty Marcy claimed he performed in 1871 is uncertain. Marcy persistently claimed that it was, even though his contemporaries never concurred. Students of hernias have conflicting opinions of Marcy's claim to priority for performing the first hernioplasty, although all accredit simple ring closure to him. The operation restores the anatomy of the deep ring by placing one or two sutures in the transverse aponeurotic arch and the iliopubic tract just medial to the spermatic cord.

Simple ring closure is the hernioplasty of choice for women with indirect inguinal hernias. After dividing the round ligament and eliminating the sac, the deep ring is abolished with a few permanent sutures.

The *Bassini-Shouldice hernioplasty* repairs the myopectineal orifice superior to the inguinal ligament—that is, the deep ring and Hesselbach's triangle—and therefore is indicated in all direct and indirect inguinal hernias. Bassini is credited with developing and performing the first modern hernioplasty. Bassini wrote six papers about his surgical technique and experiences between 1887 and 1894, but it was his 1890 paper in German that ignited the enthusiasm of surgeons worldwide. Unfortunately, from the very beginning surgeons misinterpreted Bassini's intent and modified and corrupted the operation. In North America, and to some extent in Europe, the Bassini repair consisted merely of high ligation of the sac and approximation of the conjoined tendon and the internal oblique abdominal muscle to the shelving of the inguinal ligament with interrupted sutures.

The authentic Bassini operation, which is currently performed in Europe and was detailed by Catterina (1932), includes deliberate and complete dissection to expose the anatomy in its entirety and a repair in which the internal oblique abdominal muscle, the transverse abdominal muscle, and the transversalis fascia (Bassini's triple layer) are approximated to both the femoral sheath and the shelving edge of the inguinal ligament with interrupted sutures (Fig. 34-8).

Obney and Ryan, working with Shouldice at his hernia clinic, developed de novo a hernioplasty (1950–1953) that is essentially identical to the original Bassini procedure. The dissection is exactly the same, and the repair approximates the identical myoaponeuroticofascial layers, albeit not with interrupted sutures but by precise imbrication with continuous sutures (Fig. 34-9). The Shouldice hernioplasty is, in fact, the modern equivalent of the original Bassini procedure.

The Bassini-Shouldice hernioplasty produces exceptional results and has become the worldwide standard to which other hernioplasties are compared. Nevertheless, the Bassini-Shouldice procedure has been justifiably criticized because it does not repair the femoral canal and may actually cause some femoral hernias by straining the femoral sheath. In addition, the repair is nonanatomic, because the transversalis aponeurosis is sutured to the inguinal ligament. Consequently some surgeons do not use the inguinal ligament and repair the myopectineal orifice superior to the inguinal ligament by simple edge-to-edge approximation of the transverse aponeurotic arch and the iliopubic tract. The technique is called the iliopubic tract repair to differentiate it from the Bassini and the Bassini-Shouldice procedures.

The *Cooper ligament hernioplasty* repairs the three most vulnerable areas for herniation in the myopectineal orifice—the deep ring, Hesselbach's triangle, and the femoral canal—and is therefore indicated for the three common types of hernias of the groin (Fig. 34-10). From his anatomic studies, McVay concluded (1942) that the proper structure to which the aponeurosis of the transverse abdominal muscle should be sewn was Cooper's iliopectineal ligament, where it is normally attached, not the inguinal ligament. He was unaware that Lotheissen had reported the procedure (1897) to repair a recurrent hernia in a woman without an inguinal ligament. Lotheissen immediately recognized the importance of the procedure in the treatment of femoral hernia. Lotheissen's Cooper ligament repair is currently known as the McVay hernioplasty in recognition of the man who popularized it.

In the McVay repair the transverse aponeurotic arch is sutured to Cooper's ligament medially and to the femoral sheath laterally. Exposure of Cooper's ligament and the medial border of the femoral sheath is accomplished by excision of the medial portion of the iliopubic tract (see Fig. 34-6). This extra step in the dissection is important for the accurate placement of sutures.

Relaxing incisions are mandatory because there is otherwise too much tension on the suture line. Although many enthusiasts use it routinely, McVay himself used it in only about half of the hernioplasties he performed; in those with smaller defects he performed what is essentially a Marcy repair.

Closure of the inguinal canal completes the hernioplasty and is the same in all the classical repairs. The aponeurosis of the external oblique abdominal muscle is reapproximated, which thereby incidentally reassembles the superficial ring. The distal stump of a divided cremaster muscle should be attached to the superficial ring to hitch up the testicle.

Special Considerations

Femoral hernias with small orifices in women are repaired from below the inguinal ligament with a few sutures (Bassini technique) or corked with a cylindrical plug of polypropylene mesh (see section on Lichtenstein tension-free hernioplasty) because they are rarely associated with hernias above the inguinal ligament (Fig. 34-11). Large femoral hernias in women and all femoral hernias in men, however, are repaired by the McVay Cooper ligament repair or, even better, by a properitoneal permanent prosthesis. Strangulated femoral hernias are preferably approached properitoneally because this provides direct access to the constricting femoral hernia orifice, easy release of the entrapped bowel by incision of the iliopubic tract and lacunar ligament, and ample room for bowel resection. Strangulated inguinal hernias are easily managed through a groin incision.

In indirect hernias in infants, children, and some young men, the myopectineal orifice and its transversalis fascial seal are not damaged and classical repair is unnecessary; merely eliminating the sac cures the hernia. The operation is called *high ligation of the sac* and consists of opening the inguinal canal and, without mobilization of the spermatic cord, identifying, dividing, and tracing the sac to the deep ring, where it is ligated. The operation is similar to the Ferguson and the Halsted II classical hernioplasties that are now largely obsolete.

In infant girls the fallopian tube and ovary present commonly in hernial sacs. Testicular feminization is found in 1 percent of females with hernias, especially in cases of bilateral hernias. Therefore female infants with an inguinal hernia should have a buccal smear tested for chromatin. If an infant with normal-appearing external genitalia lacks chromatin, the gonad should be left in the inguinal canal. It will serve as an estrogen generator until after puberty, when it should be removed, because of the risk of malignant degeneration. Consequently, in females, a gonad found in the hernial sac without a fallopian tube should be biopsied for identification.

In boys, exploration of the contralateral groin is commonly performed up to the age of 3 years, especially if unilateral hernia is present on the left. This policy, disputed by some, serves to avoid a second hernioplasty at a later age.

Posterior Hernioplasty

Although the posterior properitoneal approach to repairing hernias was used sporadically in the nineteenth century, credit is generally given to Cheatle (1920) and Henry (1936) for elucidating and popularizing the virtues of the approach. They used a midline incision specifically to repair femoral hernia but recognized the potential of the approach for the repair of inguinal hernias. Nyhus and associates (1959) and later Read were the next to become enthused by the approach to treat both femoral and inguinal hernias, but Nyhus was the surgeon who actually championed the method and ultimately codified it. In this approach the properitoneal space was accessed via a short transverse incision in a lower quadrant

rather than through a midline incision. The repair closes the direct space by suturing the iliopubic tract to the transverse aponeurotic arch and narrows the deep ring with a few sutures placed lateral to the spermatic cord.

The results of the properitoneal repair for primary hernioplasty even with a relaxing incision were less than satisfactory, and the procedure never became popular. An exception was the Cheatle-Henry midline approach for strangulated femoral hernias. However, the recent trend of implanting synthetic permanent prostheses for hernia repair has kindled a resurgence in the posterior properitoneal approach, and even Nyhus now buttresses his repairs with polypropylene mesh (Marlex).

Prosthetic Material for Hernioplasty

Nondegradable and biologic-tolerant synthetic mesh prostheses are readily available. Those proven useful are Marlex, Prolene, Trelex, Surgipro, Mersilene, and Gore-Tex. Each has its champions, none is perfect, and in practice the selection of the prosthesis material will be a compromise. They have, for all practical purposes, replaced the inconvenient biologic grafts such as autologous fascia lata.

Marlex, Trelex, and Prolene mesh, which resemble one another, are composed of knitted monofilament fibers of polypropylene. All are porous, slightly elastic, semirigid, and relatively heavy, and they contain plastic memory and buckle when bent in two directions at once.

Surgipro mesh is composed of knitted, braided strands of polypropylene. Its physical characteristics closely resemble those of knitted meshes of monofilament polypropylene.

Mersilene is an open-knitted mesh composed of pure, uncoated, braided fibers of the polyester Dacron. It is porous, soft, light, lacelike, supple, elastic, and without plastic memory; it has a grainy texture that prevents slippage and has only a minimal tendency to buckle when bent in two directions at once.

Gore-Tex is expanded polytetrafluoroethylene (PTFE, or Teflon). It is a nonporous, smooth, supple, fabriclike material containing through-and- through microscopic pores into which fibroblasts grow but through which serosanguineous fluid will not flow.

The prostheses made of polypropylene and polyester desirably incite a prompt fibroblast response and are rapidly integrated in the body with minimal inflammation. Gore-Tex is inert and does not incite fibroplasia or inflammation. It is not integrated in the tissues but rather is segregated by encapsulation. The process of encapsulation is slow, taking as long as 30 to 40 days. Perforated Gore-Tex is available, permitting early fixation and immobilization of the prosthesis during the process of encapsulation. The tendency of intestines to adhere to Gore-Tex is minimal. This feature, which the other permanent prostheses lack, may be advantageous for certain circumstances.

Polyester and polypropylene permanent prostheses should never contact abdominal viscera directly. They provoke binding and intimate adhesions that are difficult to divide and can cause intestinal obstruction and fistulization. Such adhesions can be prevented by interposing the omentum or an absorbable prosthesis between the permanent prosthesis and the bowel. Absorbable prostheses of knitted or woven polyglactin 910 (Vicryl) and polyglycolic acid (Dexon) are available. These prostheses also incite fibroplasia and can cause adhesions. However, they prevent the viscera from touching the permanent prosthesis during mesothelial integration of the permanent prosthesis and prevent grafting of the prosthesis to the viscera. A membranous sheet of Gore-Tex is also available. It is intended to substitute for absorbable meshes for the prevention of adhesions between viscera and polypropylene or polyester prostheses.

Infection and Synthetic Nonabsorbable Prostheses

All the synthetic materials can become sequestered, act like a foreign body and aggravate and prolong infections. Hence the risk of infection is balanced by the risk of recurrence. Theoretically, monofilament mesh ought to tolerate infection better than polyfilament mesh because bacteria can settle into and be hard to dislodge from the interstices of the polyfilament fibers. In practice this is not necessarily true, however, perhaps because the fibers are fine and the braid or twist is loose. The braided, knitted Dacron mesh Mersilene tolerates infection as well as, for example, the nonbraided, solid- fiber polypropylene meshes. If either is infected, integration rather than infection is the rule, provided suitable treatment is instituted and the prosthesis is in contact with healthy tissues and is not sequestered.

Expanded polytetrafluoroethylene (Gore-Tex) is intolerant of early infection because of the slow encapsulation of the material. If an infection develops in the space containing Gore-Tex, the material has to be removed, because there is no chance that it will become incorporated before bacteria have inhabited the microscopic spaces in the material. These spaces are too small to allow entrance of phagocytes and antimicrobial substances.

The management of infections in wounds in which there is a synthetic prosthesis is no different from that of other wounds. All that is needed is the application of sound surgical principles. Superficial infections not directly involving a prosthesis can be expected to heal.

Early infections involving a prosthesis must be treated vigorously and aggressively. The entire prosthesis must be exposed without delay. Failure to expose the entire prosthesis will lead to sinus formation. Local treatment is directed at irrigating purulent material, lysing cellular, fibrous, and fibrinous debris, and destroying the infectious agent. Saline solution, Dakin's solution, granulated sugar, and topical antimicrobial substances are all useful. Complete incorporation can be expected with meshes of both polypropylene and polyester (but not with expanded polytetrafluoroethylene) in 3 to 4 weeks, providing the mesh firmly contacts tissue and is not floating free. Systemic antibiotics, of course, are essential.

Delayed infections involving the prosthesis occur, and the interval between prosthetic implantation and infection may be months or years. In these cases and in all infected prosthetic wounds that have healed with a sinus, it is rarely possible for the prosthesis to become reintegrated, excision of the sequestered mesh is necessary. Only the sequestered mesh must be removed; the integrated mesh can remain.

Prophylaxis of infection is essential. Rigid sterile technique, precise and meticulous surgical technique, and the avoidance of seromas and hematomas with closed suction drains are important. When large pieces of mesh are used, broad-spectrum antibiotics should be administered intravenously shortly before beginning the operation and continued postoperatively until the closed suction drains, if used, are removed. Intraoperatively, topical antibiotics are also commonly used. Prophylactic antibiotics are not necessary when small prostheses of polyester or polypropylene are used.

Inflammatory granulomas are occasionally encountered during prosthetic incisional hernioplasty. When bacteria are seen in a Gram stain of an inflammatory granuloma, the prudent course is not to implant a synthetic prosthesis. A hernia is a better condition to have than an infected prosthesis. However, Gram stains are not always reliable and may be negative while the culture of the granuloma is positive and grows bacteria. Knowing this possibility, some surgeons have proceeded in the presence of a granuloma to implant prostheses of both polypropylene and polyester successfully. Nevertheless, the correct advice is not to proceed.

Anterior Prosthetic Groin Hernioplasties

Tension-Free Hernioplasties

Prosthetic soft-tissue patches have been used for years to reinforce classical repairs but without significantly improving the results. However, when the prosthesis is implanted without a formal repair, thereby obviating tension, results improve dramatically. Lichtenstein credited this idea to R. Newman of Rahway, NJ, but it was Lichtenstein and his group who championed the tension-free hernioplasty and who reported favorable results in a large series of patients (Fig. 34-12). Lichtenstein, aware of the importance of tension as a cause of recurrence, had already developed a tension-free repair for femoral hernias and for recurrent direct and indirect inguinal hernias when the defect is fibrous, circumscribed, and not too large. The technique consists of a prosthetic plug that stoppers the aponeurotic defect (see Fig. 34-11). Lichtenstein got the idea of treating femoral hernias with a plug in the femoral canal from Cheatle, who in 1921 wrote that he “blocked this opening by coiling up into a plug the internal saphena vein.”

Marlex, Prolene, Trelex, and Surgipro mesh are interchangeable and are the preferred prosthesis materials for the tension-free hernioplasties because they handle well and become quickly integrated.

The term tension-free was coined by Lichtenstein to describe his prosthetic hernioplasty, in which a formal repair of the parietal defect was not done. However, tension-free hernioplasty has become the generic term for all such hernioplasties performed through an anterior groin incision. Credit for originating the first tension-free and sutureless hernioplasty goes to Stoppa, who in 1969 introduced a properitoneal hernioplasty in which the parietal defects were not closed and peritoneal herniations were prevented by a large piece of unsutured polyester mesh. This hernioplasty is now known by the descriptive term *giant prosthetic reinforcement of the visceral sac (GPRVS)* and is discussed in the following section.

The Lichtenstein tension-free hernioplasty technique consists of a swatch of polypropylene mesh 8 × 16 cm, partway slit 1 cm from its inferior edge to accommodate the spermatic cord and fashioned for the patient. The mesh is sutured circumferentially to the internal oblique abdominal muscle, the rectus sheath, and the shelving edge of the inguinal ligament. The dissection includes a search for femoral hernias and division of the cremaster muscle, cremaster vessels, and the genital nerve to thin the cord. The tails of the mesh are drawn equally around the cord and the inferior edge of each tail is sutured to the inguinal ligament. For a successful hernioplasty the mesh must be as wide as possible, overlap the rectus sheath by more than 1 cm, extend laterally beyond the shutter mechanism, and snugly embrace the cord.

The Lichtenstein plug is cylindrical and made from a 2- to 2.5-cm strip of polypropylene mesh. For successful use of the plug the defect should be sclerotic and the plug must be wound firmly, fit snugly, and be sutured flush in place with four or more permanent synthetic sutures. The technique is especially useful in the elderly and the infirm with a single parietal defect. It is easily done with unassisted local anesthesia through a small incision directly over the defect using minimal dissection and without disturbing the spermatic cord.

Gilbert serendipitously discovered that a cone-shaped plug fashioned from a piece of polypropylene mesh approximately 6× 6 cm, when passed through the deep ring, would retain inverted indirect inguinal hernial sacs and could not be expelled on repeated Valsalva maneuvers (Fig. 34-13). Gilbert placed the cone plug through the deep ring and into the properitoneal space, where, he surmised, it unfolded like an umbrella to occlude the deep ring. The unfixed plug was reinforced with a patch similar in size and shape to that of the Lichtenstein repair but without fixation. He called the procedure the “sutureless hernioplasty” and used it with great success for indirect hernias without defective inguinal floors. Rutkow and Robbins expanded Gilbert's indications and used cone-shaped plugs in both indirect and direct inguinal hernias of all sizes. They used a patented and costly commercial plug that has no advantages over the handmade cone-shaped plugs. The cone plug is also an effective barrier when merely placed without fixation within the deep ring, and unfolding is not apparent. Indirect hernial sacs are dealt with by conventional techniques, and direct inguinal hernial sacs are merely inverted beneath absorbable sutures. Plugs positioned to invert a diffuse direct inguinal hernial sac as used by Rutkow and Robbins are ineffective because they are expellable without a covering barrier of mesh.

In the tension-free hernioplasties the ilioinguinal nerve and the genital branch of the genitofemoral nerve, when not divided, are allowed to pass with the cord through the newly fashioned deep ring in the prosthesis.

The tension-free hernioplasties are tolerant of technical leeway. For example, the mesh in the Lichtenstein tension-free hernioplasty need not always be sutured to the internal oblique abdominal muscle or even to the inguinal ligament. The tails of the mesh surrounding the cord need not always be overlapped to create a secure newly constructed deep ring, especially when cone plugs are used to retain indirect hernial sacs. In general, tension-free hernioplasties should be tailored to the problem at hand. The tension-free hernioplasty most commonly used in men by the author is Gilbert's sutureless hernioplasty, to which is added a medial point of fixation with a suture on the rectus sheath and a lateral suture to close the tails of the mesh around the cord (Fig. 34-14). The mesh is positioned to allow the inferior edge to roll on the inguinal ligament. In all instances the dissection is minimal and the cremaster muscle and neurovascular bundle are not divided unless the cord is bulky.

The tension-free hernioplasties are contraindicated for hernias resulting from a connective tissue disorder. Also they usually are not needed in women with primary indirect inguinal hernias because simple obliteration of the deep ring always produces excellent results. The tension-free hernioplasties may not be appropriate for repair of most recurrent groin hernias in men because they necessarily require redissection and remobilization of the spermatic cord, paramount causes of testicular atrophy. Stoppa's GPRVS would be a better choice because it is a posterior hernioplasty providing direct access to the parietal defects, obviating dissection of the spermatic cord and sensory nerves of the groin, and dealing more efficiently with the complex and diverse nature of recurrent hernias.

The tension-free hernioplasties have been highly touted and aggressively promoted by some of the fabricators of the prostheses and initiators of the procedures. Widespread enthusiasm for them has developed because they are easy to perform even with unassisted local anesthesia, recovery is quick, and results, thus far, are superb. Tension-free hernioplasties are especially suited to the management of simple primary hernias in men but are not the procedures of choice for complex groin hernias or for those with complications.

Posterior (Properitoneal) Prosthetic Groin Hernioplasty

Stoppa Procedure—Giant Prosthetic Reinforcement of the Visceral Sac

The properitoneal space is the logical site to implant a prosthesis. The prosthesis is held in place by intraabdominal pressure, and it is remote from the relatively avascular subcutaneous tissues and therefore relatively immune to superficial infection. The myopectineal orifice can be patched or plugged and hernioplasties buttressed with a prosthesis from the posterior approach just as they can be from the anterior (Fig. 34-15). The innovative and revolutionary properitoneal prosthetic technique introduced by Stoppa in 1969 is unique. He proposed eliminating hernias of the groin with a large unresorbable prosthesis that functionally replaces the transversalis fascia. The prosthesis adheres to the visceral sac and renders the peritoneum inextensible so that peritoneum cannot protrude through the myopectineal orifice or adjacent areas of weakness; repair of the defect in the abdominal wall is unnecessary. The operation is technically known by the descriptive phrase giant prosthetic reinforcement of the visceral sac (GPRVS) but is commonly called the Stoppa procedure.

GPRVS is performed bilaterally or unilaterally. In bilateral GPRVS a single large prosthesis is inserted in the properitoneal space of both groins through a midline or a Pfannenstiel incision (Fig. 34-16). The procedure requires general or regional anesthesia and is performed in ambulatory settings. The operation is the procedure of choice for all complex hernias of the groin such as recurrent hernias and hernias associated with connective tissue disorders. It is quick and exceptionally easy to perform, even in the obese, for whom it may be the preferred hernioplasty when they have bilateral hernias. It is the best hernioplasty for patients with primary hernias who prefer or require regional or general anesthesia.

In unilateral GPRVS a large prosthesis is inserted in the properitoneal space of a single groin. Usually it is performed through a transverse abdominal incision, but it also can be performed through an anterior groin incision, both transinguinally or subinguinally (Fig. 34-17). The abdominal incision avoids the dissection of the inguinal canal and is preferred for men with recurrent inguinal hernias after classical repair. Transinguinal GPRVS is indicated when an unanticipated complex groin hernia is encountered, and subinguinal GPRVS is indicated when an unanticipated perivascular femoral hernia is encountered. All forms of unilateral GPRVS can be done in day surgery and with local anesthesia, with the possible exception of transabdominal GPRVS, which more often than not requires general or regional anesthesia.

Ordinarily, a large prosthesis in the properitoneal space would require a slit to accommodate the spermatic cord at the deep ring (see Fig. 34-15). The slit, sutured closed around the elements of the cord, partially holds the prosthesis in place but is also a potential site for peritoneal protrusion. However, a slit in the mesh becomes unnecessary if the elements of the spermatic cord—the vas deferens and the testicular vessels—are dissected from the peritoneum to which they are attached by the transversalis fascia. The elements of the cord can then lie against the wall of the pelvis and the properitoneal prosthesis will cover them, the deep ring, and the myopectineal orifice completely. This technique, also devised by Stoppa, facilitates properitoneal prosthetic implantation (see Fig. 34-16).

The prosthetic material of choice for GPRVS is Mersilene because it is supple and elastic, stays in place without sutures, conforms to the complex curvature of the pelvis, and induces prompt fibroblastic integration. Placement of the prosthesis in GPRVS is done with long abdominal clamps. Other prostheses have been used. However, if the semirigid polypropylene meshes are substituted for Mersilene, some fixation will be necessary because they do not conform well. Gore-Tex also requires permanent fixation because encapsulation is very slow.

GPRVS is an efficient, anatomic, and tension-free repair. It is the ultimate hernioplasty. When correctly done, it cures all hernias of the groin, even prevascular femoral hernias, with very rapid recovery and minimal discomfort.

Laparoscopic Repair

Laparoscopic hernioplasty has attracted much attention and deserves comment, even though only about 5 percent of groin hernias are repaired this way. There is no question that laparoscopic hernioplasty can be done successfully in experienced hands and that in some patients it produces less postoperative pain. The question is whether or not it should be done at all, because it is inherently riskier than open hernioplasty. Prospective randomized studies comparing open repair and laparoscopic repair on a nationwide scale are about to begin. Meanwhile, data from the most experienced laparoscopic hernia surgeons of Western Europe and North America have been published, and their morbidity, mortality, and recurrence rates have been compared with those of surgeons in Western Europe and North America who have specialized in open hernioplasty. The results showed that with the exception of wound infection, the morbidity, mortality, and recurrence rates after laparoscopic hernioplasty are significantly higher than those following open hernioplasty. *The degree of postoperative discomfort should never be the motivating factor in selecting the type of operation.* Rather, the procedure should be selected that will produce the best results with the least risks. To date, this is still open hernioplasty. Furthermore, laparoscopic hernioplasty always requires general or regional anesthesia and consumes more of the hospital's resources. Most important, unlike open hernioplasty, it requires a surgeon who is very experienced and highly skilled with the technique. Open hernioplasty, preferably tension-free hernioplasty, with local anesthesia when possible, remains the procedure of choice for the majority of patients.

Complications

Ischemic orchitis, with its sequela of testicular atrophy, and residual neuralgia are the two important, although uncommon, complications unique to groin hernioplasty. They occur more frequently after anterior groin hernioplasty because the nerves and spermatic cord are necessarily dissected and mobilized. Recurrence also may correctly be considered a complication of groin hernioplasty, though surgeons traditionally have not categorized it as such.

Ischemic Orchitis and Testicular Atrophy

The clinical manifestations of ischemic orchitis develop insidiously, do not become apparent for 2 to 5 days after the hernioplasty, and are frequently misinterpreted initially. The testicle and spermatic cord become swollen, hard, tender, painful, and retracted. The process lasts 6 to 12 weeks and may resolve completely or end in testicular atrophy. Gangrene is unusual and orchiectomy is rarely necessary. The return of the testicle to normal size and shape does not mean that the process is complete, and atrophy of the testicle may not become apparent for as long as a year.

The etiology of ischemic orchitis is thrombosis of the spermatic cord, and the testicular pathology is intense venous congestion. The thrombosis is induced by surgical trauma to the cord, especially that associated with the dissection to completely remove a large indirect hernial sac. Dissection of a scrotal indirect hernial sac damages the delicate veins of the pampiniform plexus, initiates the thrombosis, and coincidentally disrupts collateral circulation (see section on dissection of classical repair).

There is no known successful treatment to prevent atrophy. Antibiotics, anti-inflammatory drugs, and massive doses of steroids are often given. Fortunately the incidence can be minimized by reducing surgical trauma to the cord by never excising the distal part of an indirect sac except when unavoidable; never dissecting beyond the pubic tubercle; and never redissecting an inguinal canal and spermatic cord in a patient who is predisposed to the complication by a previous anterior hernioplasty, vasectomy, hydrocelectomy, or other groin or scrotal surgery. In the latter situations posterior properitoneal prosthetic hernioplasty is the preferred alternative.

Neuralgia

Chronic residual neuralgia may result from surgical handling of the sensory nerves in the groin during hernioplasty or after hernioplasty from constricting scar tissue or adjacent inflammatory granulomas. The intraoperative trauma to the nerve, whether division, stretching, contusion, or suture entrapment, usually is not appreciated at the time. The pain may be localized, diffuse, projected along the course of a nerve, or referred to a nearby site. It may develop early or late, possibly weeks to months after the operation. In most cases the pain is accompanied by changes in mood and behavior. Depression, disturbed interpersonal relationships, and inability to resume work are common and complicate management. A multidisciplinary approach is beneficial.

A well-known cause of residual neuralgia is a neuroma. A neuroma results from a proliferation of nerve fibers outside the neurilemma of a partially or completely divided nerve. The pain from a neuroma varies in intensity, can be induced by changes in position, and is without spontaneous paroxysmal exacerbations. Hyperesthesia is detectable in the area of the lesion, and tapping the site may produce a severe shooting pain. The pain from deafferentation is different from the pain of a neuroma. It is burning and permanent with intermittent paroxysmal exacerbations. The area of involvement is anesthetic at first but becomes hyperesthetic. Typically the onset of deafferentation pain is delayed a week or so, and tapping the site does not produce exquisite pain.

The management of residual neuralgia is difficult. Neurolysis of the involved nerve may afford relief and is more likely to be successful if instituted early. The involved nerve may be identified by local anesthetic nerve blocks. The iliohypogastric and ilioinguinal nerves can be blocked and also divided in the groin. Blocking the genitofemoral nerve with certainty is impossible in the groin. Neuralgia from this nerve is identified by blocking L1 and L2 paravertebrally. Division of the genitofemoral nerve is performed through a flank incision or, preferably, through a laparoscope. Other adjunctive therapy includes analgesics, antidepressants, anticonvulsants, anxiolytics, and transcutaneous electrical stimulation. Injections of steroids in the involved area are sometimes helpful.

The prevention of nerve injury is important since the treatment of neuralgic complications is often unsuccessful. Fortunately they are rare. The sensory nerves should be preserved if possible; this is not always possible, and some small, not readily visible nerves are invariably divided. Occasionally major branches of the sensory nerves or the nerve trunks themselves must be divided in order to accomplish the hernioplasty. Division (usually with ligation) of the genital branch of the genitofemoral nerve is routine in most anterior groin hernioplasties. This virtually never produces genital nerve neuralgia and may suggest that complete division of the nerve is less likely to cause residual neuralgia than other types of

nerve trauma. Most nerves require ligature to control bleeding. The ligature may also confine neuroma formation to within the neurilemma.

Recurrences

Very experienced hernia surgeons using classical repairs have been reported to obtain recurrence rates in the range of 1 to 3 percent in a 10- year follow-up. Such results are exceptional, and most surgeons are unable to duplicate them. Recurrences are caused by excessive tension on the repair, deficient tissues, inadequate hernioplasty, and overlooked hernias. Recurrences predictably are more common in patients with direct hernias, especially bilateral direct inguinal hernias, direct hernias presenting on both sides of the inferior epigastric vessels, and in direct hernias combined with an indirect hernia. Indirect recurrences result from insufficient excision of the proximal end of the sac, insufficient repair of the deep ring, and continued atrophy of the shutter mechanism. Most recurrences are direct and are usually in the region of the pubic tubercle, where suture line tension is the greatest. Relaxing incisions are always helpful. Repairing bilateral inguinal hernias simultaneously does not increase suture tension and is not a cause of recurrence, as was formerly believed.

Recurrent hernias require a prosthesis for successful repair. Recurrences after prosthetic patch and plug repairs occur in areas of the myopectineal orifice that are unprotected by the prosthesis, such as the deep ring. Paraprosthetic recurrences result from prostheses that are too small. Recurrences after anterior prosthetic hernioplasty are managed properitoneally with a second prosthesis or anteriorly with a prosthetic plug.

Recurrences after *GPRVS* occur if the properitoneal space is not cleaved widely enough or if the prosthesis is too small, poorly shaped, or incorrectly implanted. Recurrences after GPRVS usually are approached anteriorly and are managed by adding a prosthetic extension to the existing prosthesis. Alternatively, another permanent prothesis can be implanted transabdominally.

UMBILICAL HERNIA

The umbilicus is one of the weak areas of the abdomen and a common site of herniation. Umbilical hernias occur more frequently in women. Obesity and repeated pregnancies are common precursors, and ascites always exacerbates the problem. Umbilical hernia in adults is acquired and has no relationship to umbilical hernia in children. Strangulation of the colon and omentum is fairly common. Rupture occurs in chronic ascitic cirrhosis, in which case an emergency portal decompression is necessary.

Umbilical hernias are common in infants and close spontaneously if the aponeurotic defect is 1.5 cm or less. Repair is indicated in infants with hernial defects greater than 2 cm in diameter and in all children with umbilical hernia still present by the age of three or four.

The classical repair for umbilical hernias is the Mayo hernioplasty. The operation consists of vest-over-pants imbrication of the superior and inferior aponeurotic segments (Fig. 34-18). Currently, the Mayo repair, although still useful, is infrequently employed. Instead, umbilical hernias with a small parietal defect are merely closed by a to-and-fro, loosely placed polypropylene suture, and those with large parietal defects are managed with a prosthesis in a repair resembling that for incisional hernia.

EPIGASTRIC HERNIA

Epigastric hernia is a protrusion of properitoneal fat and peritoneum through the decussating fibers of the rectus sheath in the midline (linea alba) between the xiphoid process and the umbilicus. Epigastric hernias often are irreducible, invariably have small aponeurotic defects, sometimes are multiple, and often produce discomfort out of proportion to their size. Repair, usually through a vertical skin incision, is easy and similar to that of umbilical hernias.

SPIGELIAN HERNIA

Spigelian hernias are ventral hernias occurring along the subumbilical portion of Spieghel's semilunar line and through Spieghel's fascia. Spigelian hernias are rare and, unless large, are difficult to diagnose because they are interparietal and contained by the aponeurosis of the external oblique muscle. Ultrasonography and computerized tomography often reveal symptomatic spigelian hernias too small to detect clinically. Large spigelian hernias may be mistaken for sarcomas of the abdominal wall. Entrapped anterior cutaneous nerves of T10 to T12 produce discomfort resembling spigelian herniation.

Spieghel's fascia is actually aponeurotic and consists of the fused aponeuroses of the internal oblique and the transverse abdominal muscles between the belly of these muscles laterally and the rectus muscle medially (Fig. 34-19). Above the umbilicus the aponeurotic fibers of these muscles crisscross and form a fairly strong barrier. Below the umbilicus the fibers are more or less parallel and may split, permitting the peritoneum and properitoneal fat to protrude through a slitlike defect but to be retained by the overlying aponeurosis of the external oblique abdominal muscle. Although spigelian hernias occur anywhere along the semilunar line, they occur most commonly where Spieghel's fascia is widest and weakest. Spieghel's fascia is widest in the area between the umbilicus and the line connecting it to the anterior superior iliac spine, and it is weakest in the area just beneath the arcuate line and above the inferior epigastric vessels. Spigelian hernias may occur inferior to the epigastric vessels, in which case they are a variant of a direct inguinal hernia.

The neck of a spigelian hernia enlarges laterally by spreading apart the broad muscles of the abdomen. The rectus muscle and sheath inhibit medial enlargement. Small spigelian hernias are simply closed, but large spigelian hernias that are in the muscles require a prosthesis.

LUMBAR HERNIA

Congenital, spontaneous, and traumatic herniations—all rare—occur through Grynfeltt's superior and Petit's inferior lumbar triangles. They are impossible to repair successfully without a prosthesis or a myoaponeurotic flap. Petit's triangle is an upright triangle bounded by the latissimus dorsi muscle, the external oblique abdominal muscle, and the iliac crest. It is covered by the superficial fascia only. Grynfeltt's triangle is an inverted triangle bounded by the twelfth rib, the internal oblique abdominal muscle, and the sacrospinalis muscle. It is covered by the latissimus dorsi.

The large, diffuse lumbar hernias occurring after kidney excision result in part from muscular paralysis, and aponeurotic defects are usually unidentifiable. They are managed as incisional hernias.

PELVIC HERNIA

Hernias occur in the obturator fossa, in the greater and lesser sciatic foramina, and in the perineum. All are rare and occur mainly in cachectic elderly patients, especially women. Of the hernias in the pelvis, the obturator hernia is most common, and when encountered it is almost always strangulated. Pressure on the obturator nerve causes pain in the region of the hip and of the knee and on the inner aspect of the thigh. This characteristic pain occurs in about half of the patients and is known as the Howship-Romberg sign. A palpable mass on pelvic or rectal examination or in the upper medial part of the thigh is occasionally noted. Prosthetic repair is normally necessary and is preferred unless septic conditions are present.

Perineal hernias occur through the pelvic diaphragm and may be anterior or posterior to the superficial transverse perineal muscle. The anterior perineal hernia is seen only in women and passes into the labia major, while the posterior perineal hernia enters the ischiorectal fossa in men and close to the vagina in women.

PARASTOMAL HERNIA

Parastomal hernias interfere with colostomy irrigations and the adhesion of stomal appliances. Paracolostomy hernias are more common than paraileostomy hernias, and both are more likely to occur when the stoma emerges through the semilunar line rather than the rectus sheath. Parastomal hernias therefore are usually lateral to the ostomy. Moving the stoma to a new location is the traditional method of management and is preferred to local repair. Local repair often fails because the belt muscles lateral to the ostomy lack sufficient aponeurosis. Among the prosthetic repairs, the preferred technique is the Leslie procedure (Fig. 34-20). This method is easy to perform, is nondebilitating, and for the author has essentially replaced transplantation of the stoma. It consists of closing the parietal defect lateral to the stoma in the subcutaneous tissues and then implanting a large piece of Mersilene that has been partially slit to accommodate the stoma and then spread out onto the anterior abdominal wall, to which it is fixed. Broad contact between the mesh and bowel is to be avoided. Mersilene is preferred to polypropylene meshes because it is soft, supple, and conforming. No sutures are placed in the intestinal wall.

INCISIONAL HERNIA

Incisional hernias are a serious surgical problem. They have a propensity to enlarge, are frequently formidable to repair, and are usually accompanied by serious associated conditions. Patients with incisional hernias are usually obese. Obesity and infection are the two principal causes of this condition. The weight of the panniculus adiposus itself pulls apart the surgical incision, and infection hampers wound healing. Hypertension, cardiac and renal disorders, diabetes, and purulent intertrigo commonly accompany obesity and increase the morbidity of repair.

Eventration Disease

The loss of integrity of the abdominal wall reduces intraabdominal pressure and causes serious disturbances, which Rives appropriately named "eventration disease." The salient feature of the syndrome is respiratory dysfunction. A large incisional hernia produces paradoxical respiratory abdominal motion similar to that of a flail chest. Diaphragmatic function becomes inefficient. The diaphragm no longer contracts against the abdominal viscera and instead forces them into the hernial sac. Appraisal of respiratory function and blood gases is essential.

The detachment of the tendinous insertion of the broad belt muscles of the abdomen aggravates midline incisional hernias. The muscles retract, pull apart the parietal defect, and cause the normally horizontal belly of the rectus muscle to assume a vertical position. Contraction of the rectus muscles then expels rather than retains the abdominal viscera. Ultimately atrophy, fatty degeneration, and fibrosis of the lateral muscles ensue and make tendinous reinsertion of the belt muscle by the midline approximation of the linea alba difficult. In some midline hernias real loss of the abdominal wall may be present as a result of infection, trauma, or repeated laparotomies. In most cases, however, the loss of substance is more apparent than real.

The viscera lose their right of domain in the abdomen in long-standing large incisional hernias. In this instance the reduction of the viscera at operation can cause death by compression of the inferior vena cava and by respiratory failure from forced elevation and immobilization of the diaphragm. The introduction of pneumoperitoneum by Goni-Moreno in 1947 made these formerly inoperable hernias reparable.

Reduced intraabdominal pressure from the hernia also causes edema of the mesentery and stasis in the splanchnic venous system and the inferior vena cava. Distention and atony of the hollow viscera occur, and decreasing the ability to raise intraabdominal pressure and producing difficulties with micturition and bowel movements. Back pain is a common

complaint; it is caused by lordosis resulting from retraction of the belt muscles and decreased efficiency of the rectus muscles.

The skin and subcutaneous tissues overlying large incisional hernias are stretched and damaged. The skin becomes atrophic, hypoxic, and devoid of subcutaneous fat. Spontaneous ulcerations develop. Typically they are solitary, occur at the apex of the hernia, and are misconstrued as a pressure sore. They resist healing and require intensive local and systemic antimicrobial therapy to prevent septic complications at the time of hernioplasty.

Obese patients with large incisional hernias are especially at risk for postoperative septic complications, respiratory dysfunction, and pulmonary emboli. Preoperative and postoperative prophylaxis for these problems is essential.

Progressive pneumoperitoneum is a useful technique to prepare patients for incisional hernioplasty because it overcomes some of the disorders of eventration disease. Pneumoperitoneum stretches the abdominal wall and intraabdominal adhesions, facilitates the return of the viscera to the abdomen, and improves diaphragmatic function. The technique of pneumoperitoneum is simple. Air is injected into the peritoneal cavity through a pneumoperitoneum needle that has been inserted with the aid of local anesthesia. The peritoneal cavity is insufflated until the patient experiences shortness of breath or shoulder pain. At first the patient may tolerate only small amounts of air. Sometimes as much as 2 to 4 L may be insufflated initially. Thereafter, air is added as needed at intervals of 1 to 3 days. The pneumoperitoneum is maintained for 10 to 20 days. The patient is ready for operation when palpation reveals flabbiness in the flanks. The inability of the patient to tolerate pneumoperitoneum is a contraindication for incisional hernioplasty.

Incisional Hernioplasty

The object of incisional hernioplasty is anatomic reconstruction of the abdominal wall. This includes closure of the parietal defect, restoration of normal intraabdominal pressure, and, in midline hernias, tendinous reinsertion of the lateral abdominal muscles. Most small incisional hernias are managed by simple closure of the aponeurotic defect. However, large incisional hernias with aponeurotic defects greater than 10 cm have a marked tendency to recur. Repairs performed by approximating aponeurosis edge to edge or by the Judd or Mayo technique of imbrication frequently fail and have recurrence rates as high as 50 percent. Consequently most incisional hernias and all recurrent incisional hernias require a prosthesis for a successful repair. Of the many proposed techniques, those developed by Rives and by Stoppa, essentially identical procedures, are preferred. The Rives-Stoppa technique is applicable to all types of abdominal incisional hernias, including postnephrectomy lumbar hernias and parastomal hernias.

In the Rives-Stoppa hernioplasty a very large Mersilene prosthesis is implanted deep to the muscles of the abdominal wall on top of the posterior rectus sheath or peritoneum (Figs. [34-21](#) and [34-22](#)). The prosthesis extends far beyond the borders of the myoaponeurotic defects and is firmly held in place by intraabdominal pressure (Pascal's principle) and later by fibrous ingrowth. The prosthesis prevents peritoneal eventration in two ways: by rendering the visceral sac indistensible and by solidly uniting and consolidating the abdominal wall. The technique thus uses the same concept and principles as Stoppa's properitoneal inguinal hernioplasty and, as noted earlier, is known technically as giant prosthetic reinforcement of the visceral sac (GPRVS).

The knitted elastic Mersilene prosthesis is arranged so that it expands vertically and is inextensible horizontally. The mesh must extend beyond the lateral borders of the aponeurotic defect by 8 to 10 cm. The distance inferiorly and superiorly is less critical, and 4 to 5 cm suffices. Above the umbilicus, the linea alba interferes with the midline placement of the prosthesis, and a slit in the middle of the superior edge of the prosthesis is necessary to allow the prosthesis to protrude upward within the rectus sheath on either side of the linea alba. Traction-fixation sutures proposed by Rives, placed 5 to 6 cm apart, fix the prosthesis circumferentially to the abdominal wall (Fig. [34-23](#)). The lateral traction-fixation sutures also stretch the retracted belt muscles and facilitate midline approximation. Fixation sutures are not needed when the prosthesis extends deep into the space of Retzius or far into the iliac fossa. When the retromuscular space is undissectable, the prosthesis must be implanted intraperitoneally. The intraperitoneal prosthesis must be prevented from touching the viscera either by the omentum or by an absorbable synthetic prosthesis. An absorbable prosthesis is used to substitute for the posterior rectus sheath when it is insufficient.

Aponeurotic closure of the parietal defect is important. The midline closure can withstand greater tension because the prosthesis, not the suture line, ultimately unites the abdomen. When necessary, tension can be reduced by 1-cm vertical relaxing incisions arranged quincuncially in the rectus sheath. Each row of overlapping relaxing incisions expands the rectus sheath by 1cm. Long relaxing incisions are not recommended above the umbilicus and should *never* be made below the umbilicus. Aponeurotic approximation is usually achievable, but when it is not, a second absorbable or nonabsorbable prosthesis inlaid in the aponeurotic defect will ensure stability of the abdominal wall during the healing process. Usually this occurs in the region of the xiphoid process or symphysis pubis. Dead space created by a large prosthesis always requires closed suction drainage to prevent seromas and hematomas and to allow quick fibrous incorporation of the prosthesis in the abdominal wall. The drains on top of the prosthesis are not removed until the drainage is essentially nil.

Infection

Infection is a serious complication and occurs in as many as 10 percent of the cases. For this reason major prosthetic incisional hernioplasties are considered "dirty cases". Management of wound infections involving a synthetic nonabsorbable prosthesis is addressed earlier in the subsection .

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 35 - Pituitary and Adrenal

William T. Couldwell
Marie F. Simard
Martin H. Weiss
Jeffrey A. Norton

HISTORICAL BACKGROUND

Although Marie correctly ascribed the clinical syndrome of acromegaly to the presence of a pituitary adenoma as early as 1889, it was not until 1893 that Caton and Paul recorded the first attempt at surgical resection of a pituitary tumor. Their two-stage lateral subtemporal procedure, as suggested by Sir Victor Horsley, was unsuccessful, because the patient died before the operation was completed. Subsequently, Horsley himself, using a lateral middle fossa approach, reported successful surgical resections in eight of ten pituitary tumors operated on between 1904 and 1906. Krause, in 1905, introduced the frontal transcranial approach to the sella turcica; this technique then provided the basis from which the majority of subsequent variations of transcranial approaches were developed.

Kiliani elaborated on Krause's technique with a bifrontal intradural approach to the sella turcica in the hopes that this procedure would offer visualization that would enable a radical intradural resection of all tumor. McArthur, in the first proposal for a predominantly extradural approach to the sella, advocated resection of the supraorbital ridge and the orbital plate to allow dissection to extend posteriorly to the level of the optic chiasm, where subsequent transverse incision of the dura provided access to the intradural compartment.

These basic approaches were modified and improved by a number of the giants of neurosurgical development in the early part of the twentieth century. Modifications described by Frazier, Dandy, Heuer, and Cushing primarily favored the direct intradural approach to the sella turcica and its contents. With the profound influence of these neurosurgeons, by the 1930s most neurosurgical approaches to the pituitary used a transcranial intradural technique, though some neurosurgeons continued to use an extradural approach. Heuer advocated a lateral anterior fossa approach using the landmarks provided by the sphenoid ridge. This approach seemed to be the shortest distance between the inner table of the skull and the sella turcica. Frazier described a frontal transcranial approach beginning at the midpoint between the midline and the lateral sphenoid wing. Cushing preferred a direct midline approach. The dominant neurosurgical teaching during the 1930s and 1940s continued to focus on a transcranial approach to the pituitary. During the 1950s Olivecrona and Ray, major figures in international neurosurgery, promulgated the use of a transfrontal intradural approach to the pituitary while developing extensive experience with hypophyseal ablation, which had become a major adjunct in the management of metastatic breast and prostate carcinoma as well as diabetic retinopathy. Ray, who performed more than 1,000 procedures on the pituitary, emphasized the need for a properly placed low bone flap in order to minimize brain retraction in gaining access to the sella and parasellar areas.

The earliest attempts at transcranial approaches to the pituitary at the turn of the century and shortly thereafter, however, resulted in a mortality rate that was generally considered prohibitive. The 20 percent mortality rate in Horsley's small series of 10 cases referred to above would certainly be unacceptable by modern standards; yet it was significantly better than the experiences of his colleagues, who reported mortality rates ranging from 50 to 80 percent. It should be recognized that this extraordinary incidence of mortality was generally accepted in the efforts at developing intracranial surgery at that time.

As a consequence of these earlier experiences, attempts were made continually to modify surgical techniques as well as adjuvant management protocols in an effort to make the surgical approach to the pituitary a more reasonable option for both surgeon and patient. In this light, Schloffer, a rhinologist from Innsbruck, Austria, recommended the use of a transsphenoidal route as an alternative and presumably safer approach to the sella turcica and its contents. In 1907 he reported the first successful removal of a pituitary tumor using the transsphenoidal approach. The technique of sphenoid sinus exposure subsequently underwent a number of modifications by interested surgeons, the culmination of which was Halsted's description in 1910 of the sublabial gingival incision for the initial stage of sphenoid sinus exposure. After initial disappointments at transcranial efforts, Cushing himself embraced the transsphenoidal approach. He described a technique that combined a number of suggestions that had been made by previous authors and used the sublabial incision described by Halsted (Fig. 35-1). Cushing also adopted the technique of submucous dissection of the nasal septum promulgated by Eiselberg and Kocher and used the headlight described by Kanavel to obtain better visualization into the depths of the operative field during operation. Using the transnasal/transsphenoidal approach during the 15-year period from 1910 to 1925, Cushing operated on some 231 pituitary tumors, with a reported mortality rate of 5.6 percent. This represented the best efforts at surgical exposure and access to the pituitary at the time.

Mortality and morbidity from the use of the transsphenoidal approach were primarily the result of infection, which was frequently associated with postoperative cerebral spinal fluid (CSF) rhinorrhea; hemorrhage and postoperative edema, in contrast, were the causes of the excessive mortality and morbidity associated with transcranial approaches. As Cushing developed expertise and confidence in transcranial surgery, he began again to use the transcranial approach to the pituitary. As he gained experience, he finally reduced his mortality rate with the transcranial approach to 4.5 percent, essentially eliminating any significant difference between the transsphenoidal and transcranial approaches. It was his opinion that visual recovery was more complete after direct decompression of the optic nerves and chiasm, which may well have been the case in those early times. In addition, Cushing recognized that the transfrontal approach allowed more extensive resection of those suprasellar tumors that were complicated by significant lateral extension. Obviously Cushing's intense interest in intracranial surgery must have contributed to his pursuit and development of transcranial approaches to the pituitary. Other distinguished surgeons, however, such as Hirsch and Hamlin, continued to use the transsphenoidal approach, with reports of excellent results. It is reported that Cushing, on the occasion of his seventieth birthday, remarked that the transsphenoidal approach, because of its many advantages, would probably return despite the liabilities that he perceived.

Because of Cushing's dominance in the evolution of neurosurgery in the United States and the outstanding results that he reported, the use of transsphenoidal operations on the pituitary diminished profoundly during this period, until Norman Dott of Edinburgh, a student of Cushing's, refocused attention on this technique. In turn, Dott reportedly introduced Guiot of France to the transsphenoidal approach, and Guiot subsequently influenced the extensive development of this technique by Jules Hardy of Montreal. Having accumulated extensive experience in transsphenoidal microsurgery, these individuals significantly altered and redefined the indications and risks related to the transsphenoidal resection of hypophyseal and parhypophyseal tumors. Hardy propounded the use of intraoperative televised fluoroscopy, the operating microscope that provided both intense focused illumination and magnification, and the design and development of microsurgical instrumentation for work within and about the sella. In addition, refinement of the techniques for closure of the opening between the sphenoid sinus and the intracranial compartment resulted in a profound decrease in the mortality and morbidity associated with the procedure.

In 1908 Horsley, working in conjunction with an engineer, Richard Clark, first proposed a system for stereotactic localization within the brain using a rigid reference frame fixed to the skull. In 1921 Hirsch suggested using this stereotactic technique to approach the sella turcica; unfortunately, this proposal did not receive general acceptance in the neurosurgical community. During the 1950s Talairach and Tournoux, working in France, and Forrest, working in Scotland, reintroduced what they believed was a significant advancement in the stereotactic approach to destruction of the normal pituitary gland. During this time, hypophysectomy for the treatment of the metastatic diseases was a frequently used procedure, and efforts to avoid more extensive surgery were frequently undertaken. Talairach, Tournoux, and Forrest noted that it was possible to advance a straight cannula through the nasal cavity directly to the confines of the sella without encountering any barrier except the floor of the wall of the sphenoid sinus in most patients. Under x-ray visualization, this cannula could subsequently be guided through the anterior wall of the sella, enabling delivery of various destructive agents directly into the gland. Such agents included radioactive materials such as yttrium-90, liquid nitrogen, or even radiofrequency current. These investigators ushered in an active era of stereotaxic pituitary surgery, and some efforts were subsequently made for treatment of pituitary tumors by these techniques. Unfortunately, initial attempts at total hypophysectomy resulted in unacceptable variability in the completeness of hypophysectomy, frequent damage to the oculomotor and optic nerves, and an unacceptably high incidence of CSF rhinorrhea and consequent meningitis. Although these techniques did not find an extensive reception for use in the management of pituitary tumors, refinement in the use of the technique of radiofrequency ablation enabled Zervas to report in 1969 no occurrence of rhinorrhea, meningitis, or local infection in any of 91 consecutive patients who had undergone stereotactic radiofrequency ablation of the normal and abnormal pituitary gland. Subsequent loss of interest in ablation of the normal gland for the treatment of metastatic carcinoma or diabetic retinopathy, however, resulted in a focused effort in transnasal transsphenoidal approach to the pituitary, which today is clearly established as the foremost approach to the sella and selected parasellar lesions.

Radical operative procedures involving the hypophysis carried a prohibitive complication rate before the development of commercially available corticosteroids in the early 1950s. Replacement of these agents in patients who are rendered hypopituitary by the development of a tumor or other mass or as a consequence of surgical extraction of such lesions is essential for successful surgical intercession; commercial availability of the corticosteroids has truly revolutionized postoperative morbidity and mortality in these efforts.

Finally, our understanding of and the diagnosis of pituitary disorders have been revolutionized by two technologies of relatively recent origin—radioimmunoassay techniques, which allow diagnosticians to determine the presence of a number of functional endocrinopathies related to pituitary problems, and high-resolution computed tomography (CT) and magnetic resonance image (MRI) scanning, which allow sophisticated imaging of the pituitary gland itself. These developments have enabled precise definition of functional endocrinopathies, which in turn can be related to visualized structural lesions in and about the pituitary. These technologies, acknowledged by the award of two Nobel Prizes, have contributed immeasurably to the emergence of pituitary surgery as one of the preeminent surgical ventures in sophisticated modern neurosurgery.

ANATOMY

Macroscopic Anatomy

From a surgical standpoint it is crucial to consider the pituitary gland as one component in a complex anatomic region replete with important neural and vascular structures. The pituitary is surrounded by the hypophyseal fossa of the sphenoid bone, known as the *sella turcica*, and is covered superiorly by meningeal tissue forming the *diaphragma sellae*. The pituitary stalk passes through the center of the sellar diaphragm; in some cases the arachnoid also may protrude through this aperture into the sella turcica. The optic chiasm lies superior and anterior to the pituitary stalk and is contiguous anteriorly with the optic nerves and posteriorly with the optic tracts. The hypothalamus, which forms the anterior and lateral walls of the third ventricle, extends from the optic chiasm anteriorly to the mammillary bodies posteriorly over a distance of approximately 10 mm. The cavernous sinuses, which anatomically are more correctly a pair of venous plexuses, border the lateral walls of the sella turcica and extend posteriorly into the petrosal sinuses. Within each cavernous sinus lies the carotid artery, the oculomotor (III), trochlear (IV), and abducens (VI) nerves, and the ophthalmic and maxillary divisions of the trigeminal nerve (V) (Fig. 35-2).

The pituitary gland itself is derived embryonically from the ectoderm. In the normal adult it measures approximately 12 mm transversely, 9 mm anteroposteriorly, and 6 mm in height. In the adult male, the gland weighs about 500 mg, and in the adult female about 600 mg. Through the course of pregnancy and during normal puberty, the gland enlarges.

The pituitary gland is composed of the anterior pituitary or *adenohypophysis*, which accounts for approximately 80 percent of the hypophyseal volume, and of the posterior pituitary or *neurohypophysis*. The adenohypophysis is divided into three regions (Fig. 35-3): (1) the *pars distalis*, which comprises the major portion of the anterior lobe; (2) the *pars intermedia*, which in the human adult is a rudimentary structure composed of colloid-filled cysts located between the anterior and posterior lobes; and (3) the *pars tuberalis*, which is an extension of the pars distalis along the outer aspect of the pituitary stalk. The neurohypophysis is also divided into three regions: (1) the *median eminence*, situated in the hypothalamus, posterior to the optic chiasm, which forms the floor of the third ventricle and is also called the infundibulum; (2) the *infundibular stem*, which is the neural portion of the pituitary stalk; and (3) the *neural lobe* or inferior portion of the neurohypophysis, also called the *infundibular process*.

Microscopic Anatomy

Vascular Anatomy

The pituitary gland receives its blood supply from the bilateral superior and inferior hypophyseal arteries, which arise from the intracranial internal carotid arteries. The anterior and posterior branches of the superior hypophyseal arteries irrigate the median eminence and the infundibular stem. Within these structures, they divide to form vascular tufts, which comprise the primary vascular plexus of the hypophyseal portal system. This capillary system drains into the long hypophyseal portal veins, which course down the pituitary stalk anteriorly to supply blood to the anterolateral portion of the adenohypophysis (Fig. 35-4). The posterior adenohypophysis is supplied by the short portal veins originating from the lower infundibular stem. The anterior pituitary has the highest blood flow of any organ in the body (0.8 mL/g/min). The long portal veins supply 70 to 90 percent of its blood flow, and the short portal vessels the remainder. Within the pars distalis, these portal veins join into the secondary capillary plexus of the hypophyseal portal system. This venous plexus drains into the dural sinuses surrounding the pituitary. The hypophysiotropic hormones originating from the hypothalamus flow down the nerve axons and are liberated in proximity to the fenestrated capillaries of the primary hypophyseal portal plexus into which they diffuse. The stimulatory and inhibitory hormones are then transported down the portal veins toward the anterior pituitary. Thus the anterior pituitary gland proper receives blood supply from two portal systems, with no certain direct arterial blood supply; this likely results in less tissue perfusion pressure and an increase in incidence of ischemic loss of function with intrasellar masses. Injury or occlusion of the hypophyseal portal veins can lead to infarction and necrosis of the anterior lobe. The posterior pituitary is supplied by the inferior hypophyseal arteries directly, and hormones are liberated into the surrounding dural sinuses.

Adenohypophysis

The anterior pituitary contains connective tissue, fenestrated capillaries, and epithelial cells organized in a glandular pattern. The distinct epithelial cell types can be characterized by their secretory protein hormones as follows:

(1) *Growth hormone (GH)* -producing cells. Somatotropes comprise approximately 50 percent of the adenohypophyseal cell population and are located in the lateral wings of the anterior lobe. GH is a 191-amino acid polypeptide hormone that opposes the effect of insulin, stimulates the uptake of amino acids, and causes a release of free fatty acids from tissue storage sites. In the liver and other tissues, GH also mediates the synthesis of insulinlike growth factors (IGFs). IGFs induce protein synthesis in the skeleton and in muscle, and glucose oxidation in adipose tissue; they also stimulate cell replication. The secretion of GH is stimulated by *growth hormone-releasing hormone (GHRH)* and inhibited by somatostatin. GH and IGF-1 stimulate the release of somatostatin, thereby down-regulating the secretion of GH. GH is secreted in episodic surges occurring every 3 to 4 h. In young people the greatest peaks occur after the onset of deep sleep. Stimuli of GH secretion include insulin-induced hypoglycemia, arginine, exercise, l-dopa, clonidine, propranolol, and GHRH.

(2) *Prolactin (PRL)* -producing cells. The mammotropes, also called lactotropes, represent 15 to 25 percent of the anterior pituitary cells and populate the lateral gland. These cells accumulate during pregnancy and lactation and following estrogen therapy. Prolactin is a 198-amino acid polypeptide known to facilitate the development of breast tissue to ensure the production of milk. Prolactin secretion is stimulated by *thyrotropin-releasing hormone (TRH)*, estrogens, stress, and exercise. Dopamine is acknowledged to be the principal *prolactin inhibitory factor (PIF)*.

(3) Adrenocorticotrophic hormone (ACTH)-producing cells. Corticotropes constitute approximately 20 percent of the adenohypophyseal cells. They lie within the mediolateral aspects of the pars distalis. ACTH is a 39-amino acid peptide that promotes growth of the adrenal cortex and the synthesis of hormones produced by this gland. ACTH also has melanotropic effects and is responsible for the pigmentation commonly seen in Nelson's syndrome and Addison's disease. ACTH is a fragment of pro-opiocortin. Both the corticotropic and the melanotropic cells cleave the prohormone into a common precursor containing 130 amino acids (Fig. 35-5). Within the corticotrope there is further cleavage to form ACTH (1-39 amino acids) and *b-lipotropin (b-LPH)* (1-91 amino acids). b-LPH is cleaved further to form g-LPH (1-58 amino acids) and b-endorphin (61-91 amino acids). Within the melanotrope ACTH is cleaved into alpha-melanocyte-stimulating hormone (a-MSH) (1-13 amino acids) and corticotropin-like intermediate peptide (CLIP). In human beings the physiologic role of the lipotropins, a-MSH, and b-endorphin is unclear. The synthesis and secretion of ACTH is stimulated by *corticotropin-releasing hormone (CRH)*. Vasopressin, though a weak ACTH stimulant, directly potentiates the effect of CRH. The inhibition of ACTH release is regulated by the negative-feedback effect of cortisol on the corticotrope directly and on the release of CRH at the hypothalamus. ACTH also regulates the release of CRH through a short-loop negative-feedback effect. The secretion of ACTH follows a circadian rhythmicity and occurs in brief episodes during the late sleep period and just before awakening. Rapid release of ACTH also occurs under the stimulation of stress such as pain, fear, noise, extreme cold, fever, hypoglycemia, and hemorrhage.

(4) *Thyroxine-stimulating hormone (TSH)* -producing cells. Thyrotropes constitute approximately 5 percent of the adenohypophysis and are located in the anteromedial region. TSH is a glycoprotein composed of two noncovalently linked moieties, the alpha and the beta subunits. The beta subunit confers biologic activity to the hormone. Thyrotropes secrete not only TSH but also the separate subunits. TSH regulates the synthesis of thyroid hormones, thyroxine (T₄) and triiodothyronine (T₃), by the thyroid gland. A euthyroid state is maintained through a balance between the stimulatory effect of TRH produced by the hypothalamus and the negative-feedback inhibition of the thyroid hormones on the thyrotrope cells.

(5) Gonadotropic-hormone (FSH/LH)-producing cells. Gonadotropes secrete both *follicle-stimulating hormone (FSH)* and *luteinizing hormone (LH)* and occupy approximately 10 percent of the anterior gland within its medial region. FSH and LH are glycoproteins formed by two noncovalently linked subunits, the above-mentioned common alpha subunit and the specific beta subunit, which confers the biologic activity to the hormones. Gonadotropin levels are elevated during the first 1 to 2 years of life and rise again with the onset of puberty. FSH stimulates ovarian follicular growth in the female and testicular growth and spermatogenesis in the male. In the female, LH promotes ovulation and luteinization of the ovarian follicle and enhances estrogen and progesterone production by the ovary. In the male, LH supports interstitial (Leydig) cell function and hence the production of testosterone by the testicle. Gonadotropin secretion in the ovulating female is stimulated by *gonadotropin-releasing hormone (GnRH)* and both positively and negatively regulated by the gonadal steroids. Short-term effects of estrogens are suppressive on basal LH release. With longer duration of exposure to estrogens, particularly at low concentrations, gonadotropin secretion is enhanced. In males, testosterone has a negative-feedback effect on hypothalamic GnRH secretion and on the pituitary gonadotrope cells.

Neurohypophysis

The posterior pituitary lacks a blood-brain barrier and consists of hypothalamus-originating neuronal axons and terminals, specialized glial cells, and blood vessels. As described above, it is divided into three anatomic regions: the median eminence, the infundibular stem, and the neural lobe. The median eminence is composed of three layers: (1) The inner ependymal zone is formed by specialized cells called tanycytes that line the inferior region of the third ventricle. These cells are connected by tight junctions that prevent the passage of CSF into the interstitial fluid of the median eminence. (2) The internal zone contains the supraopticohypophyseal tract, composed of axons of the magnocellular neurons. (3) The external zone contains the parvocellular neuron terminals that form the dopaminergic tuberohypophyseal tract. The supraopticohypophyseal tract and the tuberohypophyseal tract are present also in the infundibular stem and terminate in the neural lobe. Damage to the supraopticohypophyseal and tuberoinfundibular tracts results in atrophy of the posterior lobe as well as of the supraoptic and paraventricular nuclei.

REGULATION OF THE HYPOTHALAMIC-PITUITARY AXIS

Hypothalamic control of the pituitary gland is mediated by two separate neurosecretory systems, known as the *parvocellular* and the *magnocellular* neuronal systems. The parvocellular system consists of small neurons originating in several hypothalamic cell groups that produce the hypothalamic hypophysiotropic hormones, including GHRH, TRH, CRH, GnRH, PIF (probably identical to dopamine), and the growth hormone-inhibiting hormone known as somatostatin (SS). Their axons form the tuberoinfundibular tract, which terminates in the median eminence, where these hormones are released. These various releasing and inhibitory factors regulate adenohypophyseal release of hormones via the portal circulation.

Brain neurotransmitters such as dopamine, norepinephrine, epinephrine, serotonin, acetylcholine, g-aminobutyric acid, and opioid peptides also can influence anterior pituitary function, primarily by interacting with hypothalamic hypophysiotropic hormone-containing neurons.

The magnocellular system consists of large neural cell bodies located in the supraoptic and paraventricular nuclei. These neurons contain the peptide hormones oxytocin and vasopressin, also known as antidiuretic hormone. Their neural fibers form the supraopticohypophyseal tract, which projects down the infundibular stem and terminates in the posterior lobe, where oxytocin and vasopressin are stored and released into the systemic circulation upon appropriate stimulation. Oxytocin induces smooth-muscle contraction of the uterus, leading to the expulsion of the fetus and the placenta, and contraction of the myoepithelial cells of the breast involved in milk ejection. Vasopressin is released from the posterior lobe of the pituitary into the systemic circulation, where it binds to receptors on renal distal tubules to induce water conservation, and to receptors on precapillary arterioles to elevate mean arterial pressure. The main factor regulating vasopressin release is plasma osmolality through osmoreceptors located in the hypothalamus. As osmolality increases, vasopressin is released to ensure maintenance of serum osmolality and water balance. If a blood loss greater than 10 to 15 percent of the vascular volume is incurred, the low-pressure volume receptors located in the left atrium and in the pulmonary veins and the high-pressure baroreceptors located in the carotid sinus and in the aortic arch will mediate the release of vasopressin. In the advent of hypotension, vasopressin is also carried by the portal venous system to the anterior pituitary, where it regulates the release of ACTH.

DIAGNOSTIC STUDIES

Complete examination of sellar and parasellar regions encompasses both neuroendocrine and anatomic (radiologic) evaluations.

Neuroendocrine Evaluation

Anterior Pituitary

The advent of radioimmunoassay and immunohistochemical methods has greatly enhanced our understanding of the function of the pituitary gland in normal and in pathologic states. The evaluation of the anterior pituitary is indicated in patients who present with signs or symptoms compatible with isolated or multiple hormonal deficits, hyperprolactinemia, hyperthyroidism, diabetes insipidus, a hypothalamic disorder, or with any sellar or parasellar lesion (Table 35-1). Recognition of any potential hypopituitarism before any surgical endeavor is imperative and may have implications for avoiding complications in the perioperative period.

TSH deficiency can be diagnosed by measuring simultaneously basal serum TSH and thyroid hormone levels. A low serum T₄ in the presence of an inappropriately low TSH level suggests a central cause of hypothyroidism. To assess the TSH reserve within the thyrotropes, thereby distinguishing a hypothalamic from a pituitary defect, a TRH test may be performed. Normally, TSH and both prolactin and GH rise in response to TRH stimulation.

Dynamic tests are required to diagnose a state of ACTH deficiency, because the morning cortisol level is persistently low only when the ACTH deficiency is very severe. The CRH test may be used to distinguish a hypothalamic CRH deficiency from a pituitary ACTH deficiency. The absence of ACTH responsiveness to CRH is diagnostic of a pituitary

corticotrope deficiency. The insulin-induced hypoglycemia and the glucagon test will stimulate the entire hypothalamic-pituitary-adrenal axis. The ACTH stimulation test evaluates the capacity of the adrenal glands to secrete cortisol.

Gonadotropin deficiency can be diagnosed by measuring simultaneously basal serum FSH and LH levels and gonadal steroids, estradiol in the premenopausal female and testosterone in the male. In the event of primary gonadal failure, the lack of negative feedback by the gonadal steroids on the hypothalamic GnRH and pituitary LH- and FSH-secreting cells leads to an elevation of LH and FSH. Low circulating gonadal steroid levels associated with inappropriately low gonadotropin levels suggests a hypothalamic or pituitary disturbance. The GnRH stimulation test will evaluate pituitary gonadotrope function. Lack of a secretory response of LH and FSH to GnRH indicates a lesion at the pituitary rather than at the hypothalamic level.

GH levels in the basal state often are low in normal individuals. Measurement of the plasma level of *insulinlike growth factor-1 (IGF-1)* permits a more accurate diagnosis of a GH-deficient state, because it reflects the integrated 24-h secretion of GH. Stimulatory tests to assess the somatotrope function include sleep and exercise studies, insulin- induced hypoglycemia, and administration of arginine, l-dopa, clonidine, propranolol, or GHRH.

Posterior Pituitary

Central *diabetes insipidus (DI)* refers to a state of relative or absolute insufficient secretion of vasopressin from the posterior pituitary gland. This must be differentiated from renal DI, in which the kidney fails to respond to an appropriate elevation in serum vasopressin level. The diagnosis of central DI may be established by the water deprivation test. This study assesses the patient's ability to concentrate urine in response to an increase in plasma osmolality. Both hypothyroidism and hypocortisolism might cause a decrease in the glomerular filtration rate and thus mask a state of DI. Therefore, before the test, the patient should be euthyroid and any adrenal insufficiency should be corrected. The diagnosis of DI is based on the development during the test of abnormally concentrated plasma (osmolality > 300 mOsm/kg) and of urine that remains dilute (osmolality < 270 mOsm/kg). Also, the urine volume is not reduced to the expected degree. At the end of the water deprivation period, the administration of exogenous vasopressin will correct these abnormalities (in contrast to renal DI, in which there is resistance to exogenous vasopressin).

Radiologic Evaluation

Radiologic examination of the pituitary should provide information about the bony anatomy of the sella turcica and its surroundings as well as intrasellar contents. Depending on the clinical presentation, the use of more than one imaging modality may be indicated.

The bony anatomy of the sella turcica may be visualized by performing lateral skull x-rays, which may demonstrate enlargement of the sella in cases of intrasellar tumors. Conventional polytomography provides greater resolution. However, if more detailed evaluation of the bone at the base of the skull is necessary, CT using special windows to enhance bony detail is indicated. This study offers the additional advantage of enabling limited soft-tissue resolution of the sella, which is helpful in distinguishing empty sella from a pituitary tumor, as both may result in enlargement of the sella.

For the diagnosis and management of pituitary lesions, the contemporary imaging modalities of CT and MRI have had a revolutionary impact. MRI is superior to CT because of its inherently greater soft-tissue contrast, which allows clear visualization of the optic chiasm, optic nerves, cavernous sinuses, and carotid arteries (Fig. [35-6](#)).

High-field, thin-section MRI appears to be the most sensitive imaging method for preoperative localization of pituitary adenomas. On unenhanced images, focal glandular hypodensity identified on coronal images is the most sensitive predictor of adenoma location. Radiologic evaluation should consist of coronal, sagittal, and axial MRI, with large tumors usually having similar signal intensity to brain on T₁-weighted images. The normal pituitary gland, infundibulum, and cavernous sinuses enhance immediately after administration of gadolinium-DTPA, allowing contrast between the enhancing normal glandular tissue and the low-intensity adenomas. A T₁- weighted image following the infusion of gadolinium-DTPA is the method of choice for the delineation of intrasellar pathology. Shortly after administration, the normal vascular pituitary will increase in signal intensity, and a pituitary tumor will contrast by remaining less intense, being slower to perfuse with the contrast agent (Fig. [35-7](#)). However, after sufficient time for dye uptake into the tumor, it will increase in signal intensity accordingly. After this, the normal pituitary surrounding will “wash out” the contrast before the tumor does; thus the reverse situation will prevail, and the tumor will remain at a higher intensity than the surrounding gland for a period of time. The optimal and most consistent time for visualization is therefore in the early postinfusion period, since the kinetics of dye perfusion and clearance may be variable among patients. Because MRI also offers visualization of the major vessels, the intracavernous carotid artery's proximity to the tumor can be determined; this is especially important in the rare case of severely ectatic carotid arteries that might preclude a transnasal surgical approach because of the risk of vascular injury. In summary, MRI should be the primary imaging test used whenever suspected pituitary or parasellar lesion is to be studied noninvasively.

CT may be used if MRI is not available, with direct coronal cuts or with coronal and sagittal reconstruction of axial sections through the sellar region. On unenhanced CT a pituitary tumor usually has a density slightly less than surrounding pituitary or cavernous sinuses. Adenomas may be homogeneous, contain low-density regions of necrosis or cyst formation, or demonstrate intratumoral hemorrhage or calcification. CT also can offer relevant information regarding the bony landmarks by varying the window widths, and it may reveal sellar enlargement or sloping, thinning, or erosion of the sellar floor, all indicative of an intrasellar expansile process. In addition, the presence of calcium in some sellar or parasellar lesions (e.g., craniopharyngioma, meningioma) may be better visualized on CT images, thus providing supplemental information to that obtained by MRI.

ANTERIOR PITUITARY DISORDERS

Postpartum Pituitary Ischemia

Postpartum infarction and necrosis of the pituitary, known as Sheehan's syndrome, leads to hormonal deficiencies. The pathogenesis of this syndrome is still debated, but among the postulated causes of pituitary ischemia at the time of delivery are hemorrhage with hypovolemic shock, and pituitary portal venous thrombosis due to diffuse intravascular coagulation. Panhypopituitarism may develop over an extended period, in some cases being reported 10 to 20 years postpartum. The main clinical features include postpartum failure to lactate, postpartum amenorrhea, and progressive symptoms and signs of adrenal insufficiency and hypothyroidism. Although panhypopituitarism is common, gonadotropic function may be spared or may recover. Diabetes insipidus, although classically considered an anterior gland disease, also occurs in 1 percent of cases.

Pituitary Adenomas

Pituitary adenomas are benign tumors that may originate from any of the pituitary cell types. With today's imaging modalities and radioimmunoassay techniques we are enjoying a new era in the diagnosis and management of pituitary adenomas. Such diagnostic studies have rendered obsolete the earlier morphometric histologic classification of pituitary tumors. Pituitary adenoma is a common incidental sellar abnormality; autopsy and radiologic studies suggest that microadenomas (lesions < 10 mm in diameter) may be present in 22 to 27 percent of older individuals but that macroadenomas (> 10 mm) are quite rare (Fig. [35-8](#)). Hormone secretion by an adenoma may be asymptomatic but is helpful in the differential diagnosis of an incidental mass on MRI.

These tumors are classified according to their secretory product (if any) (Tables [35-2](#) and [35-3](#)). However, some tumors may secrete multiple hormones (plurihormonal lesions), which may complicate this classification. Another potential confounding factor is in the definition of the term “hormone production.” Classically, because cell biologists and pathologists define production as hormone synthesis that is not necessarily associated with hormone release, the endocrinologic status of the patient associated with a particular histological tumor appearance may not be correlated. From a clinical standpoint, however, the true secretion of the hormone is important, because it determines any hypersecretory endocrinopathy and provides a serum parameter for clinical diagnosis and response to treatment.

Null-Cell Adenomas

Clinical Manifestations and Diagnosis

Although tumors that secrete any of the anterior pituitary hormones may grow to sufficient size to become macroadenomas, nonfunctioning adenomas more frequently present, with symptoms of large lesions, such as headache, visual failure, or hypopituitarism. Statistically, null-cell tumors are among the most common types of adenoma found in the pituitary and occur more frequently later in life than secreting tumors. The classic visual triad produced by larger tumors is characterized by optic disc pallor, early loss of central visual acuity, and visual field defects (bitemporal hemianopsia).

In the endocrinologic evaluation of the patient in whom a nonfunctioning pituitary tumor is suspected, an important diagnostic pitfall to be avoided is the misdiagnosis of a prolactin-secreting tumor. Larger adenomas composed of cells other than lactotropes may produce a mild-to-moderate hyperprolactinemia from the so-called stalk-section effect (disconnection hyperprolactinemia-stalk compression causing loss of dopaminergic inhibition to tonic prolactin release), which must be distinguished from a true prolactinoma. Under such circumstances, it is rare to see prolactin levels in excess of 100 ng/mL, at most 250 ng/mL. Consequently, large tumors (> 2 cm) associated with prolactin levels less than 250 ng/mL (and certainly those < 100 ng/mL) should be suspected of being nonfunctional when management strategies are being planned, and one would not expect such lesions to respond to chemical reductions of serum prolactin. Nonsecreting pituitary adenomas can be defined as those that demonstrate no *apparent* clinical or biochemical abnormality indicating hormonal excess, though it must be recognized that the potential exists for the secretion of an as-yet undetectable hormone or its precursor, or an identified hormone may be synthesized in insufficient quantities to be detected by immunoperoxidase methods. Indeed, for this reason, some have considered the term “nonfunctioning” a misnomer. Tumors that secrete the alpha subunit, common to all the glycoprotein hormones, are to be distinguished from the true nonfunctioning tumors; in these cases the alpha subunit may be detected immunohistochemically in tumor specimens and biochemically in the blood and urine, but because it possesses no biologic activity it is of no endocrinologic consequence if secreted. However, the *exclusive* secretion of the alpha subunit may be difficult to determine if each of the glycoprotein hormones has not been independently assayed. For example, TSH-secreting tumors may secrete excessive quantities of the alpha subunit.

With nonfunctioning macroadenomas, varying amounts of hypopituitarism may be present, produced by compression of the gland by tumor or from interference with blood supply leading to infarction of a normally functioning gland.

Therapy

The two objectives of the treatment of *all* pituitary tumors are (1) relief of signs and symptoms attributable to mass effect and (2) correction of endocrine abnormalities—hyposecretion or hypersecretion of adenohypophyseal hormones. When nonsecreting tumors are considered, correction of endocrine abnormalities may be of lesser consequence than is hormonal excess, but it is important to realize that continued growth of the lesion may precipitate hypopituitarism.

The success of medical therapy is predicated on the presence of any associated oversecretion of anterior pituitary hormones. Thus, in the majority of cases of nonfunctioning tumors, no pharmacologic therapy is available. Approximately 10 percent of these tumors will respond to bromocriptine, a dopamine agonist, with a reduction in size. The primary mode of treatment for these tumors should therefore be considered surgical, with other options to be considered (radiotherapy, observation) in patients in whom underlying medical conditions preclude operation.

Prolactin-Secreting Adenomas

Clinical Manifestations and Diagnosis

Prolactinomas are the most common functional pituitary tumors. Although they are found pathologically at autopsy with equal frequency in men and women, clinically they present more commonly in women. In a personal series of some 392 PRL-secreting pituitary tumors operated on by one of the authors (MHW), 321 (82 percent) were in female patients. The most common presenting symptom in this group was the development of secondary amenorrhea; approximately 50 percent of those with amenorrhea had associated galactorrhea. Epidemiologically, approximately 5 percent of women with primary amenorrhea and 25 percent of women with secondary amenorrhea (other than those who are pregnant) have a PRL-secreting pituitary tumor as the cause of their clinical symptoms. Because of the readily identifiable symptoms of amenorrhea and galactorrhea, these patients generally present relatively early in the course of evolution of the tumors; this is unfortunately not true in the male population. Since the primary symptom of a PRL-secreting tumor in the male usually is a decrease in libido well before true impotence is observed, this is frequently ascribed to the “aging process” or “functional” causes. Hence males often present at a more advanced age and later in the course of their disease with chiasmal compression and visual compromise.

The diagnosis is secured by radiologic evidence of a pituitary lesion with serial elevations of serum prolactin levels. There exists a rough correlation between the size of the lesion and the serum elevation of prolactin. In addition, local invasion of the tumor into the venous cavernous sinuses is associated with a marked increase in serum PRL level. The PRL level is of great help in determining subsequent management, because the ability to successfully extirpate the tumor is reduced with large and/or invasive lesions.

Therapy

The introduction of an effective medical management potential in the treatment of PRL-secreting tumors came with Food and Drug Administration (FDA) approval of bromocriptine in 1978. The efficacy of bromocriptine (a dopaminergic agonist) in reducing serum prolactin, in addition to reducing tumor size and inhibiting further tumor growth, is well established. The central considerations in treating such patients are the patient's ability to tolerate the medication, and the understanding that in most cases the treatment must continue for the duration of the patient's life. Patients with microadenomas occasionally may be cured with chronic bromocriptine therapy.

With PRL-secreting microadenomas, therapeutic options include medical or surgical management. In our series, 225 of 262 patients (86 percent) undergoing operation for tumors less than 1 cm in size had normal (< 20 ng/mL) postoperative PRL levels (i.e., “chemical cure”). Operative mortality was zero and associated morbidity was low. Strong consideration should be given to surgical intervention in patients who have smaller tumors without significant hyperprolactinemia.

In cases of larger tumors, however, surgical cures are much less frequent (in our personal series, only 63 of 130 [49 percent] of the patients with tumors larger than 1 cm achieved chemical cure). Although no medical therapy will result in cure of a PRL-secreting macroadenoma, only a minority of these patients will remain free of their disease after surgical treatment alone. It is our practice to place *all* patients with large or invasive pituitary tumors with endocrinologically documented prolactin secretion on a trial of bromocriptine initially and to monitor clinical status and radiographic appearance accordingly. All solid primary prolactin-secreting tumors should respond to the medication, both by a reduction in tumor size and by a reduction in prolactin level. The goal of therapy should be absolute normalization of prolactin levels, as a prolonged hyperprolactinemic state may be associated with significant osteoporosis and infertility. Surgical resection of these lesions is indicated only in those patients who are intolerant of the side effects of the medication, unable to afford the cost of the medication for a prolonged period, or in whom sustained tumor reduction is not effected. After operation, if hyperprolactinemia is persistent to some extent, the patient may be able to tolerate markedly reduced doses of bromocriptine to effect long-term control, or else postoperative radiotherapy to bring the residual tumor under control should be considered.

Cushing's Disease

Clinical Manifestations and Diagnosis

Cushing's disease is the result of hypersecretion of ACTH by the pituitary. ACTH-secreting tumors are present in over 90 percent of patients with a pituitary source of excess ACTH secretion; in the remaining 10 percent the cause of excess ACTH secretion is diffuse corticotrope hyperplasia from hypersecretion of CRH. The disease affects women approximately eight times more frequently than men. Ectopic ACTH production accounts for 10 percent of ACTH-dependent hypercortisolism and should be ruled out.

Clinical manifestations of the disease include: (1) those due to *glucocorticoid excess* (Fig. 35-9)—central obesity, “moon” facies, dorsocervical and supraclavicular fat pads, proximal muscle wasting, thin skin with ecchymoses and violaceous striae, cataracts, osteoporosis, amenorrhea, diabetes mellitus, growth retardation in children, immunosuppression with fungal infections, and depression; and (2) those due to *peripheral androgen excess*—hirsutism and acne (virilization is unusual), and hypogonadism from increased negative feedback to the hypothalamus and pituitary gland.

The diagnosis of Cushing's disease is confirmed by: (1) increased basal plasma ACTH and cortisol levels in the afternoon and evening with loss of diurnal variation; (2) an elevated 24-h urinary free cortisol excretion (> 100 mg/24 h); (3) a failure of serum cortisol to suppress with the *low-dose dexamethasone suppression test* (1 mg dexamethasone is administered at 11 p.m.; an 8 a.m. plasma cortisol remains higher than 5 mg/dL); and (4) suppression of ACTH after administration of high-dose dexamethasone (overnight 8 mg dexamethasone test or standard two-day test of 2 mg every 6 h for 2 days). A failure to suppress ACTH secretion with the high-dose dexamethasone suppression test suggests the diagnosis of an autonomous adrenal tumor or ectopic ACTH secretion.

MRI study of the pituitary identifies a microadenoma in around 50 percent of the cases, a macroadenoma less frequently; or the study may appear normal (a small microadenoma below the resolution of the MRI or corticotrope hyperplasia).

Therapy

Fortunately, most patients with Cushing's disease have microadenomas that lend themselves to complete surgical resection. Our experience with these microadenomas has been gratifying; in patients with Cushing's disease who have microadenomas, the series has demonstrated a 91 percent cure rate as defined by laboratory tests. On the other hand, patients with Cushing's disease who have macroadenomas present a serious problem. These tumors are frequently invasive into adjacent dura and bone and consequently defy chemical cure (as defined by laboratory tests) by surgical means alone. For patients with elevated ACTH levels that persist after radical surgical resection, we are currently exploring stereotaxic radiosurgery as a method to eradicate residual sellar and parasellar invasion while sparing adjacent structures (as opposed to standard external beam radiotherapy). If radiotherapy fails, trials of an adrenolytic agent, mitotane, or adrenal enzyme inhibitors, such as aminoglutethimide, metyrapone, and ketoconazole will be required to control the disease. For patients who have failed all modes of nonoperative therapy, bilateral adrenalectomy is recommended.

Acromegaly and Gigantism

Clinical Manifestations and Diagnosis

Excess GH secretion produces the clinical syndrome of *acromegaly*. If the excess secretion occurs before the epiphyses of long bones have fused, the result is *gigantism*. GH-secreting pituitary adenomas are the most common cause of excess GH secretion, and the patients present with macroadenomas. Rarely, the disease is a result of the ectopic secretion of GHRH from other tumors.

With gigantism, young individuals may attain great height if the disease progresses unchecked (often greater than 7 feet). After fusion of the epiphyses, GH produces the syndrome of acromegaly, with the soft-tissue and bony enlargement in characteristic locations (Fig. 35-10). Clinical manifestations of these soft-tissue changes include coarsening of facial features, laryngeal enlargement, goiter, thick heel pads, acanthosis nigricans, cardiomegaly, and hepatomegaly. Bony changes are extensive, producing facial prognathism, enlargement of the mandible with increased spacing between the teeth, and bony enlargements of hands and feet. Concomitant with these changes are compressive neuropathies, arthropathies, and osteoporosis secondary to hypogonadism. Metabolic manifestations include associated hypertension, diabetes mellitus, and goiter; hyperhidrosis also is commonly present. Acromegaly affects men and women with approximately equal frequency.

The diagnosis is made by assessing GH secretion. A basal fasting GH level greater than 10 ng/mL is present in 90 percent of those with acromegaly. The diagnosis is confirmed by the *glucose suppression test*. In patients with acromegaly, an oral administration of 100 g glucose fails to suppress the GH level to less than 2 ng/mL at 120 min. As GH is secreted in several peaks throughout the day, a single fasting level may fail to demonstrate GH hypersecretion. The serum IGF-1 level is elevated in those with acromegaly, and this measurement may prove to be more useful because it reflects the integrated pulsatile 24-h secretion of GH and is a more reliable parameter for assessing the disease and to monitor its response to treatment. MRI or CT demonstrates the presence of a pituitary adenoma in more than 90 percent of patients with endocrinologically documented acromegaly.

Therapy

With associated diabetes mellitus, hypertension, and cardiomyopathy in a significant proportion of patients with acromegaly, oversecretion of GH should be considered a “malignant endocrinopathy” that demands complete and expedient treatment. Complete surgical removal of pituitary adenomas that secrete GH usually provides rapid control of elevated GH levels, is capable of producing long-lasting “cure,” and is relatively safe. In patients in whom total removal of the lesion is not possible (tumors with large suprasellar extent or cavernous sinus invasion) or in whom surgery is contraindicated, medical therapy should be considered. Patients with elevation of GH and IGF-1 postoperatively should also receive medical therapy. The somatostatin analogue octreotide acetate (Sandostatin) reduces tumor size in approximately 20 to 50 percent of patients, normalizes GH levels in 50 percent, and normalizes IGF-1 levels in 40 to 80 percent. Present therapy, however, demands parenteral administration or continuous infusion, which dictates careful selection of reliable patients. A long-term parenteral congener, which may be administered every 3 weeks, is currently undergoing clinical trials. Radiotherapy of these lesions, as with other pituitary tumors, is moderately effective but is not without significant adverse effects.

Gonadotroph-Secreting Adenomas

Gonadotroph cell adenomas are uncommon tumors that usually present in middle-aged men with symptoms caused by mass effect: headaches, visual disturbances, and hypopituitarism. Although gonadal function usually is normal, hypogonadism may present in men and women, and some men may present with bilateral testicular enlargement due to FSH hypersecretion. Different hormonal secretory patterns may be encountered: (1) hypersecretion of FSH (the most common anomaly), (2) secretion of biologically inactive LH associated with hypogonadism, (3) hypersecretion of intact LH associated with elevated serum testosterone levels in men and precocious puberty in children, and (4) hypersecretion of alpha subunit. In approximately 50 percent of patients, administration of TRH intravenously induces inappropriate elevations of serum LH and FSH and their subunits. The treatment of these tumors is surgical, with radiotherapy being reserved for patients who are not surgical candidates or in whom total resection is not achieved. Measuring serum LH, FSH, and alpha subunit levels postoperatively can help to determine the extent of tumor removal. For surveillance of tumor recurrence, measurement of basal or TRH-stimulated gonadotrophins or their subunits should be performed.

Thyrotroph Adenomas

TSH-secreting pituitary tumors are rare. Most patients present with invasive macroadenomas. The TSH levels are below 10 mU/L in one third of cases (170), and are usually below 20 mU/L. Most patients present with overt hyperthyroidism and a diffuse goiter. These cases can be differentiated from other causes of hyperthyroidism by a basal alpha

subunit-to-TSH molar ratio of less than 1, and exogenous TRH administration may stimulate TSH and alpha-subunit secretion. Many thyrotroph cell adenomas also secrete PRL and GH. Transsphenoidal resection is the treatment of choice. Irradiation is reserved for patients who are not surgical candidates or in whom total resection is not achieved. Irradiation appears to stabilize tumor growth and render the patient euthyroid in approximately one-quarter of cases. Octreotide has been a useful adjunctive therapy while awaiting the ablative effect of radiotherapy. This drug is effective in most patients, rendering them euthyroid and significantly reducing tumor size in many.

Surgery for Pituitary Adenomas

Preoperative Evaluation

Any lesion proximal to the pituitary or hypothalamus requires adequate endocrine evaluation preoperatively to minimize the potential for an intraoperative or postoperative catastrophe because of inadequate pituitary reserve. The two most important of these are cortisol and thyroid levels. Because of the universal use of perioperative glucocorticoids in surgery involving this area, the risk of intraoperative hypocortisolemia is generally not a major factor. However, preexisting hypothyroidism can manifest acutely during the early postoperative period, underlining the need for adequate preoperative assessment. In cases with associated hypothyroidism, reestablishment of euthyroidism requires approximately 1 week of treatment before any elective surgical procedure. For this reason complete endocrine evaluation is performed in all patients in whom a pituitary lesion is suspected; normal endocrine function preoperatively, of course, does not guarantee such a state postoperatively, but preoperative function may identify patients in whom a risk of this occurrence should be considered in the perioperative period.

An assessment of electrolyte status is also important in identifying patients with marginal diabetes insipidus who may not comment on the long-standing urinary frequency.

Transsphenoidal Approach

The transnasal transsphenoidal approach is considered the procedure of choice for surgical access to sellar lesions (Fig. 35-11). The increase in popularity of this technique may be attributed also in part to the well-recognized inadequacy of the subfrontal approach for removal of the intrasellar component of the tumor. Moreover, several reports indicate favorable results with the transsphenoidal approach in the management of visual disturbances from macroadenomas, establishing this as the approach of choice for the surgical management of most pituitary tumors, regardless of size. Our assessment of more than 200 patients who presented with visual loss from among our first 1,000 pituitary patients operated on via the transnasal transsphenoidal route yielded evidence of improved vision in 81 percent, unchanged vision in 16 percent, and worsening of vision in 3 percent. These results are similar to those of other large series that have been reported on the efficacy of the transsphenoidal approach to suprasellar tumors and are equal to or exceed the results of large series of subfrontal explorations for visual loss. In addition, there is clear documentation of the potential for *improvement in pituitary function* after transsphenoidal adenomectomy with careful preservation of normal gland in patients with preexisting hypopituitarism. The efficacy of transsphenoidal surgery in selected patients with microadenomas has been established, with some reports of tumor control in more than 90 percent of the patients. In series including larger tumors, however, a less optimistic rate of 50 to 85 percent tumor control is obtained with surgery alone.

Transsphenoidal microsurgery for both large and small adenomas performed by experienced surgeons has acceptable mortality and morbidity. Of 2,606 patients with microadenomas and 2,677 patients with macroadenomas reported in an international survey, the operative death rates were 0.27 and 0.86 percent, respectively. Direct injury to the hypothalamus seemed to be the major cause of operative death, with delayed mortality attributed to CSF leaks and their attendant septic complications or to vascular injury. Operative morbidity includes persistent or permanent diabetes insipidus, with the incidence ranging from 1.8 percent in one large series to a 17 percent incidence immediately postoperatively with large adenomas. Postoperative CSF fistulas range from 1 to 4.4 percent in different series, depending on the size of the lesion and follow-up time, but they occur most frequently with larger lesions. Other major morbidity (stroke, visual loss, vascular injury, meningitis, cranial nerve palsy) is encountered in 3.5 percent, and minor morbidity (bleeding, nasal or sinus problems, diabetes insipidus, syndrome of inappropriate antidiuretic hormone secretion [SIADH], transient cranial nerve paresis, transient psychosis) occurs in another 3.5 percent of patients. Complications amounted to a small percentage of the overall surgical experience, emphasizing the relative safety of the procedure.

The rare relative contraindications to the transsphenoidal approach include: (1) extensive lateral tumor herniating into the middle fossa with minimal midline mass; these cases may require a primary or secondary transcranial procedure to remove tumor inaccessible by a midline approach; (2) ectatic carotid arteries projecting toward the midline, which are at risk of injury using a transsphenoidal approach; and (3) acute sinusitis, which may delay the procedure for treatment of infection. Previous rhinoplasty or submucous resection may increase the difficulty in developing the dissection planes, but they invariably can be established, and thus this factor should not alone constitute a contraindication to the transnasal approach.

Characteristically these tumors are soft and friable and at surgery may herniate down through the diaphragma sellae after evacuation of the intrasellar component; this may be facilitated if the anesthesiologist performs a Valsalva maneuver intraoperatively. Other techniques to promote herniation include the infusion of air or saline through a previously placed cisternal or lumbar catheter. A pure suprasellar tumor or one that requires suprasellar access may be approached, if necessary, by carrying the bony resection anteriorly over the tuberculum sellae. Such techniques may enable removal of sizable tumors within the suprasellar cistern, subfrontal space, and cavernous sinus; many of these will herniate into the enlarged sella once the intrasellar component has been evacuated.

Once the mass has been resected, attention must be paid to obliteration of the CSF fistula if the arachnoid has been violated. During transsphenoidal procedures in which the arachnoid has been breached, a fascia lata graft of appropriate size is routinely harvested to cover the opening and held in place with a small piece of autologous nasal cartilage or Marlex mesh. The sphenoid sinus behind this graft is then packed with fat obtained at the harvest of the fascia lata graft to further buttress the graft in position. A posterior nasal pack is then placed against the sphenoidal opening to maintain hemostasis and promote sealing of the graft. Using such careful techniques of closure, the incidence of CSF leakage is less than 1 percent, and attendant infectious complications are thus avoided.

Transcranial Approach

There are instances in which the transcranial approach is desirable, as noted earlier, such as when the transsphenoidal approach is hazardous because of the presence of ectatic carotid arteries, or when the tumor spills over into the middle fossa while leaving a small, virtually normal-sized sella turcica. Such lesions generally occur because of an incompetent diaphragma sellae that enables superior and lateral growth of tumor without significant expansion of the sella. Under such circumstances, direct visualization of the tumor by the transcranial approach may be desirable. In addition, there are occasional cases in which the consistency of tumor that is encountered via the transsphenoidal approach is such as to defy an adequate resection of tumor from this approach. This may require a secondary transcranial procedure if inadequate decompression of the optic mechanism has been established.

The consistency of the tumor is usually that of a soft and friable lesion that is easily debulked with the use of curettes of variable lengths and rotations. The main difficulty encountered with the removal of the lesion is the complicating anatomy related to the vascular supply to the hypothalamus and optic chiasm, in addition to the visual structures themselves. This anatomy may preclude total removal without injury to cranial nerves or the midline neuraxis. In such cases it is certainly more prudent to remove the soft interior and leave a densely adherent capsule than to risk cranial and vascular injury for a lesion that is likely not curable by surgery alone. These patients will invariably need radiotherapy for ultimate tumor control. The goals of the operation should primarily be decompression of the optic apparatus and judicious tumor removal without exposing neural or vascular structures to undue risk of injury.

Perioperative Management

Perioperative glucocorticoids are administered to all patients; this is crucial if preoperative endocrine assessment indicates any hypocortisolemia. Methylprednisolone is given intravenously at sizable dosages of 40 mg (or 10 mg dexamethasone) every 6 h in the immediate perioperative period in those cases with neurologic compromise, usually starting the day before surgery and continuing for 1 or 2 days postoperatively, then in a tapering dosage regimen tailored to the patient's projected glucocorticoid needs as anticipated by the preoperative endocrine assessment and intraoperative findings. In patients without visual compromise, lower dosages of these high-potency glucocorticoids with a more rapid postoperative tapering may be used.

Thyroid function should be assessed preoperatively and normalized before any elective surgical intervention. The stress of surgery can provoke an acute crisis in the patient without sufficient reserve and should be a consideration in patients with an otherwise unexplained alteration in mental status postoperatively.

Serial visual field testing is routinely performed in the recovery room and intensive care unit (ICU) to monitor visual and general neurologic conditions. In the patient with preoperative visual deficit, careful monitoring in the early postoperative period is essential. Both the transfrontal and transsphenoidal routes are successful in improving vision. Immediate postoperative improvement of vision may occur, with a significant improvement usually within the first 2 weeks, but continued improvement may occur for up to 12 months. More important, any *loss* of vision in the postoperative period may indicate an evolving hemorrhagic complication, and emergent CT scanning should be performed. Evidence of hemorrhage on the postoperative scan with a progressive visual deficit would warrant emergent transsphenoidal reexploration.

Blood pressure is carefully monitored; hypotensive events are to be avoided, especially in cases with compressive neurologic deficit in which tissue perfusion is already marginal.

Urine volumes and specific gravities must be followed in concert with sodium levels in order to clearly understand the dynamics of potential postoperative diabetes insipidus. Manipulation of the normal posterior pituitary gland may produce transient fluctuations in serum sodium level.

Before discharge, a morning fasting serum cortisol level is obtained to determine the need for cortisol replacement. Intraoperative evidence of residual normal pituitary gland is also a major guide in such considerations. Thyroid evaluation is usually done at 3 to 4 weeks postoperatively, since autonomous function of the thyroid may persist for some time postoperatively.

Unless clinically indicated, postoperative imaging is not performed for at least 6 weeks postoperatively, and 3 months in most cases. This allows clearance of all operative artifactual changes that might confuse the decision about the implementation of postoperative adjuvant therapy.

Radiotherapy for Pituitary Adenomas

Primary Radiotherapy

Radiotherapy was advocated for the management of pituitary tumors as early as 1907. Radiotherapy per se, however, should not be considered a completely benign therapy or an equivalent alternative to microsurgical resection. Adverse effects from radiation in this region can range from mild to severe. Radiotherapy carries *a significant risk of worsening preexisting hypopituitarism*, with an overt 10 to 15 percent frequency of panhypopituitarism (Fig. 35-12). It also can increase the rate of atherogenesis in the major vessels in the field, and it can cause visual impairment. These complications increase as a function of total treatment dosage. The visual impairment can result from one of several mechanisms, including empty sella syndrome, treatment failure, or direct radiation damage to optic pathways. Other minor complications from radiotherapy include epilation, scalp swelling, and otitis.

In the elderly patient or any patient who has a medical illness in which general anesthesia and surgery pose a significant risk, consideration should be given to primary radiation therapy, following an appropriate clinical diagnosis of a nonfunctioning pituitary tumor. The indications for this nonsurgical approach are strengthened in cases with hypopituitarism. A dose of 4,000 cGy by external beam is considered optimal by most radiotherapists. Reported series of patients treated with radiotherapy alone describe a 50 percent recurrence rate with local control in 75 percent after salvage treatment. Other authors report a local control rate of 50 to 79 percent, with an adequate salvage in cases of recurrence. It has been suggested that late recurrence after radiotherapy, which is not uncommon, may be the result of inadequate dosage. Treatments with 4,500 cGy given in 25 fractions can result in a high (greater than 90 percent) probability of stable long-term control.

The recent development of *stereotactic radiosurgery* (image-directed focal radiotherapy) may provide a safer and more effective method of radiotherapy for pituitary adenomas. Focused radiotherapy may potentially avoid damage to midline neural and vascular structures while delivering high local therapeutic doses to the lesion with great accuracy. Stereotactic radiosurgery offers an alternative for patients who have failed previous microsurgery, and who do not require decompression of the optic apparatus or more rapid normalization of endocrine status. There are few published series with meaningful follow-up for radiosurgery of pituitary adenomas, and optimal dosimetry has not yet been defined. Early reports indicate tumor growth control rates better than 90 percent but a less optimal control rate of endocrinopathy in functional tumors. These results all indicate that stereotactic radiosurgery may be performed for pituitary adenomas with low morbidity. Additionally, these series have demonstrated that the advantage of stereotactic radiosurgery over the use of external fractionated radiation or charged particles is reduction in postoperative hypopituitarism. Present protocols require the tumor to remain a distance of more than 2 mm from the optic nerves or chiasm to avoid visual deficit from radiation.

In the asymptomatic elderly patient with a nonsecreting tumor, intact pituitary function, and no compromise of the visual system, a case can be made for merely monitoring the patient with routine clinical (visual field) and endocrine evaluation, with serial MRI or CT scanning being performed at least yearly; these tumors may exhibit a benign course without reaching symptomatic dimensions within the remaining life expectancy.

Postoperative Radiotherapy

The rationale for the use of postoperative radiation therapy is to reduce the incidence of recurrence, with several studies suggesting improved tumor control with the combination of surgery plus radiotherapy. This is especially true for large and invasive lesions, which manifest an increased rate of recurrence. This treatment does not ensure recurrence-free survival, but the time to recurrence may be prolonged. Recurrences have been reported up to 18 years after operation; published recurrence rates in series with short follow-up times may therefore be misleading.

With functioning pituitary tumors, evaluation of postoperative endocrine status may give an indication of the effectiveness of the surgical removal; however, in nonfunctioning lesions the judgment of the surgeon supplemented by postoperative imaging are the only parameters to gauge the extent of resection and therefore the risk of recurrence. The surgeon's appreciation of the totality of the resection may not be accurate in the face of an invasive tumor. The lack of a chemical marker in a true nonfunctional tumor makes assessment of cure difficult in the postoperative period. Furthermore, in contrast to prolactin- and GH-secreting tumors, no adjunctive pharmacotherapy is available.

For these reasons, the criteria for selection of patients for postoperative radiotherapy remains controversial. In general, large tumors, which frequently invade the dura and therefore defy surgical excision, should all be considered candidates for postoperative radiotherapy, especially if the patient has hypopituitarism postoperatively. Similarly, with frank cavernous sinus invasion, postoperative radiotherapy would be advocated. In cases in which tumor invasion is not evident and "total" removal has been achieved, observing the patient with routine scanning on a yearly basis can be an appropriate strategy, especially if endocrine function is intact.

OTHER LESIONS

Included in the differential diagnosis of mass lesions that can affect pituitary function are a variety of tumors or cystic lesions in the region of the midline skull base. Many of these can be differentiated from primary pituitary lesions on the basis of the clinical history, physical examination, and radiographic appearance.

Benign Pituitary Cysts

Benign pituitary cysts are a common autopsy finding. They can occasionally become large enough to become symptomatic. These represent a disparate pathologic group of entities that often require histologic evaluation to differentiate. Although pituitary adenomas may contain cystic components and represent the most common cystic lesions of the pituitary, many other cystic lesions are believed to consist of remnants of Rathke's pouch. Rathke's pouch is a transient embryologic structure (third or fourth week of embryonic life) that arises from an outgrowth of stomodeum and elongates to form the craniopharyngeal duct. With further development of the pouch wall to form the anterior lobe of the pituitary, pars tuberalis, and pars intermedia, the residual lumen is reduced to a cleft, which usually regresses. It is hypothesized that persistence and enlargement of the cleft results in a symptomatic Rathke's cleft cyst. The cyst is characteristically lined by a single layer of cuboidal or columnar epithelial cells. Most Rathke's cleft cysts are asymptomatic. Occasionally they enlarge and cause symptoms by local mass effect (hypopituitarism or visual loss). Treatment by transsphenoidal drainage usually is effective.

Less common cystic lesions in this region include arachnoid cysts, pars intermedia cysts, and cysticercal cysts.

Meningiomas

Meningiomas are common tumors that arise from cells of the meninges. Most of these tumors are well-demarcated round or oval tumors that are attached to the dura mater. The vast majority of meningiomas are benign tumors that compress, but do not invade, the parenchyma of the brain. Meningiomas can arise in several locations in and about the sella turcica, the most common of these being the planum sphenoidale, the tuberculum sellae, the anterior or posterior clinoid, or the diaphragma sellae itself. It is distinctly uncommon for such a tumor to present as a completely intrasellar mass. Clinically, patients with meningiomas in these locations present with visual loss or other cranial nerve dysfunction. In contrast to pituitary adenomas, the presentation of visual loss usually is asymmetric; physical examination often yields incongruous and asymmetric visual field defects and optic atrophy. The temporal pattern of visual loss may be acute, gradual, or fluctuating. The pattern of other cranial nerve dysfunction is variable, but involvement of the superior orbital fissure or the cavernous sinus may produce ophthalmoplegia with associated facial sensory deficits (cranial nerve V dysfunction). With larger tumors, variable amounts of pituitary dysfunction can occur from compression of the hypothalamic- pituitary axis.

On unenhanced CT meningiomas are often isodense or hyperdense compared to normal brain. After administration of contrast agent, these tumors enhance markedly on the CT scan and with gadolinium-DTPA on MRI.

The management of meningiomas is primarily surgical, with radiation therapy reserved for lesions that defy removal in younger patients.

Craniopharyngiomas

Though their precise embryologic origin is still debated, these benign squamous epithelium-lined tumors can present at any time of life. Approximately one-half of symptomatic patients present in childhood, making the lesion relatively more common in this age group. Because they are slow-growing, extraaxial tumors, the symptoms produced may be delayed until the tumor has reached considerable size. This is often the case in children, in whom the tumor can reach enormous size, often several centimeters in diameter. However, purely intrasellar tumors in older patients may be only a few millimeters in diameter and might be mistaken for a pituitary microadenoma. Because adults are much more sensitive to impairment in visual function, this is a common presentation in the older age group. Children, in contrast, present more frequently with symptoms of growth retardation secondary to hypothalamic-pituitary dysfunction (Fig. [35-13](#)), or of increased intracranial pressure, often the result of obstruction of midline CSF pathways.

Craniopharyngiomas have a characteristic appearance on skull x-ray or CT scan. The most common location for these tumors is in the suprasellar cistern, but intrasellar lesions are sometimes seen, as are tumors that lie entirely within the third ventricle. In children there is usually erosion of the sella with calcification within or above the sella. This calcification may be missed if only MRI is obtained. Both CT and MRI can demonstrate enhancement of the solid region of the tumor after appropriate contrast agent administration.

Though the primary mode of treatment of craniopharyngiomas is surgical, long-term follow-up results are disappointing, even if the surgeon estimates "total" operative removal. These tumors are somewhat responsive to radiotherapy, with several studies indicating prolonged interval to recurrence and survival after postoperative radiotherapy.

Optic Chiasm or Hypothalamic Gliomas

Astrocytoma of the optic nerve, the optic chiasm, or the hypothalamus is uncommon in adults but comprises 7 percent of intracranial tumors in children. Chiasmal and hypothalamic lesions are difficult to differentiate clinically or pathologically, and hence are often discussed together. There is an association of these gliomas with neurofibromatosis, but sporadic cases are more frequent. For lesions confined to the chiasm, the presentation is one of monocular or binocular visual impairment with or without optic atrophy. If the optic tract is involved, homonymous hemianopsia may be noted. With larger lesions, hypothalamic involvement is invariable, and children may present with endocrinopathy or lesional obstruction of CSF pathways. Endocrinopathies include precocious puberty or the diencephalic syndrome.

MRI can be helpful in distinguishing this tumor from a craniopharyngioma or a meningioma; coronal views can distinguish whether the chiasm is enlarged with tumor or displaced by another lesion in the suprasellar cistern.

These lesions defy surgical cure. Surgery is used primarily to establish the diagnosis and plan subsequent treatment, which usually is conservative management in young children and radiation in selected older patients.

Suprasellar Germinomas

Germinomas are tumors of germ cell derivation that occur primarily in adolescents and young adults. They occur most commonly within the cranium in the region of the pineal gland, the suprasellar-retrochiasmal region, within the third ventricle, the interpeduncular or quadrigeminal plate regions, or within the pituitary fossa. These tumors are not encapsulated and infiltrate so extensively that complete surgical excision is not possible without producing extensive neurologic deficits. Clinically, patients with suprasellar tumors typically present with diabetes insipidus (present in 95 percent of patients), visual difficulties, or endocrinopathy. Diagnosis usually is made with MRI, in which the tumors are intensely enhanced with administration of intravenous contrast material. Therapy usually consists of surgical biopsy or partial removal followed by radiation or chemotherapy.

Sellar Metastases

Metastases to the region of the sella turcica are uncommon and usually are the result of bony metastases to the region. A variety of tumors have been reported to metastasize to the region of the pituitary gland itself, but these are rare. When metastases do occur to the pituitary gland, the presenting symptoms are those of progressive panhypopituitarism.

Mechanism of spread to the skull base in this area is via venous routes, such as the paravertebral venous plexus, which anastomoses with the venous drainage at the base of the skull. If metastatic cells pass through dural veins, subdural seeding is possible, and contiguous spread through Virchow-Robin spaces allows intracerebral metastases to occur. Direct arterial spread may occur to the region of the posterior pituitary, usually in patients in whom metastatic spread to the lungs has already occurred (the “multistep”metastatic hypothesis).

Lymphocytic Hypophysitis

A rare, presumably autoimmune disorder, lymphocytic hypophysitis is characterized by extensive infiltration of the anterior pituitary with lymphocytes and plasma cells. The pathophysiologic process apparently begins as an acute inflammatory process with diffuse enlargement of the gland. With continued inflammation, the normal pituitary architecture becomes difficult to recognize, and the process might result in diffuse interstitial fibrosis and gland atrophy. The neurohypophysis does not appear to be primarily involved with the process. The rarity of the disorder and the lack of specific radiologic or endocrinologic features have made it difficult to diagnose without a tissue biopsy.

Empty Sella Syndrome

Though not neoplastic, the empty sella syndrome can be mistaken for an intrasellar mass on radiologic imaging. The syndrome results from a herniation of the arachnoid and subarachnoid space of the suprasellar cistern through an incompetent diaphragma sellae. This may be primary, or it may follow pituitary surgery or radiation therapy. CT or MRI will demonstrate an apparently “empty” sella, its contents the same density or signal intensity as CSF. The pituitary stalk and the compressed gland may be visualized and displaced posteriorly. Occasionally, the patient may manifest an associated compressive endocrinopathy.

TRAUMA

Though not common, the pituitary stalk is susceptible to transection in major skull fractures of the base of the skull. In up to 10 percent of fatal head injuries, the stalk is injured. Transection of the pituitary stalk at the skull base results in massive infarction of the anterior lobe, with the subsequent development of panhypopituitarism. It is likely, however, that most pituitary and hypothalamic damage results secondarily from increased intracranial pressure caused by diffuse cerebral swelling.

POSTERIOR PITUITARY DISORDERS

Diabetes Insipidus

Deficient secretion of vasopressin leads to an impairment of water conservation. Excessive volumes of urine are excreted, resulting in increased plasma osmolality and marked thirst. Dehydration can occur if the thirst mechanism is impaired or if the patient does not have access to adequate amounts of water. One-third of the cases of central diabetes insipidus are idiopathic; the other causes include tumors, granulomatous disease, or trauma destroying the hypothalamus, pituitary stalk, or posterior pituitary.

Syndrome of Inappropriate Secretion of Antidiuretic Hormone

Hypersecretion of vasopressin—syndrome of inappropriate secretion of antidiuretic hormone—(SIADH) occurs in 15 percent of hospitalized patients. It leads to impaired water excretion and often manifests few signs other than hypotonicity of body fluids. Evidence of increased total body water, such as edema and hypertension, is absent unless sodium excretion is impaired as a result of some other cause. Renal excretion of sodium is increased (> 20 mEq/L), and urinary osmolality is inappropriately elevated in relation to plasma osmolality. Plasma vasopressin concentrations are inappropriately elevated in relation to plasma hypoosmolality. Reported causes include various central nervous system disorders, neoplastic diseases, pulmonary diseases, and drugs.

Adrenal

Jeffrey A. Norton

Historical Background Embryology Anatomy Adrenal CortexPhysiology Pathology Cushing's Syndrome Adrenal Insufficiency (Addison's Disease) Primary Hyperaldosteronism (Conn's Syndrome) Hypoaldosteronism Adrenogenital Syndrome Adrenal Mass Ectopic ACTH Syndrome Adrenal MedullaPhysiology Pheochromocytoma Neuroblastoma Adrenalectomy

HISTORICAL BACKGROUND

In the mid-sixteenth century (1563), Eustachius published the first accurate anatomical drawings of the human adrenal glands and the kidneys with the proper anatomic relationships to the aorta and inferior vena cava. Nearly two and a half centuries later, in 1805, Cuvier described the anatomical division of each adrenal gland into the cortex and the medulla. In 1855 Thomas Addison described the cachexia, weakness, gastrointestinal symptoms, weak pulse, bronze skin color, and fatal prognosis associated with untreated adrenal insufficiency, a condition that still bears his name. Following Addison's treatise, Brown-Séquard demonstrated in experimental animals that the adrenal glands were essential for life.

In 1886 Frankel first described the adrenal tumor that we call a pheochromocytoma. Physiologists Oliver and Sharpey-Schafer noted that there was a substance in the adrenal medulla that dramatically raised the blood pressure of dogs, which they named “adrenalin” in 1895. Two years later Abel, at Johns Hopkins, made a similar observation, and called the substance epinephrine. In 1901 he demonstrated that epinephrine would not maintain life in an adrenalectomized animal. In 1902 Kohn demonstrated that chromaffin-positive cells were present in the adrenal medulla, the carotid body, paraganglia in the abdomen, and the organ of Zuckerkandl. In 1912 the pathologist Pick named an adrenal tumor “pheochromocytoma” for the dark color of the tumor with the characteristic chromaffin reaction. The first successful resections of a pheochromocytoma were performed in 1926 by Roux in Switzerland and Charles Mayo in the United States. Subsequently, Jaffe and others proved that pheochromocytomas were associated with paroxysmal hypertension.

Around 1912 Harvey Cushing had studied and documented the condition that now bears his name. In 1932 he reported 11 patients with moon face, truncal obesity, hypertension, polyphagia, polydipsia, polycythemia, and susceptibility to bruises and infections. Many of the patients were known to have had adenomas of the pituitary. In 1930 Hartman and associates purified adrenocortical extracts that were able to keep adrenalectomized animals alive. In 1934 Kendall, of the Mayo Clinic, obtained crystals from adrenal cortical extracts that retained adrenocortical activity. One of these crystallized extracts, compound E, was named “cortisone.” Ten years later, Kendall and coworkers were able to synthesize cortisone.

In 1952 a potent mineralocorticoid was isolated from beef adrenal gland by Grundy and associates. It was purified and called “aldosterone.” Three years later, in 1955, Conn described a patient and the clinical syndrome associated with a tumor that secreted excessive amounts of aldosterone (Conn's syndrome). In 1962, Bartter and colleagues described another form of hyperaldosteronism associated with elevated levels of renin secondary to hyperplasia of the renal juxtaglomerular apparatus. The two syndromes can be distinguished from each other by measurement of plasma renin levels, which are undetectable in Conn's syndrome.

In 1865 DeCrecchio first described the condition of congenital adrenal hyperplasia occurring in a female pseudohermaphrodite. Later, in 1887, Phillips reported four similar patients who died from salt wasting. In 1939 Butler hypothesized that the salt-wasting was from adrenal insufficiency, and in 1950 it was established that salt wasting could be treated with cortisone. Bartter and associates demonstrated that patients with salt wasting had elevated circulating levels of adrenocorticotrophic hormone (ACTH), further supporting the hypothesis of adrenal insufficiency.

EMBRYOLOGY

The adrenal gland and medulla arise from separate embryologic tissues. In the fourth week of gestation, cells destined to become the adrenal cortex develop from the mesoderm medial to the urogenital ridge. At the sixth week of gestation, these developing cells are penetrated by nerve fibers, through which medullary cells will migrate (Fig. 35-14). By the eighth week of gestation, the cortex has formed two distinct zones: a large and centrally located fetal zone, and a thin rim of cortex that will become the cortex in the adult. Throughout this period, proliferation of cells occurs largely in the outer cortex, suggesting that the same cells give rise to the fetal and the definitive zones. Around the same time, the fetal adrenal circulation is established, with several arteries entering the adrenal gland from the aorta and a central vein that drains the adrenal effluent. During the second and third months of gestation, the weight of the adrenal glands increases from 5 to 80 mg, and the adrenals then are much larger than the adjacent kidneys. After the twentieth week of gestation, growth of the adrenal cortex is dependent on stimulation from the pituitary gland. Anencephalic fetuses are born with an atrophic adrenal fetal zone. After birth, the fetal zone involutes, and the three zones of the adult adrenal cortex develop.

During the development of the fetus, primitive adrenocortical cells may widely migrate. Accessory adrenocortical tissues have been identified within the broad ligament, near the celiac axis, adjacent to the ovarian or testicular veins, and around the kidney or uterus. Less common locations include other abdominal organs and, rarely, the lungs, spinal nerves, and brain. These ectopic rests of adrenal tissue can be a significant problem because they can cause persistent or recurrent hypercortisolism in a patient with high levels of ACTH production (Fig. 35-15).

The adrenal medulla arises from the primitive sympathetic nervous system, derived from the neuroectoderm. These cells differentiate into neuroblasts that migrate ventrally from the neural crest to form the sympathetic ganglia, and into the pheochromoblasts that form catecholamine-secreting or chromaffin cells. In the fetus, neuroblasts and pheochromoblasts are found in the adrenal medulla and extraadrenal sites. The organ of Zuckerkandl, located near the origin of the inferior mesenteric artery, is the most common site for extraadrenal catecholamine-secreting tissue. The widespread occurrence of extraadrenal chromaffin cells accounts for the location of extraadrenal pheochromocytomas later in life.

ANATOMY

The adrenal glands are paired structures in the retroperitoneal space, lying at the level of the eleventh thoracic vertebra, and anteromedial to the upper pole of the kidney (Fig. 35-16). The normal adrenal gland measures about 5 by 2.5 cm and weighs from 3 to 6 g. Adrenal weight may increase by approximately 50 percent during periods of stress and during pregnancy. The adrenals are supplied by numerous small arteries arising from the aorta and the celiac, superior mesenteric, inferior phrenic, and renal arteries. The arteries are short and are located mainly on the medial side of each adrenal gland, most coming directly off the aorta. Phrenic branches are found along the superior medial margin of the adrenal gland, and renal branches along the inferior medial margin. These arteries join and anastomose over the surface of the gland, and numerous small, unbranched arteries descend through the capsule. Nitric oxide concentration is an important local factor for maintaining high local levels of blood flow to the various zones of the cortex and the medulla. Adrenal blood empties through a single central vein into the inferior vena cava on the right, and the renal vein on the left. The short, wide configuration of the right adrenal vein makes catheterization of the vein for hormone sampling difficult and results in a greater probability of life-threatening hemorrhage during removal of the right adrenal gland.

The adrenal cortex has a characteristic chrome yellow color and is divided into three zones: the zona glomerulosa, the zona reticularis, and the zona fasciculata. The outer cells of the cortex are the aldosterone-secreting cells of the zona glomerulosa. The inner cells of the cortex are the cortisol- secreting cells of the zona fasciculata. The middle cortical cells are the adrenal androgen-secreting cells of the zona reticularis. The center of the adrenal gland, the medulla, contains the catecholamine-secreting cells. Medullary cells are

chromaffin positive.

ADRENAL CORTEX

Physiology

The adrenal gland secretes five major steroid hormones: cortisol, aldosterone, progesterone, testosterone, and estradiol. The major biosynthetic pathways of adrenal steroid synthesis from cholesterol are shown in Fig. 35-17. It is advantageous to think of these pathways as three distinct points of origin: one in the zona glomerulosa, producing aldosterone, a second in the zona fasciculata, producing cortisol, and a third in the zona reticularis, called the adrenal androgen/estrogen pathway that produces testosterone, estradiol, and estrone. Each pathway requires enzyme reactions and approximately five steps from cholesterol to the end product.

Aldosterone

Aldosterone causes the kidney to retain sodium, which increases the extracellular fluid volume. Aldosterone secretion is regulated primarily by the renin-angiotensin system, with potassium, atrial natriuretic hormone (ANH), and dopamine also making important contributions. The factors that stimulate aldosterone secretion have direct effects on the kidney and vasculature and provide maintenance of the extracellular fluid volume, concentration of extracellular potassium, and blood pressure. When intake of salt is restricted and blood volume is contracted, an increase in secretion of renin, and subsequently of angiotensin II, stimulates the secretion of aldosterone (Fig. 35-18). Renin is produced and secreted by the juxtaglomerular cells, which are differentiated smooth muscle cells located in the renal afferent arteriole. An area of the distal tubule near its origin, the macula densa, lies in proximity to the juxtaglomerular cells at the hilus of the glomerulus. The four factors that control renin release by the juxtaglomerular cells are: baroreceptors located within the afferent arteriole, the macula densa, the sympathetic nervous system, and the negative feedback of angiotensin II (see Fig. 35-18). Angiotensin II and aldosterone increase the reabsorption of sodium from the renal tubule to restore the extracellular fluid volume. Conversely, when salt intake is increased and extracellular fluid volume is expanded, the formation of angiotensin II from renin is decreased, ANH is secreted by the heart, and the secretion of dopamine by the adrenals also increases. ANH and dopamine inhibit aldosterone secretion and sodium excretion by the kidney. The serum level of potassium also exerts an effect on aldosterone secretion. Hyperkalemia stimulates aldosterone secretion, but hypokalemia inhibits it.

Cortisol

Glucocorticoids serve essential functions in human beings without which life cannot be sustained (Table 35-4). Glucocorticoid secretion is pulsatile and assumes a circadian pattern, with peak secretion before awakening in the morning (Fig. 35-19). Glucocorticoid secretion is stimulated by stress and inhibited by the negative-feedback effects of cortisol (Fig. 35-20). All physiologically important effects on cortisol secretion are mediated through adrenocorticotrophic hormone (ACTH). ACTH has a short plasma half-life. The rapid metabolism of ACTH combined with the pulsatile pituitary secretion causes rapid changes in plasma ACTH levels. Several hormones can stimulate the release of ACTH from the pituitary gland, including antidiuretic hormone and epinephrine, but the most important is *corticotropin-releasing hormone (CRH)*. CRH is secreted in episodic bursts by the paraventricular nuclei of the hypothalamus, which explains the pulsatile secretion of ACTH. Plasma levels of cortisol exert a negative feedback on secretion of ACTH and CRH at the level of the pituitary and the hypothalamus, respectively.

Sex Steroids

During normal human sexual development, the adrenal gland matures through a specific process called adrenarche. Adrenarche results in adrenal secretion of dehydroepiandrosterone (DHA), DHA sulfate, androstenedione, testosterone, and estrone (see Fig. 35-17). DHA and DHA sulfate are plasma markers of adrenarche, because more than 90 percent of plasma levels come from adrenal secretion. Gonadal secretion only contributes to the plasma levels of the other steroids.

Pathology

Hyperplasia

Hyperplasia is defined as an increased number of cells and is associated with an increased function or compensatory change. In patients with Cushing's disease (excessive production of ACTH by a pituitary tumor), the adrenal gland is approximately twice normal size. The weight of a hyperplastic adrenal gland increases from a normal range of 3 to 6 g, to a range of 6 to 12 g. Microscopically, there is a widened inner zone of the compact zona reticularis and a sharply demarcated outer zone of clear cells. In patients with ectopic ACTH production by a nonpituitary tumor, the microscopic appearance of adrenal glands is similar, but the weight is greater, from 12 to 30 g. Primary pigmented micronodular adrenal hyperplasia (1- to 5-mm nodules with pigmented appearance and normal glandular weight) is a hyperplastic condition of the adrenal gland that is not associated with excessive ACTH secretion. It occurs more commonly in children and can occur in a familial pattern.

Adrenal Cortical Adenoma

Adrenal adenoma is a benign neoplasm of adrenal cortical cells that may produce symptoms secondary to excessive hormone secretion that is not controlled by normal feedback mechanisms. A cortical adenoma does not exceed 5 cm in diameter and weighs less than 100 g. Some cellular pleomorphism and tumor necrosis may be present, but usually the cells are homogeneous and bland with little mitotic activity. It is difficult to distinguish adenoma from carcinoma on the basis of cellular morphology. Adenomas produce syndromes of hypercortisolism and hyperaldosteronism but seldom produce adrenogenital syndromes. Larger tumors (> 6 cm) producing adrenogenital syndromes usually are carcinomas.

Adrenal Cortical Carcinoma

Adrenal cortical carcinoma is a malignant neoplasm of adrenal cortical cells demonstrating partial histologic and functional differentiation. Adrenal cortical carcinomas are rare, comprising 0.05 to 0.2 percent of all cancers. Women develop functional adrenal cortical carcinomas more often. There is a bimodal occurrence by age with a peak incidence of less than 5 years, and a second peak in the fourth and fifth decade of life. Adrenocortical carcinoma is described as part of a complex hereditary syndrome that includes sarcoma, breast, and lung cancer.

Adrenal cortical carcinomas are larger than 6 cm and weigh between 100 and 5000 g. Areas of necrosis and hemorrhage are common. Invasion and metastases can occur. Microscopically, the appearance is variable. Cells with large nuclei, hyperchromatism, and enlarged nucleoli are all consistent with malignancy. Vascular invasion, desmoplastic bands, and mitoses are suggestive of malignancy. Immunostaining profiles of adrenocortical carcinoma and adenoma are similar. The only reliable criterion is the presence of nodal or distant metastasis.

Cushing's Syndrome

Cushing's syndrome (endogenous hypercortisolism) is caused by the secretion of ACTH by a pituitary tumor (Cushing's disease), by the secretion of cortisol by an adrenal tumor, or by the ectopic secretion of ACTH by a nonadrenal tumor. Determining the cause of the hypercortisolism entails performing multiple tests in a logical sequence (Fig. 35-21). Treatment should aim to cure the hypercortisolism and eliminate any tumor that threatens the patient's health while minimizing the chances of endocrine deficiency or long-term dependence on medications.

Clinical Manifestations

The signs and symptoms of Cushing's syndrome are widespread and diverse, and nearly every organ in the body is affected (Table 35-5). Although hypercortisolism is the most common hormonal symptom complex for adrenal cortical tumors, Cushing's syndrome is rare, with an estimated incidence of 10 per million population. It can occur in children. Weight gain is the most common symptom of patients with hypercortisolism. Obesity is usually truncal, and patients have thinning of extremities due to muscle wasting. Increased fat in the dorsal neck region combined with kyphosis secondary to osteoporosis gives the appearance of a "buffalo hump." The face is round and "moon" shaped (Fig. 35-22). Blood pressure is mildly increased secondary to excessive mineralocorticoid secretion. Striae appear as reddish-purple broad skin marks along the flank. Hirsutism consists of excessive fine hair on face, upper back, and arms. Virilization, including clitoromegaly, deep voice, and balding, and is more commonly associated with malignant adrenal cortical tumors. Glucose intolerance with hyperglycemia is mild, and patients seldom require insulin. Muscle weakness is common and can be exacerbated by hypokalemia. Menstrual irregularity or amenorrhea is one of the most common signs of hypercortisolism; men have decreased sexual drive and impotency. Dilatation of superficial blood vessels and thinning of subcutaneous tissue can give the face a ruddy appearance. Mental changes vary from mild depression to severe psychosis. Impaired immune function may cause opportunistic infections, including cryptococcosis, aspergillosis, nocardiosis, and *Pneumocystis carinii* infection. These infections add to morbidity and may be lethal. Children may be diagnosed by an arrest of normal growth and short stature.

Diagnosis

The early diagnosis of Cushing's syndrome depends on an awareness of the different signs and symptoms associated with the disorder and the ability to discriminate these findings. The initial workup is to establish the presence of hypercortisolism. The next step is to determine whether the hypercortisolism is "pituitary-dependent" or "pituitary-independent," and the last step is to determine the exact cause of the hypercortisolism (see Fig. 35-21). Diagnostic laboratory methods yield the correct diagnosis for the presence and the cause of the hypercortisolism in nearly every case. The best screening test for the diagnosis of hypercortisolism is the measurement of 24-h urinary free cortisol. Urinary excretion of free cortisol is directly proportional to the amount of free cortisol in the plasma. As the cortisol-binding globulin becomes saturated (plasma cortisol levels of 20 mg/dL), small increases in cortisol secretion produce exponential increases in urinary free cortisol. Another test for the determination of hypercortisolism is the single-dose dexamethasone suppression test. It works because of the absence of normal feedback that occurs in all forms of hypercortisolism. Normal individuals given 1 mg dexamethasone orally at 11:00 p.m. have plasma cortisol levels below 5 mg/dL at 8:00 a.m. the next day. Patients with hypercortisolism have levels higher than 5 mg/dL. This test has a few false-negative and a few false-positive results. However, a single- dose dexamethasone test with suppression and a normal level of urinary free cortisol (< 100 mg/day) exclude hypercortisolism.

Determination of plasma ACTH levels may be useful to discriminate the cause of hypercortisolism. Patients with primary adrenal tumors or primary pigmented micronodular adrenocortical hyperplasia have undetectable or low plasma ACTH levels. Pituitary tumors have intermediate plasma levels of ACTH, and ectopic ACTH-producing tumors have very high levels. Radioimmunoassays for ACTH in plasma have been unreliable because of platelet-associated enzymes that degrade ACTH and result in falsely low levels. Samples must be collected in prechilled tubes on ice to improve reliability. Urinary 17-ketosteroid levels can help to distinguish an adrenal adenoma from a carcinoma. Low levels (< 10 mg/day) suggest an adrenal adenoma, and very high levels (> 60 mg/day) suggest adrenal cancer or ectopic ACTH syndrome.

Three diagnostic tests are most useful in determining the cause of hypercortisolism: the CRH test, the dexamethasone suppression test, and the metyrapone test. In patients with an ACTH-secreting pituitary tumor (Cushing's disease), administration of corticotropin-releasing hormone 1 mg/kg body weight increases plasma levels of ACTH and cortisol, but in patients with alcoholism, depression, or other stress diseases the response is blunted. The CRH test also may distinguish a pituitary tumor from ectopic ACTH secretion. In patients with ectopic ACTH secretion, plasma ACTH and cortisol levels after CRH administration do not increase.

The dexamethasone suppression test is the most useful test in establishing the cause of hypercortisolism and in determining whether the cause is pituitary-dependent or pituitary-independent. In normal individuals urinary levels of 17-hydroxysteroid, or free cortisol, will be suppressed by low-dose (2 mg) dexamethasone. In patients with

hypercortisolism the urinary level of free cortisol is not suppressed by low-dose dexamethasone. High-dose dexamethasone (8 mg/day) will suppress urinary levels of free cortisol to less than 50 percent of baseline levels in patients with pituitary-dependent hypercortisolism (Cushing's disease), but it will not suppress levels in patients with primary adrenal cases of hypercortisolism or ectopic ACTH syndrome. The dexamethasone suppression test has an accuracy of approximately 95 percent. The metyrapone test, which causes stimulation of ACTH release in patients with pituitary Cushing's disease, can be complementary to the dexamethasone suppression test. Combining the tests results in greater accuracy rates than are obtained with the use of either test alone.

Radiologic Evaluation

Computed tomography (CT) of the sella turcica has been reported in various studies to detect a tumor in only 0 to 15 percent of patients with pituitary-dependent Cushing's disease, and minor abnormalities in 23 to 60 percent. Most ACTH-secreting pituitary tumors are microadenomas (< 5 mm). Pituitary magnetic resonance imaging (MRI) studies, even with gadolinium, have similar resolution. In patients with pituitary-dependent hypercortisolism, CT and MRI may be normal, but bilateral petrosal sinus sampling for ACTH concentrations detects the side with a tumor in most cases. Petrosal sinus sampling is the best method of differentiating a pituitary from an ectopic ACTH-producing tumor. The study requires bilateral sampling of the inferior petrosal sinus and peripheral veins for plasma ACTH levels before and after CRH. A petrosal sinus to peripheral plasma ACTH ratio of more than 3.0 after CRH administration correctly identifies patients with Cushing's disease (sensitivity is 100 percent), with few false positive results (specificity is 100 percent). Petrosal sinus sampling also provides correct localization of the ACTH-producing microadenoma in most patients. This is a relatively new and invasive study. It requires a significant degree of expertise by the radiologist, but it is the study of choice to diagnose and localize pituitary tumors in patients with Cushing's disease.

Adrenal CT can reliably distinguish cortical hyperplasia from tumor (Fig. 35-23). CT has great sensitivity (> 95 percent), but it lacks specificity. In a patient with Cushing's syndrome, early detection of an adrenal neoplasm by CT simplifies the workup. CT can be used to image the primary tumor, plus local and distant metastases in patients with adrenocortical cancer. Approximately 15 percent of patients with Cushing's syndrome have a primary adrenal neoplasm as the source of the hypercortisolism. Diagnosis of a unilateral adrenal tumor requires the detection of a normal adrenal gland on the contralateral side. Adrenal hyperplasia also may be detected if both glands appear enlarged. MRI can add specificity to the sensitivity of CT. MRI may be able to distinguish adrenal adenoma from carcinoma, and a pheochromocytoma may be recognized by the brightness of the lesion on T₂-weighted MRI. Adrenal adenomas appear darker than the liver on T₂-weighted imaging. Carcinomas, whether primary adrenocortical or metastatic, appear as bright as or slightly brighter than the liver on T₂-weighted imaging. Pheochromocytomas appear much brighter (3×) than the liver on T₂-weighted imaging. Different adrenal sequences have been used to better distinguish adrenal masses. Although some of these sequences have been useful, none is unequivocal, and each has been associated with exceptions and false results (Fig. 35-24).

Radioisotope imaging of the adrenals with labeled iodocholesterol, such as ¹³¹I-6-beta-iodomethyl norcholesterol, can be useful in distinguishing adrenal adenoma from hyperplasia. It can help in differentiating a benign cortical neoplasm (adenoma), which usually takes up iodocholesterol, from a malignant cortical neoplasm (carcinoma), which usually does not take up the tracer. This is not absolute, because adrenal cortical carcinomas may take up iodocholesterol and be "hot" on scan. Iodocholesterol scan can be the study of choice in patients with primary pigmented micronodular hyperplasia in which CT appears normal, but bilateral adrenal uptake confirms the diagnosis. It can be used to image ectopic rests of adrenal tissue in patients who have had bilateral adrenalectomy but still have evidence of hypercortisolism. The disadvantages of radioiodocholesterol scans include exposure to radiation, limited isotope availability, and poor imaging of malignant adrenal neoplasms.

Interpretation

Once the laboratory tests have confirmed endogenous hypercortisolism, the workup should pinpoint the cause. If an adrenal cortical neoplasm is the source of the hypercortisolism, the workup will show (1) a tumor on CT or MRI (see Figs. 35-23 and 35-24), (2) low plasma ACTH levels, and (3) no suppression of urinary cortisol with high-dose dexamethasone. If there is no tumor but there are low ACTH levels and lack of suppression with dexamethasone, primary micronodular adrenal cortical hyperplasia must be excluded by iodocholesterol scan or petrosal sinus sampling for ACTH levels. If there is an adrenal tumor, dexamethasone does not suppress urinary cortisol, and ACTH levels are consistently elevated, as are urinary catecholamines. Vanillylmandelic acid (VMA) and metanephrines must be measured to exclude an ACTH-producing pheochromocytoma.

In patients with ectopic ACTH syndrome, the studies should suggest: (1) bilateral hyperplasia of adrenals on CT, (2) elevated ACTH levels in plasma, (3) no cortisol suppression with high-dose dexamethasone, and (4) no evidence of lateralization of ACTH plasma levels on petrosal sinus sampling with CRH. Small-cell lung cancer is the most common ectopic ACTH tumor. Bronchial carcinoid tumors potentially are the most common curable ectopic ACTH tumor. Lobectomy is required because a significant number of patients will have nodal metastases. If the results are inconsistent, a sampling of petrosal sinus should be analyzed for ACTH levels to determine the diagnosis of ectopic ACTH syndrome.

If the patient has a pituitary tumor, the following results are expected: (1) bilateral hyperplasia of the adrenal glands, (2) mildly elevated plasma ACTH levels, (3) suppression with high-dose dexamethasone, and (4) lateralization with petrosal sinus sampling.

Adrenal Insufficiency (Addison's Disease)

Destruction of the adrenal cortex is the most common cause of primary adrenal insufficiency, but the adrenal medulla usually is spared. Adrenal insufficiency may be secondary to an abnormality in the hypothalamus or the pituitary, the so-called secondary adrenal insufficiency. Primary adrenal insufficiency can be caused by an autoimmune adrenalitis, tuberculosis, adrenomyeloneuropathy, fungal infections, AIDS, metastatic carcinoma, familial isolated glucocorticoid deficiency, and adrenal surgery to remove a cortisol-secreting tumor or both adrenal glands. Secondary adrenal insufficiency can be caused by pituitary tumor, craniopharyngioma, pituitary surgery or radiation, sarcoidosis, histiocytosis, empty sella syndrome, hypothalamic tumors; most commonly, it is iatrogenic secondary to long-term glucocorticoid administration. Primary or secondary adrenal insufficiency may lead to adrenal crisis perioperatively if the hypothalamic-pituitary-adrenal axis cannot generate an appropriate stress level of cortisol (see Fig. 35-20). Acute adrenal insufficiency is an uncommon clinical condition, but if it is untreated the consequences may be catastrophic.

Diagnosis

The symptoms and signs of primary and secondary adrenal insufficiency include: tiredness, weakness, mental depression, anorexia, weight loss, dizziness, orthostatic hypotension, nausea, vomiting, diarrhea, hyponatremia, hyperkalemia, anemia, leukocytosis, and eosinophilia. In secondary adrenal insufficiency, hyperpigmentation also may occur. Considering the possible diagnosis of acute adrenal insufficiency can be crucial in some critically ill patients.

Spontaneous adrenal insufficiency secondary to adrenal hemorrhage or adrenal vein thrombosis must be considered in a patient with upper abdominal or flank pain, abdominal rigidity, vomiting, confusion, and hypotension. To exclude the diagnosis of adrenal insufficiency in a patient, the best diagnostic study is the corticotropin (ACTH) stimulation test. In the short ACTH stimulation test 250 mg of cosyntropin is given intravenously or intramuscularly before 10 a.m.; plasma cortisol is measured before cosyntropin is given and 60 min after. The adrenal function is normal if the basal plasma level is at least 20 mg/dL and there is an increment of more than 7 mg/dL over basal levels at 60min after cosyntropin administration.

Management

If the diagnosis of acute adrenal crisis is suspected, hydrocortisone should be given intravenously (Fig. 35-25). To verify the diagnosis, blood levels of cortisol, glucose, sodium, potassium, blood urea nitrogen, and creatinine should be determined, and a complete blood count should be performed. Treatment begins before results of the studies have returned, with the administration of a 200-mg bolus of hydrocortisone intravenously, followed by a rapid infusion of glucose and saline solution to correct hypoglycemia and dehydration. Patients who have adrenal crisis will have a low plasma cortisol level (< 18 mg/dL), hyponatremia, hyperkalemia, leukocytosis with eosinophilia, azotemia, and hypoglycemia. Another 100 to 200 mg of hydrocortisone should be administered over the next 24 h. If the patient has acute adrenal crisis, there will be a rapid response to the intravenous hydrocortisone and fluid. The high doses of hydrocortisone should be continued for several days and eventually tapered to a maintenance dose (hydrocortisone 12 to 15 mg/m² body surface area per day). If the patient has primary adrenal insufficiency (bilateral adrenalectomy), then administration of mineralocorticoid will be required when maintenance doses of hydrocortisone are reached (fludrocortisone 0.1 to 0.2 mg/day). If there is no response to the hydrocortisone bolus, adrenal crisis is unlikely, steroid treatment should be discontinued, and another source for the symptoms should be sought.

Patients with primary or secondary adrenal insufficiency must receive glucocorticoid treatment during surgical procedures (Fig. 35-26). If there is any question about the need for replacement therapy for a patient who has been on therapeutic glucocorticoids previously, a corticotropin stimulation test can be used to determine whether perioperative glucocorticoids are necessary. The amount of perioperative steroid treatment depends on the magnitude of the surgical procedure. Patients undergoing major operations, such as a Whipple procedure, coronary artery grafting, thoracotomy, or esophagectomy, should receive 100 to 150 mg hydrocortisone before operation and 50 mg intravenously every 8 h for 2 days after surgery, then 25 mg intravenously every 8h for another 3 days. After this time, the usual maintenance daily dose can be initiated. Patients undergoing surgical procedures of moderate magnitude, such as cholecystectomy, appendectomy, joint replacement, colectomy, and hysterectomy, should receive hydrocortisone 50 to 75 mg intravenously before operation and then another 50 to 75 mg intravenously over the remainder of the day of the operation. Another 50 to 75 mg hydrocortisone is administered on postoperative day one, and the usual daily steroid dose may be given on postoperative day two. Patients undergoing procedures that can be done under local anesthesia, such as hernia repair or dental procedures, should receive hydrocortisone 25 mg intravenously preoperatively only, and the usual daily dose should be administered on the first postoperative day.

Patients undergoing surgical treatment of endogenous hypercortisolism require glucocorticoid replacement. Steroids are not given preoperatively because these patients are already hypercortisolemic. Instead, hydrocortisone 100 mg intravenously is given after the removal of the adrenal tumor, the pituitary tumor, or the second hyperplastic adrenal gland. Administration of hydrocortisone is continued at a dose of 50 mg intravenously every 8 h for 2 days, and then at a dose of 25 mg intravenously every 8 h for another 3 days. The final replacement dosage of hydrocortisone is 12 to 15 mg/m² body surface area per day. This dosage may be adjusted as necessary to prevent symptoms of adrenal insufficiency. In patients who have undergone unilateral adrenalectomy for a cortisol-producing adenoma, replacement hydrocortisone continues until the hypothalamic-pituitary-adrenal axis has recovered. A corticotropin stimulation test should be performed every 3 to 6 months to assess recovery of function. When a normal response to cosyntropin is detected, the replacement dose of hydrocortisone can be discontinued. Careful follow-up is necessary, because it takes 18 to 48 months for recovery in most patients.

Primary Hyperaldosteronism (Conn's Syndrome)

Diagnosis

Excessive secretion of aldosterone is the cause of hypertension in patients with primary hyperaldosteronism. The most common cause of primary hyperaldosteronism is an aldosteronoma (aldosterone-secreting tumor); second is *idiopathic adrenocortical hyperplasia (IAH)* in which excessive aldosterone is secreted; and least common is adrenocortical carcinoma. Primary hyperaldosteronism is diagnosed primarily by an elevated plasma level of aldosterone and a decreased plasma level of renin (see Fig. 35-18). This suggests that the pathologic process is independent of renin. Diagnosis of secondary hyperaldosteronism, which occurs with renal artery stenosis, cirrhosis, and conditions of decreased kidney perfusion, depends on the detection of concomitant elevations in plasma levels of renin and aldosterone (Table 35-6).

High blood pressure measurement, low serum levels of potassium (< 3.9 mEq/L), high plasma levels of aldosterone and low plasma levels of renin are important for the diagnosis of

primary hyperaldosteronism. The symptoms of primary hyperaldosteronism may include muscle weakness, muscle cramps, polyuria, and polydipsia. These symptoms probably caused by hypokalemia and can be ameliorated with medications. Systolic hypertension may not be present, but diastolic hypertension is seen commonly.

An important alternative in the differential diagnosis is essential hypertension that is being treated with potassium-wasting diuretics. When a patient is being evaluated for primary hyperaldosteronism, all diuretics and antihypertensive medications should be stopped and a 24-h urinary excretion of potassium measurement obtained. In most patients with primary hyperaldosteronism urinary excretion of potassium will be excessive (>30 mEq/24 h). Essential for the diagnosis of primary hyperaldosteronism is concomitant measurement of plasma levels of aldosterone and renin. Patients with primary hyperaldosteronism have elevated plasma levels of aldosterone and low levels of renin. The ratio of plasma level of aldosterone to renin activity usually is greater than 30. Confirmatory evidence for the diagnosis of primary hyperaldosteronism is based on an inability to reduce plasma aldosterone levels and raise plasma renin activity after administration of the antihypertensive medication captopril. In normal individuals and in patients with essential hypertension, captopril decreases plasma levels of aldosterone and increases plasma levels of renin. In patients with primary hyperaldosteronism, administration of captopril does not change plasma levels of aldosterone and renin. A post- captopril plasma aldosterone level above 15 ng/dL and an aldosterone-to- renin ratio higher than 50 are consistent with the diagnosis of primary hyperaldosteronism.

When the diagnosis of primary hyperaldosteronism has been established, the next consideration is the cause—whether the patient has idiopathic adrenocortical hyperplasia or an aldosteronoma (see Table 35-6). The determination of the cause is critical for treatment because drug therapy is indicated for one, and surgery for the other.

CT can accurately image a high proportion of aldosteronomas (75 to 90 percent) and also can image the contralateral adrenal gland (see Fig. 35- 23), but it may miss small tumors (1 cm or less). In patients with an aldosteronoma, the opposite adrenal cortex should appear thin and atrophied on CT. The CT identification of an adenoma in one adrenal gland and a thin-appearing contralateral gland is pathognomonic for aldosteronoma.

Iodocholesterol scans with ¹³¹I-beta-iodomethyl-19-norcholesterol also can image a high proportion of aldosteronomas (90 percent). An advantage of this study over CT is that the presence of increased uptake suggests that the imaged glands are functional and responsible for the excessive steroid production. CT is more sensitive and more widely available and uses less radiation per study than iodocholesterol, but in certain patients with equivocal studies, iodocholesterol scans can be helpful. In patients with IAH as a cause of hyperaldosteronism, the iodocholesterol scan shows symmetrical uptake in both adrenal glands, but in patients with aldosteronoma, the study shows uptake only by the tumor.

Some patients with hyperaldosteronism have equivocal results in both CT and iodocholesterol scans. In these patients the study of choice for distinguishing between a tumor and hyperplasia is sampling of the adrenal veins for aldosterone (Fig. 35-27). The procedure is performed by simultaneous selective catheterization of adrenal veins and a peripheral vein. Plasma levels of aldosterone and cortisol are measured at each site before and after administration of ACTH. Aldosteronomas make aldosterone in response to ACTH. A unilateral elevation of aldosterone level or of the aldosterone-to-cortisol ratio indicates the presence of an aldosterone-secreting adenoma. Bilateral high levels of aldosterone that are similar and that are higher than peripheral levels are consistent with IAH. Adrenal venous sampling is more sensitive than CT in prospective studies.

Treatment

The treatment of primary hyperaldosteronism depends on the cause. Idiopathic adrenal hyperplasia is best managed medically with spironolactone, nifedipine, and/or amiloride, in conjunction with other antihypertensive drugs. Because aldosterone-secreting tumors usually are small and benign, the method of choice for surgical resection of these tumors is laparoscopic adrenalectomy. The laparoscopic procedure is associated with less pain, less morbidity, and shorter convalescence than open approaches, including the posterior approach. Prospective randomized studies have not been performed, but early results suggest that the tumor resection is similar and the morbidity markedly less.

It is important to know, however, whether the results of resection of an aldosteronoma have been entirely satisfactory. A high proportion of patients initially become normotensive and normokalemic, but 20 to 30 percent of patients develop recurrent hypertension within 2 to 3 years. It is hypothesized that prolonged preoperative hypertension from an aldosteronoma may alter renal function, which then leads to persistent postoperative hypertension despite complete resection of the tumor. Few cases of recurrent benign aldosteronomas have been described, and aldosterone-producing adrenocortical carcinomas are very rare (2percent of all carcinomas). Surgical removal of aldosteronoma clearly is indicated and in most patients is beneficial, if not curative.

Hypoaldosteronism

Hypoaldosteronism is a selective deficiency of aldosterone secretion without alteration in cortisol production. It results in postural hypotension and persistent severe hyperkalemia that may be associated with muscle weakness and cardiac arrhythmias. Hypoaldosteronism can result from inborn errors in aldosterone biosynthesis, failure of the zona glomerulosa secondary to an autoimmune adrenal disease in association with a critical illness, altered function of the renin-angiotensin system with hyporeninemic, hypoaldosteronism, unilateral adrenalectomy for an aldosterone-producing adenoma, and drug inhibition of aldosterone.

Hypoaldosteronism rarely has been reported after surgical excision of an aldosteronoma. After removal of an aldosterone-producing tumor, it is not routine practice to supplement with mineralocorticoids. However, some patients may develop significant hyperkalemia and hypotension lasting several days to several months after adrenalectomy. These patients should be treated with fludrocortisone (0.1 to 0.2 mg/day). This treatment will totally correct the fluid and electrolyte abnormalities and normalize the blood pressure. This treatment should be continued for several months, but it is not permanently necessary.

Adrenogenital Syndrome

Adrenogenital syndrome is commonly associated with congenital adrenal hyperplasia (CAH). CAH is a group of inherited diseases that are caused by defective activity in one of five enzymes that contribute to the synthesis of cortisol from cholesterol in the adrenal cortex (see Fig. 35-17). The term *adrenal hyperplasia* comes from the adrenal enlargement under the influence of ACTH in an effort to compensate for inadequate cortisol synthesis. The term *adrenogenital syndrome* refers to the associated finding of ambiguous external genitalia because of an incidental deficiency or excess production of adrenal androgens. Each enzyme deficiency produces a characteristic alteration in the ratio of precursor hormone to hormone that helps with the diagnosis. The different hormonal abnormalities are associated with different clinical features, including abnormal development of the external genitalia, pseudohermaphroditism, abnormal serum levels of sodium and potassium, abnormal values for blood pressure, and abnormal somatic growth.

The most common form of CAH is caused by a deficiency of the cytochrome P-450 enzyme 21-hydroxylase (over 90 percent). The remainder are caused by a deficiency of the following enzymes in order of prevalence: 11-hydroxylase, 17-hydroxylase, 17,20-lyase, and 3b- hydroxysteroid dehydrogenase (see Fig. 35-17). The mode of inheritance is autosomal recessive. The molecular genetics of most of the enzyme deficiencies have been identified, e.g., the CYP21 structural gene encoding 21-hydroxylase is part of the HLA complex on chromosome 6p21.3. The treatment for these conditions is glucocorticoid replacement therapy given in daily doses. Surgical therapy may be indicated for the ambiguous genitalia; this involves clitoroplasty and vaginoplasty in virilized females. Improved surgical techniques have provided superior results in single staged procedures by experienced surgeons.

Patients with adrenocortical carcinoma or adenoma may present with signs and symptoms of excessive sex hormone secretion that may be combined with clinical features of hypercortisolism. In children the signs of increased androgen production include: rapid growth, premature development of pubic and facial hair, acne, genital enlargement, increased muscle mass, and deep voice. The signs of increased estrogen production include gynecomastia in boys and precocious breast enlargement and/or vaginal bleeding in girls. In women the signs of excess androgen production include hirsutism, acne, amenorrhea, infertility, increased muscle mass, deep voice, and temporal balding. In men, hyperestrogenism presents with gynecomastia, decreased sexual drive, impotence, and infertility. In women, hyperestrogenism presents primarily with irregular menses in premenopausal women, and dysfunctional uterine bleeding or vaginal bleeding in postmenopausal women. The workup requires measurement of 24-h urinary 17- ketosteroids, 17-hydroxysteroids, urinary free cortisol, and, depending on virilization or feminization, serum determination of testosterone or estrogen. Virilization secondary to an adrenal neoplasm usually indicates adrenal cortical carcinoma. In the evaluation of a patient with virilization, CT or MRI of both adrenals is indicated to rule out an adrenal neoplasm.

Adrenal Mass

High-resolution CT scans have resulted in a new diagnostic problem: an asymptomatic adrenal mass seen on CT. Unexpected adrenal masses are seen in 0.6 percent of abdominal CT scans. The majority of these are benign, nonfunctional adrenal cortical adenomas, which occur in nearly 10 percent of autopsies.

A flow diagram for a suggested evaluation of an incidental adrenal mass (“incidentaloma”) is given in Fig. 35-28. Two questions arise: is it functional or is it cancer?

The surgeon should obtain a careful history and physical examination, including blood pressure, signs of Cushing's syndrome, hypertension, virilization, feminization, change in menstrual history, and evidence of occult malignancy (stool guaiac, Pap smear, anemia). Laboratory evaluation should include a 24-h urine collection for free cortisol, vanillylmandelic acid (VMA), metanephrines, and catecholamines. The 24-h urinary level of free cortisol is indicated to rule out Cushing's syndrome, and the urinary catecholamines to rule out a pheochromocytoma. The serum potassium concentration is used to exclude an aldosteronoma. Plasma levels of aldosterone and renin should be measured in any patient with hypertension and hypokalemia. Hormonal screening for an excess of androgens or estrogens is indicated only in patients with clinical signs suggestive of these disorders.

The size of an adrenal mass on CT is an important determinant of a potentially malignant tumor. Adrenal cortical carcinomas generally are larger than 5 cm in diameter, and benign lesions are less than 5 cm, but a smaller lesion should not be totally ignored. Early diagnosis may lead to discovery of a smaller adrenal cortical carcinoma that can lead to better prognosis and survival. CT can only distinguish benign from malignant neoplasms by criteria such as size, direct invasion, and detection of distant metastases. MRI of the adrenal glands has similar resolution to CT (see Fig. 35-24) but may be able to distinguish among adrenal cancer, metastases, pheochromocytoma, and adenoma on the basis of characteristics seen with different sequences.

Fine-needle aspiration for cytology of an adrenal mass has limited ability to differentiate benign from malignant primary adrenal tumors. Fine-needle aspiration might be catastrophic in a patient with an unsuspected pheochromocytoma, and therefore normal levels of urinary catecholamines are necessary to exclude a pheochromocytoma before needle biopsy is performed. In patients with suspected metastatic disease to the adrenal gland or lymphoma, needle aspiration may be helpful. In patients with known primary cancers, the presence of adrenal metastases can be confirmed by aspiration cytology. Because it cannot distinguish between benign and malignant primary adrenal neoplasms, fine-needle aspiration cytology should not be performed unless adrenal metastases are suspected.

CT and biochemical assessment should be performed on each patient with an incidentaloma. The adrenal mass diameter is carefully measured; a size greater than 5 cm is an indication for surgical resection. If the tumor is hormonally functional, adrenalectomy also is indicated. If the patient has a history of cancer, fine-needle aspiration may be considered to exclude adrenal metastases. If the mass is smaller than 5 cm and nonfunctional, a 6-month follow-up CT examination is indicated to reevaluate the mass's size. If the mass has grown, then surgical excision is necessary. If there is no change, then the mass is likely to be an adenoma, and subsequent CT scans are unnecessary.

Treatment of Adrenal Cortical Neoplasms

Adenoma

The indication for resection of an adrenal adenoma is excessive hormone production. The definitive treatment is complete removal of the adrenal gland with the adenoma. The laparoscopic approach is becoming the method of choice to remove benign adrenal tumors because it can achieve complete resection with less pain and morbidity. In patients who are undergoing resection of an adrenal tumor that produces cortisol, steroid replacement during and after surgery is required, but mineralocorticoid replacement is not. Postoperative glucocorticoid replacement (12 to 15 mg hydrocortisone/m² body surface area) may be necessary for as long as 2 years. After removal of an aldosteronoma or a virilizing adenoma, hormone replacement therapy usually is not necessary (see section [Hypoaldosteronism](#)). Surgical resection of an adenoma is curative. Larger lesions weighing 50 to 100 g that appear benign histologically (no mitoses and no vascular invasion) need careful long-term follow-up to exclude carcinoma. Laparoscopically resected tumors require close long-term follow-up to eliminate the possibility of spillage of tumor cells resulting from handling the gland during the procedure.

Carcinoma

Surgical resection is the only curative treatment for carcinoma of the adrenal cortex. If the carcinoma is intimately associated with the kidney, liver, or diaphragm on the right, or with pancreas on the left, removal of part or all contiguous structures may be necessary at the time of operation. The surgeon needs adequate imaging of the extent of disease, which can be achieved by CT or MRI. CT or MRI should include the chest to rule out metastatic disease above the diaphragm. If the right adrenal gland is involved and the inferior vena cava is compressed, an inferior vena cava contrast study or caval ultrasonography is useful to assess tumor extension into the vena cava. If resection of one kidney is indicated, an intravenous pyelogram or an intravenous contrast CT study is necessary to verify that the contralateral kidney is functioning. A complete bowel preparation is used in case the tumor is found to invade bowel. Though patients with hypercortisolism have impaired healing, adrenal tumor resection can be performed with acceptable morbidity and an operative mortality of 3 percent.

The surgical staging of adrenal carcinoma is: stage I, tumor smaller than 5 cm without local invasion, nodal, or distant metastases; stage II, same as stage I except that tumor is larger than 5 cm; stage III, tumor with local invasion or positive lymph nodes; and stage IV, tumor with local invasion and lymph nodes or distant metastases. Most patients (70 percent) present with stage III or IV disease.

The definitive initial treatment for adrenal carcinoma in all stages of disease, including locally aggressive stage III disease, is en bloc resection. This procedure may require a combined thoracoabdominal approach. Surgical resection of localized disease can be curative. Some recommend adjuvant mitotane therapy after surgical resection to improve survival, but this therapy is not proved. If complete resection of all tumor cannot be achieved, tumor debulking should be attempted to decrease the amount of cortisol-secreting tissue and to minimize complications due to tumor mass. In patients who undergo definitive resection steroid hormone levels should be monitored postoperatively. CT and MRI also are used to detect local recurrences or pulmonary metastases. If a recurrence is detected, it can be removed surgically. Prolonged disease-free intervals have been reported after resection of hepatic, pulmonary, and cerebral metastases from adrenal cortical carcinoma. When complete resection of recurrent tumor or metastases is not possible, tumor debulking may be useful in the case of selected hormonally active, slow-growing tumors. Palliation of bony metastases may be achieved by radiation therapy. Abdominal radiation therapy is not useful in reducing local recurrences.

Patients with recurrent or metastatic adrenal cortical carcinoma usually are treated with mitotane. Mitotane has direct effects on steroid metabolism and usually is associated with a decrease in urinary 17-hydroxysteroids and 17-ketosteroids and improvement in the signs and symptoms of hypercortisolism, but it seldom provides objective antitumor responses. Other chemotherapeutic agents also have been ineffective, though responses have been reported with doxorubicin, alkylating agents, cisplatin, and etoposide.

Treatment of adrenal cancer is ineffective. Surgical cure may be feasible only in stage I or stage II tumors confined to the adrenal gland. In these patients the mean duration of survival is 5 years. In patients with invasion of contiguous structures at presentation, median survival is 2.3 years. For all patients, the 5-year survival rate is 10 to 35 percent, reflecting the fact that most patients present with locally advanced or distant disease. Most clinicians still recommend aggressive surgical re-resection of locally recurrent or metastatic cancer, but even with this aggressive intervention the 5-year survival rate is 10 to 20 percent. Future challenges for better treatment of adrenal cortical carcinoma include earlier diagnosis and better chemotherapy than mitotane.

Ectopic ACTH Syndrome

The diagnosis of ectopic ACTH syndrome requires the presence of Cushing's syndrome and bilateral adrenal hyperplasia, but no evidence of a pituitary tumor. In ectopic ACTH syndrome ACTH is produced by a tumor that is not within the pituitary gland. Early indication of the disorder is hypercortisolism and severe hypokalemia (potassium < 3.3 mEq/L). The laboratory diagnosis depends on the measurement of high plasma levels of ACTH and cortisol that are unaffected by high-dose dexamethasone and do not increase with CRH. Petrosal sinus sampling for ACTH detects low plasma levels of ACTH in the venous blood draining the pituitary gland that do not increase with CRH. After the diagnosis has been made, the primary therapeutic goal is to locate and eradicate the neoplasm that is secreting ACTH. Two difficulties are localizing the tumor in some patients, and treating an aggressive, metastatic malignant tumor in others.

Ectopic ACTH-producing tumors in approximate order of decreasing incidence are: oat-cell or small-cell lung cancer, carcinoid tumor of the bronchus, thymic carcinoid, pancreatic islet cell tumor, medullary carcinoma of the thyroid gland, pheochromocytoma, midgut carcinoid, ovarian adenocarcinoma, pancreatic cystadenoma, and adenocarcinoma of unknown site. Procedures indicated to localize the ACTH producing tumor include chest and abdominal CT, chest and abdominal MRI, urinary catecholamines determination to screen for pheochromocytoma, measurement of plasma calcitonin level to rule out medullary thyroid carcinoma, and inferior petrosal sinus sampling with CRH in patients in whom a pituitary ACTH-secreting tumor cannot be excluded by other tests. Any suspicious finding in the chest or abdomen can be unequivocally confirmed by fine-needle aspiration and radioimmunoassay for ACTH on the aspirate.

The goal of surgery for patients with ectopic ACTH-producing tumors is to locate and resect (unless it is a small-cell lung carcinoma) the neoplasm. A true cancer resection (lobectomy with lymph nodes) is indicated for patients with bronchial carcinoid tumors, because one-half have lymph node metastases. These tumors are treatable, and approximately 75 percent of patients are cured by surgical resection. The proper therapy for ACTH-producing neoplasms depends on the diagnosis (the exact tumor type that is producing ACTH) and the extent of disease. Any of these tumors may be malignant and metastasize. In some patients with ectopic ACTH production, the entire tumor burden cannot be removed, and therapy should be directed toward correcting the life-threatening metabolic and hormonal abnormalities. Hypokalemia and excess mineralocorticoid activity can be managed with potassium supplementation and spironolactone. Hypercortisolism may be managed with metyrapone, amino glutethimide, or mitotane. Bilateral adrenalectomy might be indicated for patients who have tumors that cannot be identified despite numerous radiologic studies and patients who have stable but unresectable metastatic disease whose hypercortisolism cannot be controlled medically. Bilateral adrenalectomy can be performed with the posterior approach via two separate hockey-stick incisions or with the lateral laparoscopic approach through two separate sequential procedures. Patients undergoing bilateral adrenalectomy need glucocorticoid and mineralocorticoid replacement therapy postoperatively.

ADRENAL MEDULLA

Physiology

In responding to severe stress and confrontation (the so-called fight-or-flight response), the autonomic nervous system plays a critical role. The autonomic nervous system consists of the parasympathetic nervous system, which uses acetylcholine as a neurotransmitter, and the sympathoadrenal system, which uses catecholamines. The sympathoadrenal system consists of a sympathetic neuronal component with norepinephrine as the main neurotransmitter at sympathetic neuroeffector junctions, and an adrenomedullary secretory hormone epinephrine, the main hormone secreted into the bloodstream. This system influences cardiovascular, metabolic, and visceral activity in the resting organism and determines the physiologic consequences of every motion and emotion.

The chromaffin tissue of the adrenal medulla and the sympathetic ganglia originate from the neural crest (see Fig. [35-15](#)). The term *chromaffin* refers to the appearance of intracellular granules, which turn brown when stained with oxidizing agents. The small amounts of epinephrine secreted by the adrenal medulla into the bloodstream is not necessary to maintain blood pressure, but adequate release of norepinephrine at sympathetic nerve endings is critical for maintenance of normal blood pressure, especially during periods of upright posture. The major cellular supply of catecholamines is stored in secretory granules in both nerve endings and the adrenal medulla. A smaller pool of catecholamines is concentrated at the peripheral portion of sympathetic neurons and is released during stimulation. After sympathetic nerve stimulation some of the norepinephrine may spill into the circulation. It is converted in the liver by two enzymes, catechol O-methyltransferase and monoamine oxidase to normetanephrine and vanillylmandelic acid (VMA). These metabolites and the small fraction of unmetabolized norepinephrine that escapes into the bloodstream are excreted in the urine. After sympathetic nerve stimulation, most of the norepinephrine is taken up again by the neuron and stored in vesicles for future use. Monoamine oxidase in the cytoplasm of the neuron inactivates any norepinephrine that leaks out. Epinephrine released into the bloodstream from the adrenal medulla is metabolized primarily by the liver into metanephrine and VMA. These metabolites, plus small amounts of unmetabolized epinephrine, are excreted in the urine. Release of norepinephrine from neurons is primarily responsible for controlling blood pressure during upright posture. Release of epinephrine from the adrenal medulla is primarily responsible for the systemic effects of catecholamines observed during episodes of severe stress.

Pheochromocytoma

Pheochromocytomas arise from chromaffin cells that are widespread and associated with sympathetic ganglia throughout the body during fetal life. After birth, most chromaffin cells degenerate. However, most chromaffin cells in the adrenal medulla persist. This may be why 85 to 90 percent of pheochromocytomas are in the adrenal medulla, but pheochromocytomas can arise wherever there is a sympathetic ganglion. Extraadrenal pheochromocytomas may occur in any part of the body, including in the carotid body, in the heart, along the aorta (both thoracic and abdominal), and in the urinary bladder. The most common extraadrenal location is the organ of Zuckerkandl, which is near the origin of the inferior mesenteric artery to the left of the aortic bifurcation (Fig. [35-29](#)). Extraadrenal pheochromocytomas are often malignant. Bilateral adrenal pheochromocytomas usually occur in familial syndromes, including *multiple endocrine neoplasia (MEN)* type IIA and type IIB. The right adrenal gland is more commonly affected than the left.

Pheochromocytomas from symptomatic hypertensive patients usually measure 3 to 5 cm in diameter and weigh approximately 100 g. They appear tan to gray in color, and are soft and smooth. Larger tumors may be cystic from necrosis, and they may have calcification. Microscopically, pheochromocytomas are arranged in cords or alveolar patterns. Tumors generally are clearly separated from the adrenal cortex by a thin band of fibrous tissue.

The pathologic distinction between benign and malignant pheochromocytomas is not clear. Most reports suggest that only 10 percent of pheochromocytomas are malignant, but some suggest that 23 to 46 percent of pheochromocytomas are malignant. These different results may partly reflect the referral patterns of different institutions, but they may indicate a true higher incidence of malignancy. Malignant pheochromocytomas tend to be larger and weigh more, although this is not an absolute pathologic criterion. The only absolute criteria for malignancy are the presence of secondary tumors in sites where chromaffin cells are not usually found, and the identification of visceral metastases.

Associated Syndromes

Approximately 10 percent of pheochromocytomas occur as part of an inherited condition. Pheochromocytomas are associated with endocrine and nonendocrine inherited disorders.

Bilateral adrenal medullary pheochromocytomas are components of multiple endocrine neoplasias type IIA and type IIB. Familial pheochromocytoma also has been described. Affected individuals have bilateral adrenal pheochromocytomas and no other manifestation of MEN syndromes. In other families, extraadrenal pheochromocytomas have been reported in the same location in affected individuals. Pheochromocytomas occur in approximately 25 percent of patients with von Hippel-Lindau's disease and in less than 1 percent of patients with neurofibromatosis and von Recklinghausen's disease.

Clinical Manifestations

Pheochromocytomas secrete excessive amounts of unregulated catecholamines, and usually patients have significant hypertension. Pheochromocytomas can cause anxiety attacks, episodic hypertension, or sustained hypertension. The patient classically describes “spells” of paroxysmal headaches, pallor, palpitations, hypertension, and diaphoresis (Table 35-7). Some patients with pheochromocytoma have only mild labile hypertension, while others have sudden death secondary to a myocardial infarction, or cerebrovascular accident from uncontrolled episodic hypertension. Patients with pheochromocytoma may have signs of chronic hypovolemia, such as orthostatic hypotension, or lactic acidosis secondary to excessive alpha-adrenergic stimulation and vasoconstriction. The majority of these patients have mild weight loss, but obesity does not exclude the diagnosis. Diabetes mellitus may be induced by chronic excessive catecholamine secretion. Fever of unknown origin has been caused by an unexpected pheochromocytoma. Pheochromocytomas may produce multiple hormones like catecholamines and ACTH. In these patients symptoms of both pheochromocytoma and hypercortisolism are present. Other conditions may mimic pheochromocytoma and are included in the differential diagnosis (Table 35-8).

Diagnosis

The diagnosis of pheochromocytoma is based on measuring excessive excretion of catecholamines and their metabolites in the urine. The best study is a 24-h urine collection for catecholamines, metanephrine, and vanillylmandelic acid. The 24-h urinary level of VMA has the greatest sensitivity (97 percent) and the best specificity (91 percent) (Fig. 35-30). Because pheochromocytomas secrete catecholamines episodically, 24-h urinary measurements are more sensitive than a plasma level, which reflects only the level at the moment at which blood is sampled. Elevated levels of norepinephrine more commonly occur with extraadrenal pheochromocytomas because of the physiology of catecholamine secretion, and elevated levels of epinephrine commonly occur with adrenal pheochromocytomas.

The clonidine suppression test can be used as an additional study to determine whether a patient has a pheochromocytoma and is recommended in the diagnosis of a pheochromocytoma when plasma catecholamine levels are mildly elevated (500 and 2000 pg/mL). In normal individuals and in patients with idiopathic hypertension, clonidine suppresses plasma levels of catecholamines, and in patients with pheochromocytoma it does not. Few false-positive or false-negative results have been reported. The author recommends relying on urinary levels of catecholamine metabolites.

Localization Studies

CT and MRI have been effective for localization of pheochromocytomas (Fig. 35-31). These studies are noninvasive and sensitive and are able to detect tumors approximately 1 cm in diameter. MRI may be more specific because of its ability to distinguish various types of adrenal tumors on the basis of imaging characteristics, but CT has several advantages over MRI. CT scans are less expensive and more available than MRI. A major advantage of MRI is that there is no radiation exposure, and it can be used to image pheochromocytomas during pregnancy. MRI has accurately imaged an intrapericardial pheochromocytoma and distinguished it from the cardiac chambers.

A nuclear medicine imaging study for the localization of pheochromocytomas is scanning with ¹³¹I labeled metaiodobenzylguanidine (MIBG). MIBG is similar to norepinephrine and is taken up and concentrated in catecholamine-synthesizing tissue. The sensitivity of MIBG scanning is between 78 and 86 percent in sporadic pheochromocytoma, approximately 91 percent in malignant pheochromocytoma, and 94 percent in familial pheochromocytoma. MIBG scanning is safe, noninvasive, and efficacious for the localization of pheochromocytomas, including those arising in extraadrenal sites, and malignant disease. False-positive results with MIBG scintigraphy are rare (tumors such as medullary thyroid carcinoma and neuroblastoma may take up the tracer), which accounts for the high specificity (98 to 100 percent) of the study. False-negative results can occur and have an incidence of 13 to 20 percent, which lowers sensitivity. False-negative results may be more common in patients with multiple tumors and metastatic disease.

Treatment

Preoperative Preparation

When the diagnosis of pheochromocytoma has been made and the tumor has been localized, an essential component of preoperative preparation is alpha-adrenergic blockade. Patients receive phenoxybenzamine 10 mg orally two or three times a day (see Fig. 35-30). If tachycardia develops (heart rate > 100 beats/min), beta-adrenergic blocking agents (propranolol) are indicated. Propranolol should never be started before alpha blockade, because unopposed vasoconstriction may worsen hypertension. Phenoxybenzamine increases the total blood and plasma volume in patients with pheochromocytoma. Lactic acidosis can occur in patients with pheochromocytoma, and therefore measurement and correction of arterial blood pH should be performed before the induction of anesthesia.

Alpha-methyltyrosine is another drug that is useful for preoperative preparation of these patients. It is a competitive inhibitor of tyrosine hydroxylase, the rate-limiting step in catecholamine biosynthesis. It reduces catecholamine production by 50 to 80 percent in patients with pheochromocytoma. The usual dose of 250 mg four times a day may be increased to a maximum dose of 3 to 4 g per day. The calcium antagonist nifedipine with phenoxybenzamine or nicardipine alone (60 to 120 mg/24 h) has been used to control hypertension in patients with pheochromocytoma. Both of these newer drug strategies appear to work as well as phenoxybenzamine.

Intraoperative Management

If a patient with pheochromocytoma is elderly or has had cardiac complications, right heart catheter monitoring should be maintained during surgery. On the morning of surgery, an arterial catheter and peripheral intravenous catheters are inserted. Arterial CO₂ is measured to rule out acidosis. Despite preoperative preparation, during manipulation of the tumor marked changes in blood pressure can occur. Hypertensive episodes are controlled with alpha-adrenergic blocking agents, such as phentolamine (Regitine) or smooth muscle relaxers, such as sodium nitroprusside. Nitroprusside usually is chosen because of its rapid onset and short duration. With the use of intravenous nitroprusside, blood pressure can be continuously titrated to acceptable levels. The use of preoperative preparation and intraoperative regulation of blood pressure with nitroprusside greatly reduces operative morbidity and mortality.

The operation is performed via a transabdominal approach, with a bilateral subcostal or a long midline incision. Preoperative localization studies can effectively guide the exploration, but the entire abdomen is nevertheless carefully visualized and palpated. Some have argued that localization procedures are so sensitive and specific that focused direct approaches may be preferred. Small intraadrenal pheochromocytomas have been removed using laparoscopic techniques. Laparoscopic procedures decrease pain and shorten the recovery time, but laparoscopic adrenalectomy is still considered experimental for resection of pheochromocytoma.

Extraadrenal pheochromocytomas can be difficult to find (see Fig. 35-29). The most common locations of intraabdominal extraadrenal pheochromocytomas are in the hilar region of the kidneys, and in the chromaffin tissue along the aorta from the celiac axis to the aortic bifurcation (Fig. 35-32). The organ of Zuckerkandl near the inferior mesenteric artery is the most common extraadrenal location, and pheochromocytomas have been removed from the urinary bladder. Multiple locations, metastatic potential, and multiple tumors support the need for a complete exploration of the entire abdominal cavity. The “rule of tens” may be of value in the management of pheochromocytomas; 10 percent are malignant, 10 percent are extraadrenal, and 10 percent are bilateral. Bilateral adrenal pheochromocytomas are widely believed to occur primarily as part of a familial syndrome. In patients with MEN type II, some suggest that nearly 100 percent have or will develop bilateral benign adrenal medullary pheochromocytomas, and others suggest that the incidence of bilaterality, although high, may be significantly less (70 percent). The question of whether or not to remove both adrenals is controversial; the author recommends removing only the adrenal gland with an identifiable tumor.

Malignant Pheochromocytoma

Malignant pheochromocytomas are thought not to occur in MEN syndromes and are present in approximately 10 percent of patients with pheochromocytoma. Some reports indicate that substantially more than 10 percent of sporadic pheochromocytomas may be malignant. Pathologic analysis has not been helpful in predicting which tumors are malignant. Patients who develop metastases might not develop them for many years after surgery (Fig. 35-33). After removal of a pheochromocytoma, most recommend lifetime follow-up with measurement of blood pressure and urinary levels of catecholamines. It appears that with careful follow-up the true incidence of malignant pheochromocytoma may be greater than 10 percent.

The basic principles in treatment of malignant pheochromocytoma are to resect recurrences or metastases whenever possible and to treat hypertensive symptoms by catecholamine blockade. Painful bony metastases respond well to radiotherapy. Localized or solitary soft-tissue masses, even when metastatic to the liver or lung, should be resected surgically. Combination chemotherapy with cyclophosphamide, vincristine, and dacarbazine, a regimen that is effective for metastatic neuroblastoma, has also been beneficial for patients with metastatic pheochromocytoma. The regimen has a 60 percent response rate, including some complete responses. Survival data of patients with malignant pheochromocytoma are difficult to obtain because of the rarity of the tumor. In two large series, the 5-year survival rates for patients with malignant pheochromocytoma were 36 and 60 percent.

Neuroblastoma

Neuroblastoma is the fourth most common pediatric malignancy. The median age at diagnosis is 2 years. Half of all malignancies diagnosed within the first year of life are neuroblastomas, and the mortality of children with neuroblastoma diagnosed in the first year of life is lower than that of older children. Neuroblastoma has been associated with genetic diseases, including neurofibromatosis, Beckwith-Wiedemann syndrome, and trisomy 18. Neuroblastoma has a biologic characteristic unique among human cancers, namely, the capacity to spontaneously differentiate and regress. It is believed that some children with this life-threatening disease are spontaneously cured by their immune system, and these malignant tumors then become benign ganglioneuromas.

Neuroblastoma is believed to arise from the embryonic neural crest. It is called a small round blue cell tumor because it is composed of dense nests of cells separated by fibrovascular bundles. The characteristic electron microscopy finding is dense core secretory granules. Neuroblastoma may be difficult to distinguish from other small round blue cell tumors of childhood, which include Ewing's sarcoma, lymphoma, rhabdomyosarcoma, and other peripheral nerve tumors. An important adjunct to the diagnosis of neuroblastoma is the detection of tumor markers in the serum and urine. The most commonly used marker is urinary excretion of catecholamines. Elevated 24-h urinary levels of homovanillylmandelic acid and VMA are detected in 65 and 90 percent, respectively, of patients with neuroblastoma. Levels of catecholamines in urine are correlated with prognosis, i.e., higher levels are found in patients with more extensive disease.

Neuroblastoma can occur anywhere along the sympathetic nervous system. The most common sites of tumor are in the adrenal gland (40 percent), the paraspinal ganglia (25 percent), the thorax (15 percent), and the pelvis (5 percent). Thoracic primary tumors are more common in children younger than 1 year of age. Metastatic tumor is detected in approximately 50 percent of infants and 67 percent of older children at the time of diagnosis. Metastases usually involve lymph nodes, bone marrow, bone, liver, and subcutaneous tissue. Pulmonary metastases are rare.

Clinical Manifestations

The most common presenting sign and symptom is an abdominal or flank mass, which must be distinguished from Wilms' tumor. Thoracic neuroblastoma commonly presents as a posterior mediastinal mass identified on a routine chest radiograph. Thoracic tumors may cause signs and symptoms of respiratory distress or spinal cord compression. Neuroblastoma in the neck presents with a cervical mass; the diagnosis can be supported by the signs of Horner's syndrome, suggesting that the mass involves the sympathetic ganglia. Pelvic neuroblastoma commonly involves the organ of Zuckerkandl and may present with urinary symptoms or as a pelvic mass.

There are unusual symptoms and signs associated with neuroblastoma: (1) Some neonates with neuroblastoma have subcutaneous blue-tinged nodules, the so-called "blueberry muffin" sign of neuroblastoma. (2) Some have acute cerebellar and truncal ataxia, which is associated with neurologic involvement but favorable outcome. (3) Some present with severe secretory diarrhea similar to the adult Verner-Morrison syndrome; in these patients the neuroblastoma is producing vasoactive intestinal polypeptide (VIP), which can be measured in the plasma. (4) Neuroblastoma may present with periorbital metastases that cause proptosis and ecchymosis; this is secondary to invasion and involvement of the retro-orbital tissues. (5) Some patients have diffuse bone marrow involvement with neuroblastoma without an obvious primary tumor.

Radiology and Staging

CT is the best imaging study for patients with neuroblastoma and should be performed to determine the extent of disease. CT detects abdominal neuroblastoma with a high degree of accuracy, and calcifications are identified in most tumors. MRI has similar accuracy to CT, and it may be able to image intraspinal extension and vascular invasion more effectively.

Useful nuclear medicine studies include the technetium diphosphonate bone scan and ¹³¹I-MIBG because of its preferential uptake by cells with adrenergic secretory granules. Bone scan is the best study to determine bony involvement, and MIBG scan can image some metastatic disease that is not expected.

The goal of the imaging studies is to stage the total extent of tumor. Prognosis and treatment depend on the stage of disease at presentation. The most widely used staging system is the international staging classification for neuroblastoma (Table [35-9](#)).

Treatment

Treatment depends on the stage of disease. For patients with localized disease (stage 1 and 2), surgical resection is the mainstay of treatment. In patients with cervical, mediastinal, and pelvic tumors complete resection usually is feasible. Most localized abdominal neuroblastomas involve major vessels, which is stage 2A disease with localized tumor that cannot be completely excised. In unresectable abdominal tumors, biopsy specimens are taken, and then the tumor is treated with radiation or chemotherapy; a subsequent reattempt to remove it might be necessary. If complete resection can be achieved in this fashion, prognosis is the same as with initial resection. In patients with metastatic disease that can be controlled by chemotherapy, resection of the primary tumor may prolong survival. Patients with functional neuroblastomas may need to undergo debulking surgery to alleviate the signs and symptoms of hormonal excess, such as hypertension or diarrhea. If all distant tumor is eliminated by chemotherapy, residual primary tumor mass is excised, although this has not demonstrated improved survival.

Radiation therapy has clearly defined roles in children with lymph node metastases (stage 2B and 3). It also has been useful in infants with spinal cord compression. Neuroblastomas are highly chemoresponsive, and sometimes they are curable with chemotherapy. Response rates for single agents are: cyclophosphamide 59 percent, cisplatin 46 percent, epipodophyllotoxins 30 percent, vincristine 20 percent, dacarbazine 14 percent, melphalan 24 percent, and ifosfamide 20 percent.

The 5-year survival rates for stage 1 and 2 disease are 90 and 85 percent, respectively. For patients with stage 3 disease, the 5-year survival rate is 60 percent, and for patients with stage 4 disease it is 20 to 40 percent. Although surgery is curative for patients with localized, completely resectable neuroblastoma, and there are active chemotherapeutic drugs for patients with distant disease, the overall survival rate of children with neuroblastoma needs to be improved.

ADRENALECTOMY

Posterior Approach

This approach is used for resection of small benign tumors of the adrenal gland, e.g., an aldosteronoma or a cortisol-secreting adenoma. It also can be used for resection of hyperplastic glands in hypercortisolemic patients with primary pigmented micronodular hyperplasia, and for adrenalectomy in patients with Cushing's disease in whom transphenoidal surgery has failed. The patient is placed supine on the operating table and general anesthesia is induced. The patient is then turned to the prone jackknife position (Fig. [35-34](#)). The chest and the pelvis are padded by rolls so that the abdominal contents fall anteriorly away from the retroperitoneum. A curvilinear hockey-stick incision (bilateral incisions if bilateral adrenalectomy is indicated) is fashioned from the tenth rib to the posterior superior iliac spine (Fig. [35-35](#)). The superior part of the incision is 3 fingerbreadths from the midline, and the lower part is 4 fingerbreadths from the curvilinear or hockey-stick shape. The incision is made through the skin, subcutaneous tissue, and the latissimus dorsi muscle to the posterior lamella of the lumbodorsal fascia. This is incised, and the sacrospinalis muscle is retracted medially. The twelfth rib (and occasionally the eleventh rib as well) is removed subperiosteally. The anterior lumbodorsal fascia is incised, and the intercostal neurovascular bundle is clamped and divided. The retroperitoneum is entered. The kidney, which is surrounded by Gerota's fascia, is palpated at the depth of the incision. Care is taken to separate the pleura from the diaphragm without entering the thoracic cavity. The diaphragm is divided with the Bovie electrocautery knife to expose the adrenal gland, which lies within the fat at the superior edge of the incision. Gerota's fascia is incised. The kidney is identified and retracted inferiorly with the hand. After division of Gerota's fascia at the superior edge of the incision, the chrome yellow adrenal cortex is visualized.

The adrenal gland is freed with blunt dissection. On its medial, inferomedial, and superomedial margins are the arterial feeding vessels, which are small and numerous. These vessels are identified with blunt dissection, hemoclips are applied, and the vessels are divided. The adrenal vein is ligated toward the end of the resection. On the left side, the adrenal vein may be found at the inferomedial border of the adrenal gland, diving away from the operating surgeon as it empties into the renal vein (Fig. [35-36](#)). On the right side, the adrenal vein may be found on the superomedial border of the adrenal gland, where it empties directly into the inferior vena cava. The right vein is wide and short, and if torn can cause life-threatening hemorrhage. The adrenal vein is ligated with a 2-0 silk tie, and the adrenal gland is removed (Fig. [35-37](#)).

At the time of closure, the diaphragm is not reapproximated. If the thorax has been entered, pleural air is evacuated with a tube while the lungs are hyperinflated by the anesthesiologist, and the pleural opening is closed with a running suture. The anterior and posterior lamella and the lumbodorsal fascia are closed; the skin is closed. In patients with Cushing's syndrome, the closure should be meticulous because wound complications are common.

Laparoscopic Approach

The laparoscopic approach to adrenalectomy has developed as a result of technologically improved equipment and the quest for less invasive and less painful operative procedures. This method of adrenalectomy is suited for small aldosteronomas or cortisol-secreting adenomas. It also is indicated for resection of hyperplastic adrenal glands. Although laparoscopy has been used in a few patients with small pheochromocytomas, it is not recommended because of the need for more tumor manipulation with the instruments and the inability to control the adrenal vein at an early stage of the procedure. Experience in open adrenalectomy is recommended for surgeons who choose to perform laparoscopic adrenalectomy, because difficult laparoscopic dissections can require emergent conversion to open procedures.

The patient is positioned in the lateral decubitus position with the affected adrenal side superficial for laparoscopic adrenalectomy. The operating table is flexed, and the kidney rest is extended. An inflatable bean bag is used to secure the patient in the operating lateral position. Generous padding is used, especially in patients with hypercortisolism. A transperitoneal approach is used from this lateral position. Four ports are placed transversely across the upper abdomen approximately 2 to 4 fingerbreadths below the costal margin from the midclavicular line to the posterior axillary line (Fig. [35-38](#)). The abdominal cavity is entered through the anterior port site and insufflated with CO₂ gas to a pressure of 12 to 15 mmHg. A videoendoscope is inserted through one port and operating instruments that include retractors, dissecting forceps, grasping forceps, irrigation, electrocautery, laparoscopic ultrasound probe, and harmonic scalpel are inserted through other ports. On the left side the splenic flexure of the colon must be mobilized, and the spleen and the tail of the pancreas are mobilized anteriorly. The adrenal gland is identified with the use of the ultrasound probe by scanning over the superior and medial border of the kidney (Fig. [35-39](#)). On the right side, the adrenal gland is exposed by dissecting and mobilizing the right lobe of the liver anteriorly. The ultrasound probe can be used to determine the precise location of the adrenal gland. The gland is dissected and removed in a fashion similar to the open procedure, described previously, except that the ultrasonic coagulation-cutting instrument (harmonic scalpel) is used to induce hemostasis in the small arterial branches to the gland. Large hemoclips are applied to the adrenal vein (Fig. [35-40](#)). The resected gland is placed in a bag to avoid spillage while it is manipulated through the small incisions. The small skin and fascial incisions are closed. The laparoscopic approach to adrenalectomy is associated with less pain and a faster return to normal function than the open procedures. In experienced hands, the conversion rate from laparoscopic to open procedures is acceptably low.

Anterior Approach

The anterior approach for adrenalectomy is the recommended approach for potentially malignant adrenal tumors and pheochromocytomas. This approach is chosen because it provides access to the entire abdomen for accurate staging of any lymph node or liver metastases. It also allows better control of the inferior vena cava and other great vessels in the resection of malignant tumors, and it allows the surgeon to visualize and palpate along the sympathetic chain, which is a common extraadrenal site of pheochromocytoma. Some authors have stated that the use of newer localization studies has reduced the importance of this indication.

The abdominal cavity is entered through a midline incision or a bilateral subcostal (chevron) upper abdominal incision (Fig. [35-41](#)). A fixed upper abdominal retractor, e.g., the Thompson or the Omni, aids exposure. The left adrenal gland is approached by dividing the ligaments between the spleen and the diaphragm and mobilizing the spleen and the tail of the pancreas out of the retroperitoneum. Alternatively, the left gland may be approached by opening the lesser sac and mobilizing the splenic flexure. Once the lesser sac is entered, the pancreas can be elevated by dividing the inferior border and identifying the adrenal gland, which is posterior to the pancreas. The complete mobilization of the spleen and the pancreatic tail is preferred because it avoids the need to work in a deep hole with poor visualization. The left adrenal vein runs along the inferomedial border and empties into the left renal vein. With the anterior approach, the adrenal vein can be easily ligated early in the procedure, and doing so is preferable in resecting pheochromocytomas. After ligation and division of the left adrenal vein, hemoclips are applied to the small arteries that course along the medial border of the adrenal gland, and the arteries are divided (Fig. [35-42](#)).

The right adrenal gland can be exposed by dissecting along the inferior vena cava underneath the liver, or mobilizing the entire right lobe of the liver as is done during a liver resection. The second method is preferred because it allows better exposure of the right adrenal gland, which is nestled underneath and to the right of the inferior vena cava. The right adrenal vein easily can be exposed after mobilization of the right lobe of the liver at the superior medial border of the gland, where it drains into the inferior vena cava (Fig. [35-43](#)). After ligation and division of the vein, hemoclips are applied to the arteries along the medial border, and the arteries are divided.

The open anterior approach to the adrenal gland is associated with more pain and longer hospitalization than the laparoscopic approach. It is indicated for larger tumors and

pheochromocytomas, though some pheochromocytomas are resected laparoscopically.

Thoracoabdominal Approach

The thoracoabdominal approach to adrenalectomy is indicated for large, clearly malignant tumors that may involve adjacent organs, such as the liver, kidney, pancreas, spleen, or the inferior vena cava. The patient is positioned with the side of the tumor upward so the chest and abdomen can be entered. The incision is through the eleventh intercostal space, across the costal margin, and into the abdomen. Each adrenal gland is exposed in the same manner as previously described for the anterior approach. The resection proceeds in a similar fashion. The thoracoabdominal approach is associated with more postoperative pain than the other approaches, but it is indicated for some tumors because of their size and direct invasion of contiguous structures.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 36 - Thyroid and Parathyroid

Gregory P. Sadler
Orlo H. Clark
Jon A. van Heerden
David R. Farley

HISTORICAL BACKGROUND

Goiters (from the Latin *guttur*, throat) have been known since 2700 b.c., long before the thyroid gland was recognized. The thyroid gland was first documented by the Italians of the Renaissance period. Leonardo da Vinci originally depicted the thyroid in his drawings as two separate glands on either side of the larynx. The term thyroid gland (Greek *thyreoeides*, shield-shaped) is attributed to Thomas Wharton in his *Adenographia* (1656), although Bartholomeus Eustachius had used the description previously; his work, however, was not published until the eighteenth century. In 1619, Hieronymus Fabricius ab Aquapendente recognized that goiters arose from the thyroid gland. It was Albrecht von Haller in 1776 who classified the thyroid as a ductless gland. Many functions were imaginatively ascribed to the thyroid gland, including lubrication of the larynx, providing a reservoir of blood to prevent engorgement of the brain, or beautifying women's necks.

Treatment of goiter was varied; marine preparations, such as burnt seaweed, were among the most effective. In 1811 Bernard Courtois discovered iodine in the ash of burnt seaweed. Surgery of goiters was hazardous, with an exceedingly high complication and mortality rate. The first accounts of thyroid surgery were given by Roger Frugardi in 1170. Failing response to medical treatments, two setons were inserted at right angles into the goiter and tightened twice daily until the goiter separated. The open wound was then treated with caustic powder and left to heal.

Thyroid surgery continued to be hazardous (mortality over 40 percent) until the mid-nineteenth century, when advances in general anesthesia (1840s), antisepsis (1860s), and hemostasis (1870s) enabled surgeons to perform thyroid surgery with significantly reduced mortality. The most notable thyroid surgeons were Emil Theodor Kocher (1841–1917) (Fig. 36- 1) and C.A. Theodor Billroth (1829–1894), who performed thousands of operations with increasingly successful results. As patients survived longer, however, problems emerged that had not been previously encountered. After total thyroidectomy, patients became myxedematous with cretinous features; the changes were more noticeable in children. Kocher coined the term “cachexia strumipriva” and wrongly attributed it to operative tracheal trauma giving rise to chronic asphyxia. Felix Semon suggested that myxedema was secondary to the loss of thyroid function, a view originally treated with skepticism. This was later proved true by Victor Horsley’s studies on monkeys undergoing total thyroidectomy.

The first successful treatment of myxedema was achieved in 1891 by George Murray when he prepared an extract of sheep's thyroid that he injected subcutaneously into a patient. The following year, Edward Fox demonstrated that oral therapy in the form of “half a sheep's thyroid, lightly fried and taken with currant jelly once a week” was equally effective.

Few of Billroth's patients developed myxedema, but William Halsted suggested that this was because of a difference in operative technique. Kocher was extremely neat and precise, operating slowly in a bloodless field. He removed all the thyroid, and his patients developed myxedema but rarely suffered laryngeal nerve damage or postoperative tetany. Billroth, however, worked rapidly and with less concern for hemorrhage. He often removed the parathyroid glands but left more thyroid tissue and therefore encountered postoperative hypoparathyroidism but rarely myxedema. In 1909 Kocher received the Nobel Prize for medicine in recognition “for his works on the physiology, pathology, and surgery of the thyroid gland.”

EMBRYOLOGY

A clear understanding of the developmental embryology and anatomy of the thyroid gland is essential for the clinician performing a thorough physical examination of the gland and aids in evaluating diagnostic images. Knowledge of possible developmental anomalies and the thyroid gland's relationship to the parathyroid glands and other neck structures is vital in performing safe and effective thyroid operations.

The thyroid gland originates from the base of the tongue in the region of the foramen cecum. Embryologically, it is an offshoot of the primitive alimentary tract. The endoderm cells in the midline of the floor of the pharyngeal anlage thicken and form a median thyroid anlage, which migrates caudally into the neck (Fig. 36-2). The anlage descends along a tract that runs anterior to the structures that form the hyoid bone and the larynx; it is composed of epithelial cells that provide the follicular cells of the thyroid. As it descends, it is joined laterally by a pair of components originating from the ultimobranchial bodies of the fourth and fifth branchial pouches. These lateral components supply the C cells of the thyroid, which secrete calcitonin. When the C cells become neoplastic, the result is medullary carcinoma of the thyroid. An understanding of this anatomy explains why medullary carcinoma usually is located in the upper poles of the thyroid and virtually never in the isthmus or pyramidal lobe. The thyroid gland forms follicles by the end of the tenth week of gestation and concentrates iodine and produces colloid by the end of the twelfth week.

ANOMALIES

Rarely, the thyroid gland, whole or in part, descends more caudally. This results in thyroid tissue located in the superior mediastinum behind the sternum, adjacent to the aortic arch or between the aorta and the pulmonary trunk, within the upper portion of the pericardium, or in the interventricular septum. The following types of anomaly can be encountered.

Pyramidal Lobe

The migratory tract of the developing thyroid gland is known as the thyroglossal tract or duct. Normally the duct atrophies, although it may remain as a fibrous band. In about 80 percent of people, the distal end that connects to the thyroid persists as a pyramidal lobe projecting up from the isthmus, lying just to the left of the midline (Fig. 36-3). In the normal individual the pyramidal lobe is not palpable, but in disorders resulting in thyroid hypertrophy (e.g., Graves' disease, diffuse nodular goiter, or lymphocytic thyroiditis), the pyramidal lobe usually is enlarged and palpable.

Lingual Thyroid

The median thyroid anlage sometimes fails to develop, resulting in athyreosis, or it may develop but fail to descend, leading to a lingual thyroid (Fig. 36-4). Lingual thyroid is estimated to occur in 1 in 3000 cases of thyroid disease. It occurs more commonly in females, and some develop hypothyroidism. In these patients, the lingual thyroid is the only functioning thyroid tissue, although a normally situated thyroid also may be present.

Presentation usually is dependent upon the size of the lingual thyroid. An asymptomatic posterior lingual mass may be discovered because of physiologic thyroid hyperactivity. If tumor formation occurs the patient presents with symptoms of a posterior oral swelling. If the thyroid tissue continues to enlarge, symptoms such as a choking sensation, dysphagia, dyspnea, and dysphonia may predominate.

Diagnosis is established by scanning with radioiodine (123I) (Fig. 36-5) or technetium (99mTc). Treatment consists of thyroid suppression with thyroxine; operation for symptoms or an enlarging mass is rarely necessary and may result in hypothyroidism.

Malignancy is rare, occurring in less than 3 percent of patients with symptomatic lingual thyroids. Diagnosis in these cases may be established by *fine-needle aspiration cytology (FNAC)* or biopsy.

Thyroglossal Duct Cyst

Thyroglossal duct cysts are midline structures containing thyroid epithelium; they may occur anywhere along the course of the thyroglossal duct, though typically they are found between the isthmus of the thyroid gland and the hyoid bone (Fig. 36-6). The cysts usually cause few symptoms but may become infected, prompting the patient to seek medical advice.

Diagnosis may be established by asking the patient to protrude his or her tongue; when the tongue is protruded, the thyroglossal duct cyst moves upward. Treatment is by surgical excision and should include the thyroglossal duct remnant. As the duct may pass anteriorly to, posteriorly to, or through the hyoid bone, the central portion of the hyoid bone is removed to minimize the possibility of recurrence (the Sistrunk procedure).

About 1 percent of thyroglossal duct cysts contain thyroid cancer, and approximately 25 percent of patients with thyroglossal duct cysts that contain papillary cancer have papillary cancer elsewhere within the thyroid gland. Occasionally squamous cell carcinomas develop in thyroglossal duct cysts. Medullary thyroid cancers are not found in thyroglossal duct cysts.

Lateral Aberrant Thyroid

Lateral aberrant thyroid tissue is rare. It is believed that the so-called “lateral aberrant thyroid” is almost always a well-differentiated papillary carcinoma (exhibiting a follicular pattern) that has metastasized to a cervical chain lymph node, replacing the node with tumor. Diagnosis of lateral aberrant thyroid should direct the clinician to search for the primary thyroid tumor, which is almost always present in the ipsilateral lobe of the thyroid. In some patients the primary thyroid cancer is microscopic. Normal ectopic thyroid tissue may be present in the neck; it is always in the central neck (the migratory path of the normal thyroid), it is not situated in lymph nodes, and it is benign.

ANATOMY

The normal adult thyroid gland is light brown in color and firm in consistency, weighing 15 to 20 g. It is formed by two lateral lobes connected centrally by an isthmus. The lobes are approximately 4 cm long, 2 cm wide, and 20 to 40 mm thick, with the isthmus 2 to 6 mm thick. The lateral lobes run alongside the trachea, reaching the level of the middle thyroid cartilage superiorly. Laterally, the lobes are adjacent to the carotid sheath and the sternocleidomastoid muscles; anteriorly, they are adjacent to the strap muscles (sternothyroid and sternohyoid). In approximately 80 percent of individuals, a pyramidal lobe is present, usually just to the left of the midline, extending upward from the isthmus along the anterior surface of the thyroid cartilage. It is a remnant of the thyroglossal duct (see Fig. 36- 3).

The four parathyroid glands usually are closely related to the thyroid gland, found on the posterolateral surface of the lobes, within 1 cm of the inferior thyroid artery in 80 percent of individuals. The upper parathyroid glands are more dorsal or posterior and usually are situated at the level of the cricoid cartilage. The lower parathyroid glands are more variable in position but usually are anterior to the recurrent laryngeal nerves. The thyroid gland is enveloped by a loosely connecting fascia that is formed from the partition of the deep cervical fascia into anterior and posterior divisions. The thyroid is attached to the trachea and suspended from the larynx. It moves upward with elevation of the larynx on swallowing. The true capsule of the thyroid is a thin, fibrous layer, densely adherent, that sends out septa that invaginate the gland, forming pseudolobules. Thyroid nodules are palpable in about 4

percent of adults; smaller, occult nodules can be detected by ultrasound or at postmortem examination in more than 50 percent of older adults.

The thyroid gland has an abundant blood supply provided by four major arteries. The paired superior thyroid arteries arise as the first branch of the external carotid artery, approximately at the level of the carotid bifurcation, and descend several centimeters in the neck to the superior pole of each thyroid lobe. Here the arteries divide into anterior and posterior branches as they reach the gland. The paired inferior thyroid arteries arise from the thyrocervical trunk of the subclavian arteries and enter the gland from a posterolateral position. Occasionally a fifth artery, the thyroidea ima, is present, originating directly from the aortic arch or the innominate artery and ascending in front of the trachea to enter the gland in the midline inferiorly. A rich venous plexus forms under the capsule and drains to the internal jugular vein on both sides via the superior thyroid veins (which run with the superior thyroid artery) and the middle thyroid veins, which can vary in number, passing from the lateral aspect of the lobes. The inferior thyroid veins leave the inferior poles bilaterally, usually forming a plexus that drains into the brachiocephalic vein. Lymphatic drainage of the thyroid gland is primarily to the internal jugular nodes. The superior pole and medial isthmus drain to the superior groups of nodes, and the inferior groups drain the lower gland and empty into pretracheal and paratracheal nodes.

Innervation of the gland is by sympathetic fibers from the superior and middle cervical sympathetic ganglia. The fibers enter with the blood vessels and are vasomotor in action. Parasympathetic fibers are derived from the vagus nerve and reach the gland via branches of the laryngeal nerves.

Microscopically, the thyroid is divided into lobules that contain 20 to 40 follicles. There are roughly 3×10^6 follicles in the adult male thyroid gland. The follicles are spherical and average 30 μ m in diameter. Each follicle is lined by cuboidal epithelial cells and contains a central store of colloid secreted from the epithelial cells under the influence of the pituitary hormone, thyroid stimulating hormone (TSH). The second group of thyroid secretory cells are the C cells or parafollicular cells, which contain and secrete the hormone calcitonin. They are found as individual cells or clumped in small groups in the interfollicular stroma, abutting between follicular cells. They are located in the upper poles of the thyroid lobes, reflecting their origin as neuroectodermal cells derived from the ultimobranchial bodies, and are part of the amine containing precursor uptake decarboxylase (APUD) series described by Pearse.

Laryngeal Nerves

It is important to note the close relationship of the thyroid gland to the recurrent laryngeal nerves and the possible variations in the course of the recurrent nerves. The recurrent laryngeal nerves supply the intrinsic muscles of the larynx, and damage to one of them leads to ipsilateral vocal cord paralysis. Similarly, the external branch of the superior laryngeal nerve, which innervates the cricothyroid muscle, also is at risk during thyroid surgery. Damage of either nerve may result in a disability of phonation.

Identification of the nerves, rather than attempting to avoid them, should be standard practice for the surgeon. The recurrent laryngeal nerves originate from the vagus nerves. On the right side, the recurrent nerve originates where the vagus nerve crosses the first part of the subclavian artery; the nerve loops under the subclavian artery and ascends slightly obliquely to enter the larynx at the level of the cricoid cartilage and posterior to the cricothyroid muscle. The left recurrent nerve branches from the vagus as it crosses the aortic arch and loops posteriorly around the ligamentum arteriosus before it ascends medially in the tracheoesophageal groove to enter the larynx opposite the contralateral nerve. The variable course taken by the recurrent nerves is demonstrated in Fig. 36-7. The right recurrent nerve is in the tracheoesophageal groove in 64 percent of people, compared to 77 percent on the left. The nerve is lateral to the trachea on the right in 28 percent of people and in 17 percent on the left. In a minority of people the nerve is anterolateral to the trachea (right 8 percent, left 6 percent), exposing it to accidental division during subtotal lobectomy. A misconception is that the recurrent laryngeal nerves run behind the inferior thyroid artery, but this is true in only 53 percent of people on the right and 69 percent on the left. In others the nerve's course is anterior to the artery (right 37 percent, left 24 percent), or between branches of the artery (right 7 percent, left 6 percent). Failure to identify the course of the nerves can lead to accidental damage.

The recurrent laryngeal nerves are not always recurrent; in about 1 percent of people one of the nerves is nonrecurrent. This occurs almost exclusively on the right in association with a vascular anomaly of the right subclavian artery; rarely, it occurs on the left with dextrocardia or situs inversus. In these situations, the nerve arises from the vagus to run directly to the larynx, often in close proximity to the superior thyroid vessels, and may be at risk when these vessels are transected (Fig. 36-8).

The superior laryngeal nerve arises from the vagus near the base of the skull and descends medial to the carotid vessels. At the level of the hyoid bone it divides into two branches, one sensory (internal branch), and the other motor (external branch). The external branch runs on the lateral surface of the inferior constrictor muscle and descends to innervate the cricothyroid muscle (Fig. 36-9). This muscle alters vocal cord tension and affects the pitch of the voice. In most instances the nerve runs in close proximity to the superior pole vessels, and in 21 percent of people it is closely related to the vessels and is at significant risk if it is not identified at operation. To avoid injury, the superior pole vessels should be individually ligated and divided low on the thyroid gland and dissected laterally to the cricothyroid muscle.

PHYSIOLOGY

Through release of its principal hormones, thyroxine (T₄) and triiodothyronine (T₃), the thyroid gland influences the metabolic rate of all tissues. Increased secretion increases the metabolic rate; conversely, the rate decreases when secretion is decreased. Release of T₄ and T₃ is stimulated by the anterior pituitary hormone thyrotropin or *thyroid-stimulating hormone (TSH)*. Secretion of TSH is directly suppressed by T₄ and T₃ (a negative feedback loop). TSH release also is stimulated by the hypothalamic hormone *thyrotropin-releasing hormone (TRH)*. Thyroid hormone production is influenced by numerous physiologic, pathologic, and pharmacologic factors.

Iodine Metabolism

The formation of thyroid hormones is dependent on the availability of exogenous iodine. The average daily iodine requirement is 0.1 mg. Iodine is found principally in fish, milk, and eggs. In the United States, iodine is routinely added to bread and salt in order to reduce the frequency of iodine deficiency. Iodine is rapidly converted to iodide in the stomach and jejunum and is absorbed into the bloodstream within 1 h; and from there it is distributed uniformly throughout the extracellular space (Fig. 36-10). Iodide is actively transported into the thyroid follicular cells by an ATP-dependent process. The thyroid-serum iodine ratio under normal conditions is about 50:1, and most of the body's store of iodine is found in the thyroid gland (90 percent). Thyroid-serum ratios can be as high as 500:1 in certain instances, such as iodine deficiency or Graves' disease.

One-third of the loss of iodine from the plasma is accounted for by thyroid concentration, and the other two-thirds through renal excretion. In studies involving radiolabeled iodine, all the iodine is concentrated within the thyroid or excreted in the urine within 48 h, and the plasma and tissues are mostly cleared of iodide. Evidence of labeled iodine in serum is accounted for by secretion from the thyroid gland in the form of thyroid hormone.

Synthesis of Thyroid Hormone

Steps in the synthesis of thyroid hormone are: (1) active trapping and concentration of iodide in the follicular cell; (2) rapid oxidation of iodide to iodine; (3) linkage of iodine with tyrosine residues in thyroglobulin; (4) coupling of these iodotyrosines (monoiodo- and diiodotyrosine) to form the active thyroid hormones T₄ and T₃. Active accumulation of iodide in the thyroid gland is stimulated by TSH, acting via a specific membrane receptor located in the thyrocyte plasma membrane. This mechanism is probably through changes in cyclic adenosine monophosphate (cAMP). Once inside the thyroid cell, the iodide diffuses through the cytoplasm to the apical membrane. It remains in its free state for a short time before being oxidized by peroxidase and hydrogen peroxidase. Iodine rapidly links to tyrosine residues present in abundance in thyroglobulin, a thyroid-specific protein, resulting in the formation of two separate molecules, *monoiodotyrosine (MIT)* and *diiodotyrosine (DIT)*. Two molecules of DIT combine to form tetraiodothyronine, or thyroxine (T₄); a molecule of MIT and DIT combine to form 3,3',5-triiodothyronine (T₃) or 3,3',5'-triiodothyronine, reverse T₃ (rT₃). The coupling steps are catalyzed by peroxidase in the presence of H₂O₂ and also are rate dependent on TSH.

When iodide transport is defective or when oxidation to iodine is impaired because of disease or pharmacologic agents, goiter or hypothyroidism may result. The antithyroid drugs (propylthiouracil, methimazole, and carbimazole) inhibit the oxidation of iodide to iodine by competitive inhibition of peroxidase and also may interfere with the coupling reaction. In high doses iodide also inhibits iodine trapping. It also has an antithyroid action by inhibiting the proteolysis involved in the release of thyroid hormone. Potassium iodide tablets often are administered to people exposed to radiation leaks involving radioactive forms of iodine, such as nuclear accidents, because it blocks trapping by the thyroid gland.

Storage, Secretion, and Metabolism of Thyroid Hormone

T₄ and T₃ are bound to thyroglobulin and are stored in the colloid of the thyroid follicles. Release of the active hormones is by a process of endocytosis. The colloid is taken up by the follicular cell as discrete packets (endosomes), which then fuse with lysosomes containing hydroxylases. Hydrolysis results in production of all component parts, T₄, T₃, rT₃, MIT, and DIT. Through a process of deiodination most of the iodide is released from MIT and DIT and reused in the follicle. The iodothyronines are more resistant to this process and are secreted; these steps also are TSH dependent.

The active thyroid hormones circulate in the plasma attached to plasma proteins, principally the carrier proteins, *thyroid hormone-binding globulin (TBG)*, thyroid hormone-binding prealbumin (TBPA), and albumin. About 99.98 percent of thyroid hormone circulates in the plasma bound to protein, and the remaining 0.02 percent is unbound and is the free, *active physiological fraction*. In some conditions TBG may be increased, usually as a result of estrogen effects of pregnancy or the contraceptive pill. This results in a higher circulating amount of T₄ because of increased serum binding capacity. In this situation, *active free T₄* levels remain unaltered.

T₃ is the more potent of the two thyroid hormones (rT₃ is biologically inert), although its circulating plasma level is much lower than that of T₄; the ratio is 10:1 to 20:1. T₃ is less tightly bound to protein in the plasma than T₄, and so it enters tissues more readily. T₃ is three to four times more active than T₄ per unit weight, with a half-life of about 1 day, compared to about 7 days for T₄. Though the thyroid gland produces some T₃ and rT₃, it is known that 75 percent of T₃ is produced by the extrathyroidal conversion of T₄ to T₃ in the peripheral tissues. Almost 85 percent of T₄ is converted peripherally to metabolically inert rT₃ or T₃. Some studies suggest that T₄ is a prohormone and that T₃ is the only hormone acting at the cellular level.

Molecular Basis of Thyroid Hormone Action

Thyroid hormones are transported across the plasma membrane of tissues by an ATP-dependent transport system. Uptake by the tissue is rate-limited by the amount of free hormone available at the tissue level. T₃ is the active hormone, and its activity is mediated through T₃ receptors located in the cell nucleus. The receptors bind to regulatory genes and modify the expression of these genes.

T₃ receptors belong to a group of hormone-responsive nuclear transcription factors. There are two types of T₃ receptor genes, a and b, located on chromosomes 17 and 3. Expression of T₃ receptors is tissue specific. T₃ receptors a1, a2 and b1 mRNA are expressed in almost all tissues, but some T₃ receptors are expressed only in certain tissues, e.g., b2 is found only in the brain. The brain contains mostly areceptors, the liver b receptors, and cardiac muscle expresses both.

Deiodination and Excretion

Deiodination of thyroid hormones is effected by three different types of deiodination enzymes, which are tissue specific. The released iodine is returned to the blood, where it reenters the metabolic pool. The residual T₃ and T₄ are conjugated with glucuronic acid, which renders the hormones water soluble and facilitates excretion in urine and bile, or

sulphate. Some of the excreted iodothyronines are reabsorbed from the small intestine, constituting the enterohepatic circulation. About one-third of total body clearance is effected through the bile, but up to 50 percent of the thyroxine may be reabsorbed. Significant amounts of thyroid hormone and iodine may appear in the milk of lactating mothers.

Regulation of Thyroid Activity

The principal homeostatic control of thyroid hormone secretion is the hypothalamic-pituitary-thyroid axis. The basophil cells of the anterior pituitary produce TSH, which directly regulates thyroid function. TSH acts on the thyroid cell to promote thyroid hormone production at all levels, enhancing iodine uptake, increasing synthesis, and raising secretion of T₄. TSH also has a secondary action on thyroid gland growth, increasing cellularity and vascularization of the gland. Secretion of TSH is regulated at two levels. Thyrotropin-releasing hormone (TRH), is produced by the hypothalamus and reaches the gland via the hypophyseal portal system to stimulate TSH release (Fig. 36-11). TRH binds to high-affinity TRH receptors on the anterior pituitary cells. Originally it was thought that TRH exerted its action of TSH release via adenylate cyclase and cAMP, but now it is believed that postreceptor activation is via the phospholipase-C–based hydrolysis of inositol phospholipids, leading to Ca²⁺ and diacylglycerol activation of protein kinase C. Release of TRH from the hypothalamus is suppressed by T₃, acting in a feedback loop. TRH has been shown to be equipotent in stimulating release of prolactin from the pituitary and TSH.

More important to thyroid hormone regulation is the direct feedback exerted on the pituitary by the level of thyroid hormone in the blood. Raised levels of thyroid hormone suppress TSH and TRH secretion, and lowered levels promote secretion. Iodine deficiency increases the goitrogenic effects of TSH on the thyroid.

ASSESSMENT OF PATIENTS WITH THYROID DISEASE

Thyroid disease may be divided into two types: problems relating to function (hyperthyroidism/hypothyroidism) and thyroid masses. The two types are not mutually exclusive and patients frequently present with both problems.

History

Obtaining an accurate history is essential in assessing thyroid disease. Symptoms such as dysphagia, dyspnea, and choking are frequently encountered in patients with goiter and may be exaggerated by patients raising their arms above their heads (Pemberton's sign). Pain is uncommon. Localized pain may suggest malignancy, especially medullary thyroid cancer, whereas pain radiating to the ear often is observed in patients with thyroiditis or hemorrhage within the thyroid gland. A change in the character of the voice should also be of concern because it may suggest involvement of the recurrent laryngeal nerves in a malignant process, with vocal cord paralysis. A past history of exposure to radiation, family history of benign or malignant thyroid disease, living in an iodine- deficient area, or ingestion of goitrogenic drugs also are significant.

Physical Examination

Thyroid masses rise on swallowing; most thyroid swellings are accurately discernible by observing the patient swallow. Failure to observe before palpating the thyroid gland may lead to missing a large retrosternal goiter arising from beneath the sternum and clavicles.

Palpation usually is performed from behind while the patient is sitting in a chair with the neck slightly extended and should include palpation of the gland while the patient swallows. A landmark is the cricoid cartilage; the isthmus almost always crosses a fingerbreadth below the cricoid. The normal thyroid gland usually is not palpable unless the patient has a particularly thin neck. The thyroid gland may be diffuse and bilaterally enlarged (goiter), as encountered in conditions such as Graves' disease (hyperthyroidism), Hashimoto's thyroiditis, or multinodular goiter. A unilateral mass may be palpated, as in a colloid nodule, follicular adenoma, or carcinoma.

The cervical chain of lymph nodes should be assessed as well as the nodes in the posterior triangle. The jugular nodes immediately adjacent to a thyroid nodule often are involved in patients with a papillary thyroid cancer. A Delphian node should be palpated for just above the thyroid isthmus and cricoid cartilage.

Fine-Needle Aspiration Cytology (FNAC)

Fine-needle aspiration cytology is a simple and low-risk technique that is an integral part of thyroid assessment in the outpatient setting for patients with thyroid nodules. A 23-gauge needle is inserted into the thyroid swelling, and several passes are made while aspirating the syringe. Cells are placed on prelabeled dry glass slides; some are then immediately placed in 70% alcohol while others are air dried. These slides are stained by Papanicolaou or Wright's stains and observed under the microscope. Skilled cytopathologists can accurately diagnose the majority of thyroid diseases using this technique, with a high degree of specificity. This test is less accurate in patients with thyroid nodules and a history of familial nonmedullary thyroid cancer and in patients with a previous history of exposure to low-dose therapeutic radiation. Benign and malignant thyroid tumors are common in such patients, and the tumors usually are multifocal.

Tests of Thyroid Function

Thyrotropin (TSH, reference range 0.15–4.2 mU/mL)

Thyrotropin secretion from the anterior pituitary is controlled via a negative feedback loop by serum T₃ and T₄ levels. In cases in which high T₃ and T₄ levels are encountered (Graves' disease, toxic nodular goiter), TSH levels will be accordingly low and may be undetectable. When T₄ levels are low (primary thyroid destruction, e.g., end-stage Hashimoto's thyroiditis), TSH levels will be correspondingly high. Many clinicians believe that the circulating level of TSH is the single most sensitive test of thyroid function.

Older radioimmunoassays have been replaced with more sophisticated immunometric assays, using monoclonal antibodies that target two separate sites (increasing specificity) on the TSH molecule. One monoclonal antibody is labeled with a nonradioactive marker, allowing readings with an accuracy down to 0.005 mU/mL. A reference range of normal TSH levels has been established in euthyroid patients, against which levels from the test subject can be compared.

Total Thyroxine (TT₄, reference range 55–150 nmol/L) and Free Thyroxine (FT₄, 12–28 pmol/L)

Total thyroxine concentration reflects the fraction of T₄ bound to TBG and other carrier proteins in the serum and also the amount of free T₄ in circulation. T₄ production from the thyroid is dependent on TSH from the pituitary and an adequate intake of iodine in the diet. When T₄ production from the thyroid is increased, the bound and free T₄ levels rise (FT₄ remains in equilibrium with bound T₄), resulting in an increase of TT₄. When T₄ production from the thyroid decreases, bound and FT₄ levels drop, which leads to decreased TT₄ levels. Conditions leading to a changes in the level of TBG (e.g., estrogen intake) can alter the level of TT₄ (as binding sites increase), but are not reflected by changes in circulating FT₄, and so the individual remains euthyroid.

Free T₄ estimates are not performed as a routine screening tool in thyroid disease. Use of this test is confined to cases of early hyperthyroidism in which TT₄ levels may be normal but FT₄ levels are raised. In patients with end-organ resistance to T₄ (Refetoff syndrome) T₄ levels are increased, but TSH levels usually are normal.

Total Triiodothyronine (TT₃, reference range 1.5–3.5 nmol/L) and Free Triiodothyronine (FT₃, 3–9 pmol/L)

Levels of total T₃ or free T₃ are not used as a routine investigation of thyroid function. FT₃ is most useful in confirming the diagnosis of early hyperthyroidism, in which levels of FT₄ and FT₃ rise before TT₄ and TT₃. Most T₃ production comes from the peripheral conversion of T₄, and this process may be inhibited by conditions such as starvation illness (low T₃ syndrome) or by the effect of certain drugs (e.g., propranolol). There is a rare condition of T₃ thyrotoxicosis in which levels of TT₄ in the hyperthyroid patient are normal and radioiodine uptake is normal but the TT₃ level is raised. This condition is more common in patients from endemic goiter areas and in patients with small solitary thyroid nodules.

HYPERTHYROIDISM/THYROTOXICOSIS

Thyrotoxicosis is the clinical syndrome that results when excessive levels of active thyroid hormone are secreted into the circulation. There are many causes of thyrotoxicosis, but two predominate: Graves' disease (diffuse toxic goiter) and toxic solitary or multinodular goiter (Plummer's disease). The rarer conditions causing thyrotoxicosis are listed in Table 36-1. Conditions resulting in increased thyroid hormone production, such as Graves' and Plummer's disease, or secondary hyperthyroidism because of a TSH-secreting pituitary tumor should be distinguished from conditions in which there is a leak of thyroid hormone, i.e., patients with subacute painless or painful thyroiditis. Hyperthyroidism also can result from taking thyroid hormone (factitious hyperthyroidism), from struma ovarii, from increased secretion of human chorionic gonadotropin with a molar pregnancy, and from rare metastatic thyroid cancers that secrete thyroid hormone.

Graves' Disease (Diffuse Toxic Goiter)

Graves' disease is the most common form of thyrotoxicosis. Although originally described by the Welsh physician Caleb Parry in a posthumous article in 1825, the disease is known as Graves' disease after Robert Graves, a physician from Ireland, who described three patients in 1835. Graves' disease is about six times more common in women, and although it may develop at any age, it is most prevalent in young adults (20 to 40 years of age). Associated extrathyroidal manifestations of this autoimmune disease include exophthalmos, pretibial myxedema, dermopathy, acropachy, and vitiligo.

Pathogenesis and Pathology

Graves' disease is an autoimmune disorder in which pathogenic thyroid- stimulating antibodies or immunoglobulins are directed at the TSH receptor on thyroid follicular cells. Binding of the antibodies stimulates the receptors and leads to excess thyroid hormone secretion, which characterizes the condition. Originally the responsible antibody was thought to be long- acting thyroid stimulating antibody (LATS), described by Adams and Purves in 1956. It is apparent now that a whole family of antibodies contribute to the development of the disease. Thyroid-stimulating immunoglobulins (TSI) or antibodies (TSAb) attach to and stimulate the TSH receptor, and TSH- binding inhibiting immunoglobulins (TSII) or antibodies (TBIA) block the TSH receptor. Current practice is to group all these antibodies together under the term thyroid receptor antibodies (TRAb).

What initiates Graves' disease and antibody production is unclear. One theory suggests a defect in the suppressor T lymphocytes allowing helper T cells to stimulate the production of TSI from helper B cell clones. Another theory is that an immune response is launched to altered antigens on the follicular cell surface, an observation supported by the fact that Graves' disease and ophthalmopathy occur more frequently in patients who have been irradiated to the head and neck. Genetic factors also are clearly involved; identical twins have a 50 percent chance of developing the condition if the twin has it, compared to a 30 percent chance in fraternal, nonidentical twins. This is probably through increased frequency of leukocyte antigen expression (HLA-b8 and DR3 in Caucasians and HLA- Bw35 in Japanese).

Macroscopically, the thyroid gland in patients with Graves' disease is diffuse and smoothly enlarged, and the gland's vascularity also is increased. Microscopically, the gland is hyperplastic, and the epithelium is columnar, with minimal colloid present. The nuclei exhibit mitosis, and papillary projections of hyperplastic epithelium are common. There may be aggregates of lymphoid tissue, and vascularity is markedly increased.

Clinical Features Common to All Forms of Thyrotoxicosis

The clinical symptoms and signs of thyrotoxicosis are the same in patients with Graves' disease and toxic nodular goiter, except that patients with Graves' disease usually have

more severe hyperthyroidism and have extrathyroidal manifestations of disease. Attention should be paid to a family history of autoimmune thyroid disease, including Graves' disease, Hashimoto's thyroiditis, and other autoimmune disorders.

Manifestation of the increased caloric turnover may be evident. Patients develop heat intolerance, increased thirst, sweating, and weight loss despite adequate caloric intake. Women may develop amenorrhea and decreased fertility and have an increased incidence of miscarriage. Cardiovascular manifestations are tachycardia or atrial fibrillation. In cases in which high-output cardiac failure ensues, signs and symptoms of congestive cardiac failure such as dyspnea and peripheral edema or even anasarca may become evident. Adrenergic stimuli may be particularly distressing, and fatigue, agitation and excitability, disturbed sleep pattern, emotional lability, hyperkinesis, and tremor may be present. In marked cases, psychosis can develop. Diarrhea or increased bowel frequency are the most common gastrointestinal manifestations and run an intermittent course during the disease.

On physical examination, weight loss and facial flushing may be evident. The skin may be warm and moist, and patients often have inappropriate sweating in a cool environment. African-American patients often note darkening of their skin. Examination of the pulse usually reveals tachycardia or atrial fibrillation (the latter is especially apparent in the elderly). Cutaneous vasodilation leads to a widening of the pulse pressure and a rapid falloff in the transmitted pulse wave (collapsing pulse). A fine tremor, muscle wasting, and proximal muscle group weakness with hyperactive tendon reflexes often are present.

Clinical Features Specific to Graves' Disease

Graves' disease is characterized by the classic triad of goiter, thyrotoxicosis, and exophthalmos. These features may occur singularly or in any combination. Additionally, patients present with a goiter that is characteristically diffuse, enlarged, and smooth. Evidence that the whole gland is enlarged is demonstrated by enlargement of the pyramidal lobe, which can be palpated as it crosses the cricoid cartilage (Fig. 36-12). Patients with Graves' disease also may have onycholysis or thyroid acropathy, hair loss, pretibial myxedema (3 to 5 percent) (Fig. 36-13), and gynecomastia (3 to 5 percent). An audible bruit resulting from markedly increased vascularity of the gland can be heard over the gland in up to 50 percent of patients. Splenomegaly also may be present.

Exophthalmos may be present in association with thyrotoxicosis (Graves' ophthalmopathy) or as an isolated condition with no evidence of thyrotoxicosis (euthyroidal or ophthalmic Graves' disease). The condition is characterized by: (1) spasm of the upper eyelid, with retraction revealing the sclera above the corneoscleral limbus (Dalrymple's sign) and lid lag (von Graefe's sign); (2) external ophthalmoplegia; (3) exophthalmos with proptosis; (4) supraorbital and infraorbital swelling; and (5) congestion and edema of the conjunctiva (chemosis) (Fig. 36-14). The exophthalmos is a result of increased retro-orbital tissue and can be assessed objectively with an exophthalmometer (Hertel), which measures the distance from the lateral bony orbital margin to the anterior surface of the cornea. Protrusion may lead to ophthalmoplegia, an inability to move the eyeball (upper rotation being most commonly restricted), leading to diplopia. If proptosis is progressive, optic nerve damage and blindness may occur, usually preceded by decreasing visual acuity and increasingly impaired color vision. This condition is commonly referred to as malignant exophthalmos. An urgent ophthalmic opinion should be sought. Marked protrusion can result in chemosis, in which the sclera and conjunctiva become inflamed, with itching, lacrimation, photophobia, and, eventually, ulceration.

The pathogenesis of ophthalmopathy is controversial; the cross-reaction of the thyroid antigen and ocular muscle antibodies is a possible explanation. Continued hyperthyroidism and hypothyroidism aggravate exophthalmos and should be avoided. Histologically, a diffuse lymphocytic infiltration of the retro-orbital tissues occurs, followed by fibroblast activation with glycosaminoglycan (a mucopolysaccharide) production leading to edema and fibrosis.

Diagnostic Findings in Graves' Disease

Thyrotoxicosis is characterized by an autonomous thyroid function and decreased or undetectable level of TSH in association with elevated concentrations of circulating T₃ and/or T₄. Raised levels of circulating thyroid autoantibodies are usually detected in the serum. A radioactive thyroid scan with ¹²³I is characterized by diffuse uptake throughout the gland. An uptake of 45 to 90 percent is usually observed (Fig. 36-15).

Treatment of Graves' Disease

Three treatment modalities are available for patients with Graves' disease: medical management in the form of antithyroid drugs, thyroid ablation with radioactive ¹³¹I, and subtotal or total thyroidectomy. The treatment chosen depends on the age of the patient, the severity of the disease, the size of the gland, any coexistent pathology, including associated ophthalmopathy, and other factors such as patient's preferences and pregnancy.

Antithyroid Drugs

The hyperdynamic peripheral adrenergic effects of thyrotoxicosis can be alleviated by administering beta-blocking agents. These drugs have the added effect of decreasing the peripheral conversion of T₄ to T₃. Propranolol is the most commonly prescribed medication. It reduces the heart rate, controls tremor, and to some extent relieves the agitation that these patients have. Beta blockers have no apparent effect on the overall remission rate of thyrotoxicosis.

The main antithyroid drugs are propylthiouracil (PTU) and methimazole (Tapazole) in the United States and carbimazole (in the United Kingdom). These drugs act by inhibiting the organic binding of thyroidal iodine and also inhibit the coupling of iodotyrosines. Propylthiouracil also influences the extrathyroidal conversion of T₄ to T₃. These medications have no effect on the underlying cause of the disease, although there is evidence that propylthiouracil decreases thyroid autoantibody levels.

These drugs also can cross the placenta, inhibiting fetal thyroid function, and they are excreted in breast milk. Side effects of treatment include skin rashes (1 percent), fever, peripheral neuritis, polyarteritis, granulocytopenia (which is reversible on discontinuing treatment), and, rarely, agranulocytosis (1:250). In rare instances, aplastic anemia, which has a poor prognosis, has been documented. Patients should be monitored for these possible complications and warned to stop medication and seek medical advice should they develop a sore throat or fever.

Standard medical treatment is to start the patient on 100 to 300 mg propylthiouracil three times daily, or 10 to 30 mg methimazole, initially three times daily and then once daily, or 40 mg carbimazole daily. Beta blockers are often used initially, before the diagnosis is made, to treat tachycardia and may be added for symptomatic relief. Patients are observed regularly on an outpatient basis, and the dose of antithyroid medication is titrated as needed in accordance with TSH and T₄ levels. Most patients have improved symptoms in 2 weeks and become euthyroid in about 6 weeks. The regimen described here is in wide use, though some physicians add thyroxine 0.05 to 0.10 mg to prevent hypothyroidism (the blocking/replacement regime). The length of treatment with antithyroid drugs is controversial. For patients with small, diffusely enlarged glands or larger glands that decrease in size in response to treatment with antithyroid medication, the relapse rate after treatment for 12 to 18 months is about 50 percent. Patients with larger diffuse glands or toxic nodular goiter develop recurrent hyperthyroidism when the antithyroid medication is discontinued, and hence definitive treatment with thyroidectomy or radioiodine therapy is indicated.

Radioactive Iodine Therapy (¹³¹I)

Most patients in the United States undergo treatment with radioiodine. The major advantages of this form of treatment are the avoidance of a surgical procedure and the concomitant risks of recurrent laryngeal nerve damage and hypoparathyroidism, reduced overall treatment costs, and ease of treatment. The major disadvantage is the high incidence of hypothyroidism requiring lifelong thyroxine replacement therapy, the slower correction of the hyperthyroidism, and a higher relapse rate after initial treatment, necessitating further therapy. Radioiodine therapy also has more of an adverse effect on ophthalmopathy than does thyroidectomy.

Patients most suitable for ¹³¹I therapy are those with small or moderate- sized goiters, those who have relapsed after medical or surgical therapy, and those in whom antithyroid drugs or surgery are contraindicated. Younger patients (under 35 years of age) usually are treated with thyroidectomy, and older patients are treated with ¹³¹I. Radioiodine therapy is contraindicated in women who are pregnant or breast-feeding. Relative contraindications are ophthalmopathy (in which progression of eye signs has been documented), patients with isolated thyroid nodules or toxic nodular goiters, and young age (i.e., especially children and adolescents). Although there is no evidence of long-term problems with infertility or increased incidence of cancer in children who have been treated with ¹³¹I, most specialists are reluctant to treat children in this manner and suggest thyroidectomy (usually near-total) for this age group. Children treated with radioiodine for Graves' disease have an increased risk of developing hyperparathyroidism.

Patients should be euthyroid before ¹³¹I therapy and should stop all antithyroid drugs for 2 to 3 weeks before treatment in order to allow for adequate uptake into the thyroid. Treatment is provided in the form of a drink of ¹³¹I sodium iodide, the dosage of which usually is calculated with a formula based on gland volume and ¹³¹I uptake; the typical initial dose is about 10 mCi of ¹³¹I (approximately 8500 cGy). Cure rate after initial therapy is dosage dependent; with 5 mCi, cure rate is 70 percent; with 10 mCi, 87 percent; and with 15 mCi, 96 percent. The higher the initial dose, the earlier the onset and the higher the incidence of hypothyroidism.

After standard treatment with radioiodine most patients become euthyroid within 2 months. Approximately 15 percent of patients are hypothyroid at 1 year, with a 3 percent increment each year thereafter. Six months after radioiodine treatment, 50 percent of patients are euthyroid, and the remainder are hyperthyroid or already hypothyroid. Patients need long- term follow-up with TSH levels monitored on a regular basis. Close monitoring is essential, because hypothyroidism and recurrent hyperthyroidism aggravate Graves' ophthalmopathy.

The complications of ¹³¹I treatment include: (1) exacerbation of thyrotoxicosis with arrhythmias; this usually becomes apparent within 10 days and may be a particular problem in the elderly, precipitating cardiac failure or death; (2) overt thyroid storm (rare but potentially life threatening); (3) hypothyroidism; (4) risk of fetal damage in patients who are pregnant (women are advised not to become pregnant for 6 months to 1 year after treatment); (5) worsening of eye signs, noted to be more common after ¹³¹I treatment than after surgery (33 percent compared to 16 percent); and (6) hyperparathyroidism.

Surgical Treatment

Surgery is advised when radioiodine treatment is contraindicated, such as for young patients, patients with Graves' ophthalmopathy, pregnant patients, patients with suspicious thyroid nodules in Graves' glands, and patients with large toxic nodular goiters with relatively low levels of radioiodine uptake. Thyroidectomy is the treatment of choice in patients with very large goiters and severe thyrotoxicosis at initial presentation. There is a higher failure rate with ¹³¹I treatment in these groups, necessitating additional therapy. In the United States radioiodine is the usual treatment for patients over 35 years of age with Graves' disease; in the United Kingdom and many other countries thyroidectomy is more frequently used because it is associated with less hypothyroidism and more rapid correction of hyperthyroidism. The objective of thyroidectomy for Graves' disease should be the complete and permanent control of the disease with minimal risk of morbidity in terms of nerve and parathyroid damage.

Patients should be euthyroid before operation with antithyroid drugs that should be continued up to the day of surgery. Many physicians prefer to treat patients with Lugol's iodine solution (3 drops twice daily) in the 10 days before operation, and some use propranolol. Preoperative treatment with iodine reduces the vascularity of the gland. All these measures decrease the risk of thyroid storm, which can be precipitated by surgery in unprepared patients.

Whether subtotal, near-total, or total thyroidectomy should be performed is controversial. The most commonly undertaken procedure, and perhaps the safest in terms of morbidity,

is bilateral subtotal thyroidectomy, in which about 1 to 2 g of thyroid tissue is left on both sides, or a total lobectomy on one side and a subtotal thyroidectomy on the other side (Hartley-Dunhill procedure), leaving about 4 to 5 g of thyroid tissue.

Total thyroidectomy can be performed with minimal risk of morbidity and is the operation of choice in patients with coexisting eye disease. Catz and Perzik reported no progression in 66 of 70 patients with total thyroidectomy. Similarly, Winsa and colleagues reported that ophthalmopathy stabilized or improved in 96 percent of patients 6 months or more postoperatively, which may be the result of removal of the antigenic stimulus. In their series of patients undergoing total thyroidectomy for Graves' disease, 21 of 25 patients not previously treated with ¹³¹I had normalization of TSH-receptor antibodies (TRAb) at 2.5 years.

Advantages of thyroidectomy over radioiodine treatment are: immediate cure of disease and decreased long-term incidence of hypothyroidism. Initial series probably overstated the incidence of hypothyroidism because they failed to account for later recovery of thyroid function. Other advantages include a decreased number of outpatient visits and the potential removal of a coexisting thyroid carcinoma. Disadvantages are: possible recurrent laryngeal nerve injury (approximately 1 percent), hypoparathyroidism (usually transient in approximately 13 percent and permanent in 1 percent), hematoma, and hypertrophic scar formation.

Recurrent thyrotoxicosis usually should be managed by radioiodine treatment, because reoperation carries a higher morbidity risk; when tissue has been left on one side, the risk of complications is less. Long-term follow-up should be maintained for all patients, with clinical review and yearly TSH measurement to detect the possible late onset of hypothyroidism or recurrent hyperthyroidism.

Treatment of Exophthalmos

The severity of Graves' ophthalmopathy is independent of thyrotoxicosis; data suggest, however, that recurrent hyperthyroidism and hypothyroidism aggravate the eye problems. Some reports suggest that total thyroidectomy alleviates the eye disease. It is unproved whether total thyroidectomy is preferable to near-total or subtotal thyroidectomy. Total thyroidectomy should be undertaken only in patients with severe exophthalmos when they are well prepared.

Severe or malignant exophthalmos is rare. Treatment is essentially symptomatic; steroid eye drops or systemic steroids (60mg prednisolone daily) should be used initially to alleviate chemosis. When symptoms are more severe upon awakening, patients should tape their eyes closed at night, and the head of the bed should be elevated. Patients whose eyes are worse during the day should wear glasses to protect the eyes from sun and wind and should use artificial tears to protect against drying.

Lateral tarsorrhaphy to oppose eyelids helps to alleviate drying and subsequent chemosis and corneal ulceration. In extreme situations, retro- orbital radiation or orbital decompression may be necessary to save vision.

Toxic Nodular Goiter

Toxic nodular goiter, also known as Plummer's disease, is a consequence of one or more thyroid nodules trapping and organifying more iodine and secreting more thyroid hormone independently of TSH control. Toxic nodular goiter occurs most often in areas of endemic goiter. It has been documented that most "hot" or "autonomous" thyroid nodules have TSH- receptor (common) or *gsp* (less common) mutations.

Hyperthyroidism in patients with toxic nodular goiter is milder than in patients with Graves' disease, and the condition is not accompanied by the extrathyroidal manifestations of ophthalmopathy, pretibial myxedema, vitiligo, or thyroid acropathy. Ingestion or administration of iodides, e.g., iodine supplements or intravenously administered contrast agents, may precipitate iodine-induced hyperthyroidism (Jod-Basedow phenomenon).

Patients with toxic multinodular goiter (MNG) are older at presentation than those with Graves' disease. The thyroid-gland goiters characteristically have one or more nodules on palpation. Symptoms such as dysphagia and dyspnea may be present. Some goiters are retrosternal. Symptoms are often mild, and atrial fibrillation in the elderly is frequently the only clinical finding apart from the goiter. The diagnosis is suggested by the history and physical examination and confirmed by documenting a suppressed serum TSH level and raised thyroid hormone level. Antithyroid antibodies usually are not present.

Therapy with antithyroid medication or beta blockers alleviates symptoms but usually is less effective than in patients with Graves' disease. Radioiodine therapy is not as effective as in Graves' disease because of lower uptake, and hence these patients require larger doses of radiation. ¹³¹I uptake is localized to one or more autonomous toxic nodules, and the remaining thyroid tissue is suppressed. ¹³¹I ablation may be used in patients who are unsuitable for surgery, but because of the high failure rate with this treatment, thyroidectomy is considered the treatment of choice. For solitary nodules, nodulectomy or thyroid lobectomy are the treatments of choice, because cancer is rare. For toxic multinodular goiter, lobectomy on one side and subtotal lobectomy on the other side is recommended for most patients, negating the need for bilateral reoperation in cases of recurrent disease.

Thyroid Storm

Thyroid storm is life-threatening but is rarely encountered during thyroidal—or other—surgery. Most patients with thyroid storm have had known or unknown untreated hyperthyroidism, and thyroid storm is precipitated by an infection (typically pharyngitis or pneumonitis), labor, administration of iodine (such as amiodarone), or after ¹³¹I treatment.

Signs and symptoms resemble those of severe thyrotoxicosis, with profound tachycardia, fever, and confusion. Disorientation associated with dehydration from vomiting, diarrhea, and fever may occur and, in extreme cases, adrenergic hyperactivity can lead to overt mania; coma may result as a late event.

The best management is prophylaxis. Patients with hyperthyroidism should be euthyroid before operation. The history and examination of patients admitted for procedures requiring a general anesthetic should identify undiagnosed hyperthyroidism. In cases of thyroid storm, patients can be treated in the acute phase with a combination of fluid replacement, antithyroid drugs, beta blockers, sodium iodate solution or Lugol's iodine solution, hydrocortisone, and a cooling blanket. Sedation may be necessary in cases of agitation with hyperactivity. Aspirin should be avoided because it increases free thyroid hormone levels. In extreme cases peritoneal dialysis or hemofiltration may be effective in lowering serum T₄ and T₃ levels.

HYPOTHYROIDISM

Hypothyroidism is the clinical syndrome that arises when there is a deficiency in the circulating levels of thyroid hormone. In neonates the disease is termed *cretinism* and is characterized by neurological impairment and mental retardation. Early treatment lessens the neurological deficits. Hypothyroidism also may be associated with Pendred's syndrome (deafness and hypothyroidism) and Turner's syndrome. In adults, onset of symptoms is insidious and the patient may be unaware of changes. Causes of hypothyroidism are listed in Table 36-2. The two principal causes of hypothyroidism in the United States are autoimmune thyroiditis and iatrogenic mechanisms such as thyroidectomy, radiation treatment, or medications. Iodine deficiency and dyshormonogenesis are other causes of hypothyroidism and goiter.

Clinical Manifestations

When the thyroid gland fails to develop or function in utero, children are born with cretinism and characteristic facies similar to those of Down syndrome and dwarfism (Fig. 36-16). Failure to thrive is apparent, and mental retardation often is severe. Immediate treatment with thyroid hormone at birth can lessen the neurological and intellectual deficits. Hypothyroidism at birth also can occur because of blocking antibodies from the mother. Hypothyroidism developing in childhood or adolescence is termed *juvenile hypothyroidism*; these children appear younger than their chronologic counterparts and may develop abdominal distention, umbilical hernia, and rectal prolapse. Mental performance may be impaired, but severe retardation is uncommon. Hypothyroidism secondary to autoimmune thyroiditis is far more prevalent in females (80 percent of cases). In adults symptoms in general are nonspecific, including tiredness, weight gain, cold intolerance, constipation, and menorrhagia.

Myxedema is the term given to severe hypothyroidism. In these patients facial features change because of the deposition of glycosaminoglycans in the subcutaneous tissues, leading to facial and periorbital puffiness. The skin becomes rough and dry and can develop a yellowish tinge from reduced conversion of carotene to vitamin A. Hair loss may be marked, with characteristic loss of the outer two-thirds of the eyebrows; remaining hair becomes dry and brittle. Enlargement of the tongue may impair speech, which is already slowed, in keeping with the impairment of mental processes. Untreated dementia may develop (*myxedema madness*). Abdominal symptoms may predominate. Patients may complain of a nonspecific, dull abdominal pain accompanied by distention and constipation. Libido and fertility are impaired in both sexes.

Cardiovascular changes include bradycardia and cardiomegaly, and a pericardial effusion might be present. Hypotension may be evident with a reduced cardiac output, and some patients develop shortness of breath and pulmonary effusions. Cardiac failure is uncommon. When hypothyroidism occurs as a result of pituitary failure and low TSH levels (secondary hypothyroidism), features of hypopituitarism may be present, such as pale, waxy skin, loss of body hair, and atrophic genitalia.

Laboratory Findings

Hypothyroidism is characterized by low circulating levels of T₄ and T₃. Raised TSH levels are found in primary thyroid failure, whereas in secondary hypothyroidism TSH levels are low. Secondary hyperthyroidism is rare and can be diagnosed by measuring TSH after a TRH challenge. The TSH level is low and does not increase in response to TRH. Autoimmune thyroid disease is characterized by the presence of thyroid autoantibodies (antithyroglobulin, antimitochondrial, or anti-thyroid-peroxidase [anti- TPO]). Other findings in hypothyroidism include anemia, diminished voltage with flattening or inversion of T waves on electrocardiogram, slow alpha waves with loss of amplitude on electroencephalogram, and raised levels of serum cholesterol (>300 mg/dL). In myxedema, comatose patients also have hyponatremia and CO₂ retention.

Treatment

Treatment of hypothyroidism is simple, inexpensive, and effective. Thyroxine is the treatment of choice and is administered in dosages varying from 50 mg to 200 mg per day. Patients are instructed to take tablets in the morning, usually without other medications, or at mealtime to assure good absorption and to avoid any sleep interference.

Young and otherwise healthy individuals tolerate initial starting doses of 100 mg of thyroxine per day, but elderly patients, patients with coexisting heart disease, and patients with profound hypothyroidism are less tolerant of thyroxine and should be started on a lower dose, such as 25 mg to 50 mg, slowly increasing the dose over weeks to months to attain a euthyroid state. An electrocardiogram should be obtained before treatment of patients with severe hypothyroidism for comparison if chest pain develops. Thyroxine dosage is titrated against TSH levels, which should return to normal. Thyroxine supplementation also must be determined by the clinical response of the patient. Whether or not patients with subclinical hypothyroidism (normal T₄, slightly raised TSH) should be treated is controversial. Evidence suggests that patients with subclinical hypothyroidism and increased antithyroid antibody levels should be treated, because they progress to more severe hypothyroidism. Patients with mild hypothyroidism may benefit from small doses of T₄, as the hypercholesterolemia, which accompanies hypothyroidism in this group of patients, is improved by therapy.

Patients who present with myxedema coma, in contrast to the patients with mild to moderate hypothyroidism, require emergency treatment with large doses of intravenous thyroxine

(400mg) followed by 100 mg/day. These patients usually are hyponatremic and hypocapnic and need careful monitoring in an intensive care unit.

THYROIDITIS

Autoimmune Lymphocytic Thyroiditis (Hashimoto's Thyroiditis)

Chronic lymphocytic thyroiditis, more commonly known as Hashimoto's thyroiditis or disease, after the physician who first described the condition in 1912, is an autoimmune thyroid disease and is the most common cause of hypothyroidism. It is ten times more common in women and more prevalent in the 30- to 60-year-old age group, with a prevalence of about 20 cases per 1000 women and an annual incidence of 1 to 2 new cases per 1000 women in the population. Autoimmune thyroiditis may be familial; up to 50 percent of first-degree relatives of patients with chronic autoimmune thyroiditis have thyroid antibodies inherited as a dominant trait. Chronic autoimmune thyroiditis is encountered in children but is rare in those under 5 years of age. In adolescents 40 percent of goiters are from autoimmune thyroiditis. Other predisposing conditions to autoimmune thyroiditis include Down syndrome, familial Alzheimer's disease, and Turner's syndrome. It is more common in areas of iodine excess. Studies suggest that thyroid cells in Hashimoto's thyroiditis have increased FAS receptors and that interleukin-1 induces abnormal FAS expression and triggers apoptosis or increased programmed thyroid cell death.

Pathology

In Hashimoto's disease the thyroid gland typically is firm and mildly enlarged. The enlargement usually is symmetrical. Frequently the pyramidal lobe also is enlarged. Histologically, there is follicular and Hürthle cell hyperplasia associated with lymphocytic and plasma cell infiltration and formation of lymphoid follicles. The disease is usually focal but gradually extends to involve the whole gland. Epithelial cell degeneration occurs with fragmentation of the basement membrane, and remaining epithelial cells enlarge and demonstrate oxyphilic changes (Hürthle or Askanazy cells). As lymphocytic infiltration progresses, the thyroid tissue degenerates and may be replaced by fibrous tissue.

Clinical Manifestations

Approximately 20 percent of patients with Hashimoto's thyroiditis present with signs and symptoms of hypothyroidism; a few patients present with hyperthyroidism (Hashitoxicosis). Most patients are euthyroid when the diagnosis is made. The most common presenting symptom is a tightness in the throat, often associated with a painless, nontender enlargement of the thyroid gland. Compression of the trachea or a recurrent laryngeal nerve is rare. Rapid enlargement of the thyroid gland should raise suspicion of thyroid lymphoma or carcinoma. Palpation usually demonstrates a diffusely enlarged, firm, often granular thyroid gland; in some cases the gland also is nodular. Usually the pyramidal lobe is enlarged. Evidence of other autoimmune conditions, such as disseminated lupus, rheumatic arthritis, and myasthenia gravis, may be present.

Diagnostic Findings

In early Hashimoto's thyroiditis, patients may present with a transient rise in serum thyroid hormone levels, but as the disease progresses, the serum TSH level rises as serum T₄ and T₃ levels fall. The diagnosis is confirmed by the presence of circulating antithyroid antibodies. These antibodies are directed against the membrane-bound enzyme involved in thyroid hormone synthesis, thyroid peroxidase (TPO), formerly called antimitochondrial antibodies, in almost 100 percent of patients and against thyroglobulin in about 50 percent of patients. FNAC examination of the thyroid gland occasionally is useful in confirming the diagnosis of Hashimoto's thyroiditis and in patients in whom malignancy is suspected.

Treatment

In the absence of compressive symptoms, patients demonstrating goiter, with or without evidence of hypothyroidism, are best treated with thyroid hormone. Reduction in thyroid goiter size with thyroxine treatment is variable but is more commonly seen in younger patients. Surgical intervention is indicated for patients complaining of obstructive symptoms, for cosmetically unacceptable goiters, or when thyroid cancer (other than lymphoma) is found. Thyroxine therapy with long-term follow-up monitoring of TSH levels is recommended.

Subacute Thyroiditis (De Quervain's Thyroiditis)

Subacute thyroiditis, also known as de Quervain's, granulomatous, or giant cell thyroiditis, is an uncommon, acute inflammatory disease of the thyroid. It is thought to be precipitated by a viral infection, although the exact cause is unknown. It is commonly encountered in North America but is relatively rare in the United Kingdom and Europe. The disease may be responsible for up to 10 percent of patients with hyperthyroidism in the United States. It affects women five times more often and usually is seen in patients 20 to 40 years of age.

Clinical Manifestations

Patients usually present with fever, malaise, and unilateral or bilateral thyroid pain and a recent history of an upper respiratory tract or viral infection may be given. Some patients complain of the symptoms of thyrotoxicosis, including palpitations, sweating, and heat intolerance, which are caused by the release of thyroid hormones from disrupted follicles in the inflamed thyroid gland. Palpation of the thyroid gland may reveal a tender, firm gland with mild unilateral or bilateral enlargement.

Pathology and Diagnostic Tests

Histologically, the disease is characterized by an acute inflammatory reaction of the thyroid gland. Degenerative thyroid follicles are surrounded by giant cells forming granulomas, which may be demonstrated on FNAC. Laboratory investigations demonstrate an elevated erythrocyte sedimentation rate (ESR) associated with a neutrophilia. Thyroid function tests usually show elevated levels of thyroid hormones (T₄ and T₃) with suppression of TSH. As the disease resolves, thyroid hormone levels return to normal, although the TSH level can remain low for some time. In contrast to Graves' disease, radioiodine uptake in the acute stage of the disease is low or negligible, because the released thyroid hormone, as result of inflammation, suppresses the serum TSH concentration.

Treatment

Usually treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) for pain relief is all that is necessary. Treatment with NSAIDs should be continued for several weeks after the disease has resolved in an effort to prevent recurrence. Beta blockers (e.g., propranolol) in the initial stages of the disease can be useful for relief of thyrotoxic symptoms. In the more severe cases it might be necessary to prescribe steroids for short periods. Prednisolone 40 mg once daily for 1 to 2 weeks, followed by a gradual reduction of the dose over the ensuing month, is recommended in such cases.

The disease usually lasts 1 to 6 weeks and resolves spontaneously. In some cases the disease lasts from several weeks or months and runs a course alternating between bouts of exacerbation followed by periods of remission. Most patients have complete resolution of the disease, although 10 percent of patients experience permanent hypothyroidism and require thyroxine replacement therapy.

Riedel's Thyroiditis

Riedel's thyroiditis is a rare disease of the thyroid characterized by a marked dense, invasive fibrosis that may extend beyond the thyroid capsule and involve surrounding structures. Fibrosis may involve the strap muscles, blood vessels, trachea, esophagus, and, on occasion, the parathyroid glands, which leads to hypoparathyroidism. Severe cases can result in the patient's becoming hypothyroid. The cause of the condition is unknown, but it may be part of a more generalized condition known as fibrosclerosis that causes fibrosis in other parts of the body, including the retroperitoneum, mediastinum, lacrimal glands, and bile ducts (sclerosing cholangitis).

Patients usually present with symptoms of compression such as hoarseness, stridor, and dyspnea. In more progressive cases involving the esophagus, dysphagia may be present. There often is rapid enlargement of the thyroid gland, which on palpation is "woody," hard, and nontender. Laboratory investigations usually are normal.

Riedel's thyroiditis resembles anaplastic thyroid cancer, except that the goiter is smaller. Diagnosis usually is established by FNAC, although open biopsy occasionally is needed. Treatment with tamoxifen and steroids often is helpful. Isthmectomy to relieve compressive symptoms or to establish the diagnosis is necessary in some patients. Most operations are difficult because of the loss of tissue planes and should only be embarked on by experienced surgeons. Thyroxine replacement therapy is necessary in patients with hypothyroidism.

Acute Suppurative Thyroiditis

Acute suppurative thyroiditis is rare. It is predominantly a disease of childhood or adolescence and is invariably associated with an acute upper respiratory tract infection. The disease is manifested by acute thyroid pain associated with dysphagia, fever, and, occasionally, rigors. The most common bacterial causative agents are streptococci, staphylococci, and pneumococci, but it also can be caused by *Escherichia coli* and *Coccidioides immitis*.

Suppuration usually is unilateral but may extend into the deep spaces in the neck, invading the trachea, esophagus, or mediastinum. FNAC with smear and culture is diagnostic. Treatment consists of intravenous antibiotics and drainage of any abscess. Thyroid lobectomy rarely is required. Most patients recover completely and are euthyroid.

GOITER

Simple or nontoxic goiter is an enlargement of the thyroid gland in a euthyroid patient, not associated with any neoplastic or inflammatory process. It may be diffuse and symmetrical or nodular. Several forms of goiter have been described.

Familial Goiter

Familial goiters usually are regarded as goiters caused by an inherited enzymatic defect (dyshormonogenesis) that may cause impairment of iodine accumulation, organification, or coupling of iodotyrosine in the thyroid gland. The inborn error of metabolism generally is inherited as an autosomal recessive trait, but dominant traits have been described. Familial goiters usually are associated with hypothyroidism, although patients may remain euthyroid. Familial goiter also is associated with deafness (Pendred's syndrome).

Endemic Goiter

Endemic goiter is defined as thyroid enlargement affecting a significant number of inhabitants of a particular locale. The most important factor in the development of this condition is iodine deficiency. It is most commonly encountered in mountainous areas where the iodine content of drinking water is particularly low. Most countries throughout the world have had one or more areas where endemic goiter was encountered; in the United States it was formerly in the Midwestern mountainous regions. Administration of iodine, usually as an additive in table salt, has proved successful as a prophylaxis in reducing the incidence of this condition.

Sporadic Goiter

Sporadic goiter is the term given to a goiter for which no definitive cause can be established. It excludes goiters caused from thyroiditis and neoplasia as well as endemic goiter.

Pathology

The thyroid gland may be diffusely enlarged and smooth, or enlarged and markedly nodular. In the early stages of the disease, the gland may be hyperplastic and diffusely

enlarged, a condition that may be reversed by the administration of iodine or thyroid hormone. Nontoxic nodular goiter is a multinodular gland in which the nodules vary considerably in size and number. Nodules are filled with gelatinous, colloid-rich material, and scattered between nodules are areas of normal thyroid tissue. Gross or microscopic cyst formation may be present, with evidence of degeneration, hemorrhage, and calcification.

Clinical Manifestations

Most patients with goiters are asymptomatic. The most common symptom is a sensation of pressure in the neck coupled with a mass. If the goiter enlarges significantly, patients may complain of compressive symptoms such as dysphagia or dyspnea. Paralysis of a recurrent laryngeal nerve is rare and should raise the suspicion of malignancy. On occasion a recurrent laryngeal nerve is stretched over a rapidly enlarging thyroid nodular cyst and ceases to function. Goiters may extend into the thorax and become retrosternal (Fig. 36-17), which may be associated with an impedance of venous return in the jugular veins (Fig. 36-18) and consequent facial flushing. Such flushing is accentuated by the patient's raising his or her arms above the head (positive Pemberton's sign). Sudden pain, frequently associated with rapid enlargement of the thyroid gland, usually is related to hemorrhage into a colloid nodule or cyst.

Examination reveals a diffusely enlarged, soft thyroid goiter in patients with simple goiter or an enlarged gland with nodules of varying size and firmness in multinodular goiter (Fig. 36-19). In patients in whom one nodule predominates, or is painful, or has recently enlarged, FNAC is recommended because it is sensitive and specific in the diagnosis of colloid nodule.

Results of laboratory investigations usually are normal, although in patients over the age of 60 years with long-standing multinodular goiters (>17 years), a significant number develop thyrotoxicosis (Plummer's disease). Most of these hyperthyroid patients have a suppressed TSH and increased T₃ level but normal T₄ (T₃toxicosis).

Treatment

Most euthyroid patients with small, diffuse, simple goiters need no treatment. If the goiter is of significant size, treatment with thyroxine may depress TSH stimulation of the thyroid gland and reduce hyperplasia, decreasing goiter size or preventing increases in size. Endemic goiter is managed by the administration of iodine. Surgery should be reserved for patients with cosmetically unacceptable goiters, compressive symptoms, or retrosternal goiters or when malignancy is suspected or demonstrated on FNAC.

SOLITARY OR DOMINANT THYROID NODULE

Solitary thyroid nodules are present in about 4 percent of the population of the United States. Thyroid cancer has an incidence of 40 new cases per 1,000,000 people. The vast majority of thyroid nodules are benign and do not require removal. The physician or surgeon should be able to perform an accurate clinical assessment of any thyroid nodule, appreciate the risk factors of thyroid cancer, and be able to evaluate which patients would benefit from surgery.

Risk Factors for Cancer

Two groups are at risk for having carcinoma of the thyroid: patients exposed to low-dose radiation to the head and neck regions and patients in whose family another member has developed thyroid cancer. Medullary thyroid cancer, which is the greatest risk factor because it may be inherited as an autosomal dominant trait, can be diagnosed by testing for a RET point mutation. Approximately 6 percent of patients with papillary thyroid cancer have familial disease. Papillary thyroid cancer also occurs more commonly in patients with Cowden's syndrome and familial polyposis.

Irradiation of the Head and Neck

A history of low-dose ionizing radiation (<2000 rad) to the thyroid gland places the patient at increased risk for developing thyroid cancer. The risk increases linearly from 6.5 to 2000 cGy; beyond this the incidence declines as the radiation causes destruction of the thyroid tissue, although patients who receive 4000 cGy for treatment of Hodgkin's disease have been shown to have a higher incidence of thyroid cancer. Low-dose therapeutic radiation was used to treat conditions such as "enlarged thymus" in an effort to prevent "sudden crib death" (150 cGy), tonsils and adenoids to prevent the need for tonsillectomy (750 cGy), acne vulgaris (1500 cGy) and other conditions such as hemangioma, ringworm, and scrofula.

There is a 40 percent chance that patients presenting with a thyroid nodule and a history of radiation have thyroid cancer. Of those patients who have thyroid cancer, the cancer is located in the nodule in 60 percent of patients and in the remaining 40 percent of patients the cancer is in another area of the thyroid gland. In children living near the Chernobyl nuclear power plant, where an accident caused radiation leakage in 1986, there has been a 60- fold increase in thyroid cancer, with most cancers occurring in those who were infants at the time of exposure.

Thyroid cancer in patients with a history of radiation exposure tends to be of the papillary type and frequently is multifocal with a higher incidence of lymph node metastases. To prevent higher doses of radiation exposure to the thyroid, scanning with ¹²³I or ^{99m}Tc has largely replaced scanning with ¹³¹I in most hospitals.

Evaluation

Most thyroid nodules are benign and are colloid nodules or adenomas. To distinguish between suspicious thyroid nodules that may be malignant and benign thyroid nodules, certain issues should be considered.

Because 40 percent of patients with a solitary nodule or multiple thyroid nodules with a history of previous head and neck irradiation will have cancer of the thyroid gland, a history of external radiation to the head or neck is important. The age and gender of the patient are also important. Thyroid nodules in children and elderly patients are more likely to be malignant. Solitary nodules also are more common in men over 40 and women over 50 years of age. Signs that should prompt concern are rapid enlargement of an old or new thyroid nodule, symptoms of local invasion (e.g., unilateral vocal cord paralysis), or compressive symptoms (e.g., dysphagia and dyspnea from invasion into the trachea and esophagus).

Approximately 15 percent of solitary thyroid nodules are malignant, and in hard solitary thyroid nodules the risk of malignancy is two or three times higher. When the apparently solitary nodule is found to be a dominant nodule in a multinodular thyroid gland, the incidence of thyroid cancer in this nodule is less than 5 percent. An exception to this is a history of previous head and neck irradiation or a history of familial thyroid cancer. Lesions that are hard, gritty, or fixed to surrounding structures such as the trachea or strap muscles are probably malignant. The presence of palpable cervical lymphadenopathy adjacent to a thyroid nodule should be suspected as a carcinoma. Cervical lymphadenopathy because of metastatic thyroid cancer may be present when no thyroid nodule is palpable. These patients almost always have a thyroid cancer that usually is present in the ipsilateral thyroid lobe. This cancer may be microscopic and demonstrated only after careful sectioning of the thyroid gland.

Investigations

Fine-Needle Aspiration Cytology

The procedure of choice in evaluating thyroid nodules is fine-needle aspiration cytology (FNAC). The test is fast, is minimally invasive, has few risks, causes little discomfort to the patient, and has been shown to be specific and sensitive. It has replaced radionuclide and ultrasonographic imaging in the preoperative evaluation of thyroid nodules.

Ninety percent of nodules can be categorized into the following groups: *benign*, 65 percent; *suspicious*, 15 percent; *malignant*, 5 percent; and *nondiagnostic*, 15 percent. The incidence of false-positive results is about 1 percent, and false-negative results, about 5 percent.

FNAC is the investigation of choice in patients with thyroid nodules, but it has limitations. The proportion of patients with suspicious cytology results ranges from 11 to 20 percent. Most have follicular or Hürthle cell neoplasms, and 20 percent of these patients have a thyroid malignancy. There is no accurate method of predicting which of these patients has a thyroid cancer, because the diagnosis depends on demonstrating capsular or vascular invasion, which is not possible on FNAC. The technique also is less reliable in patients who have previous irradiation to the head and neck or have a family history of thyroid cancer, as many of these patients have multifocal lesions.

When a cyst is encountered on FNAC it should be drained completely, which is curative in about 75 percent of simple cysts (Fig. 36-20), although some require a second or third aspiration. If the cyst persists after three attempts at aspiration, unilateral thyroid lobectomy is recommended. In cases in which the cyst is larger than 4 cm in diameter or is complex (i.e., has solid and cystic components), thyroid lobectomy is recommended, because these cysts have a higher incidence of malignancy (15 percent).

Thyroid Imaging

Ultrasonographic Imaging

Ultrasound evaluation of the thyroid gland is helpful for detecting nonpalpable nodules and for differentiating solid from cystic nodules. It provides a noninvasive and inexpensive method of following the size of suspected benign nodules diagnosed by FNAC. This is especially useful in determining whether the nodule or the surrounding normal thyroid tissue decreases in size when the patient is treated with thyroid hormone. Ultrasonographic imaging has no role in screening for thyroid nodules in asymptomatic patients.

CT and MRI

Computed tomography and magnetic resonance imaging usually are unnecessary in the evaluation of thyroid tumors except for large or retrosternal lesions and for assessing suspected invasion into surrounding structures. MRI is more accurate than CT in distinguishing recurrent or persistent thyroid tumor from postoperative fibrosis.

Thyroid Isotope Scanning

Scanning the thyroid with ¹²³I or ^{99m}Tc can indicate the functional activity of a nodule and of the thyroid and correlate the location of palpable nodules with the nodules seen with scanning. Scanning with ¹³¹I has been largely replaced by ¹²³I or ^{99m}Tc scans to lower the dose of radiation delivered to the thyroid gland during the investigation. Nodules that trap less iodine than the surrounding thyroid tissue are termed "cold," nonfunctional, or hypofunctional (Fig. 36-21). Almost 85 percent of nodules on scanning are cold, and these lesions have a 10 to 25 percent chance of malignancy. Of the 5 percent of nodules shown to be "hot" on scanning, approximately 1 percent are malignant. Thyroid scanning is recommended in the assessment of thyroid nodules only in those patients who have follicular thyroid nodules on FNAC.

Laboratory Findings

Thyroid function tests are not useful in the assessment of patients with thyroid nodules, because most patients with thyroid cancer are euthyroid. In those patients presenting with hyperthyroidism and a solitary nodule, the chances of the nodule's being malignant are very low. Serum thyroglobulin levels cannot differentiate benign from malignant thyroid nodules unless the levels are extremely high, in which case metastatic thyroid cancer should be suspected. Thyroglobulin levels are useful in following patients who have undergone total thyroidectomy for thyroid cancer, excluding medullary thyroid cancer, and also for monitoring patients with nodules being followed nonoperatively. Serum calcitonin levels should be obtained in patients with a family history of medullary thyroid cancer or *multiple endocrine neoplasia type II (MEN II)* and where FNAC demonstrates medullary

thyroid cancer to be present. Patients who are RET oncogene positive should always have a 24-h urine collection with measurements of levels of vanillylmandelic acid (VMA), metanephrine, and catecholamine to rule out a coexisting pheochromocytoma.

Treatment

The algorithm for the diagnosis of a thyroid nodule is shown in Fig. 36-22. When a diagnosis of colloid nodule is made cytologically, thyroidectomy is not necessary except for cosmetic or symptomatic reasons, and these patients can be safely observed. A second FNAC is recommended 6 to 8 months after the initial FNAC if the nodule enlarges. Patients with benign nodules should undergo ultrasound determination of the size of the lesion and a baseline thyroglobulin level should be obtained. Patients may be prescribed thyroxine in doses sufficient to maintain a serum TSH level between 0.1 and 1.0 mU/mL. Approximately 50 percent of these nodules significantly decrease in size in response to the TSH suppression of this regimen. Thyroidectomy should be performed if a nodule enlarges or develops in a patient whose serum thyroglobulin increases despite TSH suppression therapy with thyroxine. An exception to this general rule is the patient who has had previous irradiation of the thyroid gland. In these patients total or near- total thyroidectomy is recommended with FNAC because of the high incidence of thyroid cancer (40 percent).

Thyroid lobectomy is recommended in patients in whom a thyroid cyst has recurred after three aspirations and in patients with a cyst that is larger than 4 cm or is found to be a complex cyst on ultrasound examination. Signs and symptoms such as pressure, hoarseness, dysphagia, vocal cord paralysis, dyspnea, adjacent cervical lymphadenopathy, or recent rapid increase in size of a thyroid nodule (unless FNAC demonstrates hemorrhage) all favor thyroid lobectomy. Patients with a thyroid nodule and a family history of papillary or Hürthle cell cancer probably warrant thyroidectomy. Patients with medullary carcinoma or MEN II with an elevated basal or stimulated calcitonin level or positive RET oncogene also should have a total thyroidectomy, even in the absence of a palpable mass.

MALIGNANT TUMORS

In the United States thyroid cancer occurs in about 40 per 1,000,000 people each year and accounts for less than 1 percent of all malignancies. Thyroid cancer is responsible for six deaths per million persons annually. Most patients present with a palpable swelling in the neck, which initiates assessment through a combination of history, palpation, and FNAC examination. Diagnosed correctly and managed properly, most patients with differentiated thyroid carcinoma of follicular cell origin are curable.

Several different types of primary thyroid carcinoma are recognized, although 90 to 95 percent of thyroid carcinomas are *differentiated* tumors of follicular cell origin (differentiated thyroid cancer) giving rise to papillary, follicular, or Hürthle cell carcinoma. About 6 percent of patients with papillary and Hürthle cell cancer have familial disease; follicular and anaplastic thyroid cancer are more common in patients from iodine- deficient areas. Medullary thyroid cancers account for about 6 percent of thyroid cancers, and approximately 30 percent of these tumors are familial.

Molecular Basis of Thyroid Tumors

Oncogene (*onc*) is the title given to a gene that contributes directly to tumor genesis. Cellular oncogenes (*c-onc*) or proto-oncogenes are cellular genes that can give rise to oncogenes when they are altered or when their expression is modified. The modified genes may encode for a mutated receptor on the cell surface, resulting in an abnormal growth signal being transmitted to the nucleus through overexpression of the receptor or by the receptor's being continually (constitutively) activated. Inappropriate cell division results and may lead to malignant transformation. The fact that these genes have survived in the evolutionary process and play a role in cell transformation suggests that proto-oncogenes are involved in the normal control of cell growth and differentiation.

Several oncogenes are involved in thyroid tumor genesis. Mutations in TSH-receptor and Gs genes are found in hyperfunctioning thyroid nodules, for example, and probably constitutively activate the adenylate cyclase–protein kinase A signal transduction pathway, leading to a well- differentiated tumor. An oncogene that has a significant role in the development of papillary thyroid cancer is the RET oncogene. Rearrangement of this gene occurs in *papillary thyroid cancer (PTC)* and is more common in childhood thyroid cancers. Point mutations involving RET have been demonstrated in patients with familial medullary thyroid cancer, MEN IIA, and MEN IIB.

The RET proto-oncogene encodes for a tyrosine kinase receptor on the cytoplasmic membrane; the ligand has been identified as glial cell line–derived neurotrophic factor (GDNF). RET is expressed during embryogenesis in the nervous and excretory systems. Disruption of RET results in developmental abnormalities in these systems, including the enteric nervous system (Hirschsprung's disease). Tumors arising from the neural crest cells (medullary thyroid cancer and pheochromocytoma) also contain RET point mutations. The RET gene is located on chromosome 10, and rearrangement of RET by fusion with heterologous genes creates transforming oncogenes, which have been implicated in the pathogenesis of PTC. These oncogenes are RET/PTC1, RET/PTC2, and RET/PTC3.

Not all patients with PTC possess the RET/PTC gene, and expression markedly varies geographically, e.g., being low in Japan (3 percent) and high in Italy (34.6 percent). It also is more common in thyroid cancers in children that develop after radiation exposure. In vitro irradiation induces RET/PTC formation, and 66 percent of the papillary thyroid cancers that were removed from patients living near Chernobyl demonstrated RET/PTC1 and RET/PTC3.

A second oncogene associated with PTC is TRK-A, found on chromosome 1, which encodes for a cell surface receptor for nerve growth factor. Mutated *ras* oncogenes have been demonstrated in a wide variety of human tumors and occur more frequently in human thyroid tumors of follicular origin. Point mutations in *ras* occur in thyroid adenomas and multinodular goiters, as well as follicular carcinoma, and are believed to be an early mutation.

A molecular defect associated with the development of follicular adenoma is allelic loss of genes on the short arm of chromosome 11 (11q). Deletion of 3p accompanies transformation from follicular adenoma to a follicular adenocarcinoma.

The p53 gene is a tumor suppressor gene encoding for a protein (p53) that acts as a transcriptional regulator. Its function is to maintain genomic integrity. Point mutations in the p53 gene result in a protein that lacks this ability, which in turn results in proliferation of malignant cells. A mutated p53 gene is found in most anaplastic thyroid cancers and in thyroid cell lines.

Papillary Carcinoma

Papillary carcinoma is the most common of thyroid malignancies, accounting for 80 percent of all thyroid cancers. It is the predominant thyroid cancer in children (75 percent) and in people who were previously exposed to radiation in the neck (85 to 90 percent). Papillary carcinoma occurs more often in women, with a 2:1 female-to-male ratio, and the mean age at presentation is 30 to 40 years.

Pathology

Papillary thyroid cancers, including mixed papillary follicular carcinoma and the follicular variant of papillary carcinoma, usually are hard and whitish and remain flat on sectioning with a blade, rather than bulging as does normal tissue or benign nodular lesions. Macroscopic calcification, necrosis, or cystic change may be apparent. Smaller tumors occasionally may be present inside thyroid cysts.

Histologically, papillary carcinomas may exhibit papillary projections, a mixed pattern (papillary/follicular structures), or a pure follicular pattern with intranuclear inclusions. The diagnosis is established by characteristic cellular features. Cells are cuboidal with pale, abundant cytoplasm, crowded nuclei, and intranuclear cytoplasmic inclusions, the so-called *Orphan Annie cells* (Fig. 36-23). A characteristic fibrovascular stroma with calcium deposits (*psammoma bodies*) may be present. Tumors that exhibit papillary features such as *mixed tumor* or the *follicular variant* of papillary carcinoma should be classified as papillary carcinomas because they act biologically as papillary carcinomas. Multifocality in papillary carcinoma is common (30 to 87.5 percent) and is a minor risk factor in prognosis. Papillary cancers exhibit a propensity for lymphatic spread within the thyroid and to local lymph nodes (paratracheal and cervical chain). Papillary tumors may invade local structures such as the trachea, esophagus, and recurrent laryngeal nerves and rarely are encapsulated. Blood-borne metastases usually are a late feature.

Macroscopically, there are three recognized forms of papillary thyroid cancer, each based on the size and extent of the primary disease. *Minimal or occult/microcarcinoma* tumors originally included papillary cancers up to 1.5 cm in diameter. They now are defined as tumors of 1 cm or less in size with no evidence of local invasiveness through the thyroid capsule and that are not associated with lymph node metastases. They are nonpalpable and usually are incidental findings at operative, histologic, or autopsy examination. Studies have demonstrated occult papillary thyroid cancer to be present in 2 to 36 percent of thyroid glands removed at autopsy. Recurrence rate in patients with tumors 1.5 cm or smaller after removal is about 7 percent and the mortality about 0.5 percent. *Intrathyroidal* tumors are greater than 1 cm and are confined to the thyroid gland, with no evidence of extrathyroid invasion. *Extrathyroidal* tumors are locally advanced with invasion through the thyroid capsule into adjacent structures. All types of primary thyroid cancers can be associated with lymph node metastases and invasion into intrathyroidal blood vessels or occasionally distant metastases. Long-term prognosis is better for intrathyroidal lesions.

Other types of papillary carcinoma of the thyroid are: tall cell, columnar, diffuse sclerosing, clear cell, trabecular, Hürthle cell, and poorly differentiated variants. Most of these variants behave more aggressively. Angioinvasion also predicts more recurrences and a worse prognosis, but coexisting lymphocytic thyroiditis suggests fewer recurrences and a better prognosis. There usually is no significant difference in management.

Clinical Manifestations

Most patients are euthyroid and present with a slow-growing painless mass in the neck. Accompanying symptoms, which usually are associated with advanced locally invasive disease, may include dysphagia, dyspnea, and hoarseness. Ipsilateral enlarged cervical lymph glands also may be present and most commonly are encountered in younger patients. Nodal involvement may be more apparent than the primary lesion; the so-called *lateral aberrant thyroid* usually is a cervical lymph node that has been invaded by metastatic cancer.

Distant metastases are uncommon at initial presentation (1 to 15 percent), depending on whether detected by physical examination and symptoms or by radioiodine scanning and thyroglobulin determination after thyroidectomy. Distant metastases to the lung are more commonly encountered in children, although up to 20 percent of all patients ultimately develop distant disease (Fig. 36-24).

Thyroid cancer often is suspected by history or physical examination. In most cases the diagnosis is easily established in the outpatient setting by FNAC, which is specific and sensitive in papillary, medullary, and anaplastic thyroid cancer. Radioiodine thyroid scans are not necessary in preoperative evaluation. CT and MRI are used selectively in patients with extensive local or substernal disease or lymph node involvement.

Prognostic Indicators

Ideal treatment objectives for thyroid cancer are detailed in Table [36-3](#). In order to compare the merits of different surgical therapies, a method of accurately assessing various prognostic factors is necessary.

AGES Scale

This is a postoperative prognostic scale of risk of dying from papillary cancer. Factors taken into account are Age, pathologic tumor Grade, Extent of disease and Size of tumor. Two groups are identified with this system: low-risk patients (young, with well-differentiated tumors, no metastases, and small primary lesions) and high-risk patients (older, with poorly differentiated tumors, local invasion, distant metastases, and large primary lesions). Of 860 patients in one study, 85 percent were in the low- risk group, with a cancer risk mortality of 2 percent at 25 years; the cancer risk mortality in the high-risk group was 46 percent.

MACIS Scale

This more sophisticated postoperative scale is a modification of the AGES system. Factors assessed are distant Metastases, Age at presentation (<40 or >40 years), Completeness of original surgical resection, extrathyroidal Invasion, and Size of original lesion (in cm) (Fig. [36-25](#)).

DNA Ploidy

Measurement of DNA ploidy is of some use in prognosis because aneuploidy is noted in 10 percent of all papillary carcinomas but is present in 25 to 50 percent of patients who die from thyroid cancers. Aneuploid DNA measurements are encountered infrequently in papillary thyroid cancers among survivors.

The most significant single prognostic indicator overall is distant metastases, especially to bone. Local invasion of the primary tumor through the thyroid capsule into the adjacent structures increases the mortality tenfold over matched patients with intrathyroidal tumors.

The other frequently used classification systems are AMES (Age, Metastases, Extent, Size) and TNM (Tumor size, Nodal involvement, and Metastases to distant sites). Patients less than 45 years of age without distant metastases usually are considered to be at low risk.

Surgical Treatment

When patients are found to have a *minimal* papillary thyroid carcinoma in a thyroid specimen removed for other reasons, unilateral thyroid lobectomy and isthmectomy usually is sufficient surgical treatment unless there is angioinvasion or the tumor is at the margins of resection. Similarly, in rare cases of encapsulated papillary thyroid carcinoma, total lobectomy is all that is needed. In all other patients with papillary thyroid cancer, total or near- total thyroidectomy is the procedure of choice.

The Case for Total Thyroidectomy

There are several arguments for the treatment of papillary carcinoma by total thyroidectomy: multifocal disease, decreased incidence of local recurrence, reduced risk of anaplasia in any residual tissue, facilitation of diagnosing unsuspected metastatic disease by radioiodine scanning or treatment with ¹³¹I (reducing incidence of distant recurrence), and greater sensitivity of blood thyroglobulin levels to predict recurrent or persistent disease.

Multifocal disease in papillary carcinoma is well documented (up to 85 percent). Unilateral lobectomy therefore may leave residual disease in the contralateral lobe. Recurrence rates in the contralateral lobe after thyroid lobectomy range from 4.2 to 26 percent. In one study, the recurrence rate was 19 percent at 10 years after lobectomy versus 11 percent in similar patients treated with total thyroidectomy.

Whether total thyroidectomy has any survival advantage over near-total thyroidectomy (total lobectomy with subtotal resection of the contralateral lobe) is unclear. Studies from the Mayo Clinic demonstrated that in low- and high-risk AGES categories local recurrence was higher after lobectomy than after total or near-total resection. Total thyroidectomy did not have better survival rates than near total thyroidectomy. It is suggested that subtotal resection on the side opposite the tumor decreases the chances of damage to the parathyroid and the recurrent laryngeal nerve. This argument is valid in the case of patients whose parathyroid glands are anteriorly located on the thyroid, so that some thyroid tissue is left to preserve the blood supply to the parathyroid glands. When the parathyroid glands are situated off the thyroid gland, there is little benefit in leaving thyroid tissue. In experienced hands the incidence of permanent hypoparathyroidism or nerve damage after total thyroidectomy is about 1 percent. Differentiation of the residual thyroid cancer to anaplastic carcinoma occurs in about 1 percent of patients and is almost always fatal.

Total thyroidectomy, by removing all residual normal thyroid tissue, facilitates the uptake of radioactive ¹³¹I extrinsic to the gland and helps identify and ablate metastatic thyroid cancer. Once all thyroid tissue has been removed, thyroglobulin levels should remain below 3 ng/mL; by monitoring thyroglobulin levels, disease progress can be followed. When thyroglobulin levels increase, recurrent disease must be present and appropriate screening with ¹³¹I should be done.

The patient with a thyroid nodule that is suspected to be papillary thyroid cancer should have FNAC performed. When papillary thyroid cancer is diagnosed, the definitive operation can be done without confirming the diagnosis on histologic examination. Patients with a nodule that might be papillary cancer should be treated by thyroid lobectomy, isthmectomy, and removal of any pyramidal lobe or adjacent lymph nodes. When intraoperative frozen-section examination of a lymph node or of the primary tumor confirms carcinoma, completion total or near-total thyroidectomy should be performed. If a definitive diagnosis cannot be made, the operation is terminated. When final histology confirms carcinoma, completion total thyroidectomy usually should be performed. For patients who have minimal papillary thyroid cancers confined to the thyroid gland without angioinvasion, no further operative treatment is recommended. Treatment with thyroid hormone to suppress TSH is recommended. Patients at low risk should have serum TSH levels between 0.1 and 1.0 mU/mL, whereas patients at high risk should have serum TSH levels suppressed to less than 0.1 mU/mL.

Local recurrence is a serious complication, with a disease-related mortality rate of 33 to 50 percent. Patients with nodal recurrence usually do better than those with tumor recurrence in the thyroid bed or with distant metastases. Lymph node metastases in the lateral neck in patients with papillary carcinoma usually should be managed with modified radical neck dissection, and en-bloc dissection of all fibrofatty tissue and lymphatic tissue. The jugular lymph node chain is removed, while preserving the sternocleidomastoid muscle, the internal jugular vein, and the spinal accessory nerve (which are all taken in a radical dissection). Dissection of the posterior triangle and suprahyoid dissection usually are not necessary but should be performed when appropriate. Prophylactic neck node dissection is not necessary in patients with papillary thyroid cancer, because micrometastases appear to be ablated with radioiodine therapy. Survival rates for the different treatment modalities are summarized in Fig. [36-26](#).

Follicular Carcinoma

Follicular carcinoma, which accounts for about 10 percent of thyroid malignancies, is decreasing in incidence in the United States. This cancer occurs more often in women, with a female-to-male ratio of 3:1, and it presents in an older age group, mean age at presentation of 50 years, compared to a mean age of 35 in patients with papillary carcinoma. Follicular carcinoma occurs more frequently in iodine-deficient areas, and a rare form of familial disease is reported in patients with dys hormonogenesis.

Pathology

Follicular carcinoma usually is solitary and approximately 90 percent are surrounded by a tumor capsule. Follicular carcinoma differs from papillary carcinoma in that vascular invasion and hematogenous spread to bone, lung, and liver is seen more commonly than lymphatic spread, which usually is encountered in late stages. Lymph node metastases occur in less than 10 percent of patients. Histologically, follicles are present, but the lumen may be devoid of colloid. Architectural patterns depend on the degree of differentiation demonstrated by the tumor.

Two types of follicular carcinoma are recognized. In *minimally invasive tumors* there is evidence of invasion into but not through the tumor capsule at one or more spots. Previously, these tumors may have been reported as *atypical adenomas*, although there are reports of metastases occurring, and they should be regarded as low-grade carcinomas. In the second type, *frankly invasive tumors*, there is evidence of vascular invasion or tumor invasion through the tumor capsule. Tumor infiltration and invasion may be apparent at surgery, with tumor thrombus in the middle thyroid or jugular veins.

Clinical Manifestations

Follicular cancers usually present as solitary thyroid nodules, occasionally with a history of rapid size increase, and long-standing goiter. When hemorrhage into the nodule has occurred, pain may be a presenting feature, but usually the tumors are painless. At initial presentation cervical lymphadenopathy is uncommon, although distant metastases are more frequently encountered than papillary carcinoma. Rarely, follicular cancers may be hyperfunctioning (1 percent), and patients may present with signs and symptoms of thyrotoxicosis.

Unless distant metastases have been confirmed as follicular thyroid cancer, definitive preoperative diagnosis usually is not possible in follicular thyroid carcinoma. This is because FNAC is unable to distinguish between follicular cells from a benign follicular adenoma and from a carcinoma. Approximately 20 percent of all FNAC aspirates showing follicular cells are carcinomas; the remaining 80 percent are adenomas.

Surgical Treatment and Prognosis

Patients diagnosed by FNAC as having a follicular lesion should undergo thyroid lobectomy, including the isthmus and the pyramidal lobe. Intraoperative frozen-section examination usually is not helpful but should be performed when there is evidence of capsular or vascular invasion or when adjacent lymphadenopathy is present. Total thyroidectomy should be performed when thyroid cancer is diagnosed, except in patients with minimally invasive follicular cancers because the prognosis in these patients is so good. In older patients, when follicular neoplasms diagnosed by FNAC are greater than 4 cm and macroscopically suspicious for cancer, some surgeons proceed with total thyroidectomy, because the risk of cancer is about 50 percent. A diagnosis of invasive carcinoma necessitates completion total thyroidectomy primarily so that ¹³¹I can be used to detect and ablate metastatic disease. Prophylactic nodal dissection is unwarranted because nodal involvement is infrequent, but therapeutic neck dissection is recommended.

The cumulative percentage of patient mortality from follicular thyroid cancer is about 15 percent at 10 years and 30 percent at 20 years. Factors that significantly worsen long-term prognosis are; age over 50 years at presentation, tumor size greater than 4 cm, higher tumor grade, marked vascular invasion, extrathyroidal invasion, and distant metastases at the time of diagnosis.

Hürthle Cell Carcinoma

Hürthle cell tumors are considered by the World Health Organization classification to be a variant of follicular cell neoplasms. Tumors contain sheets of eosinophilic cells packed with mitochondria. They are derived from the oxyphilic cells of the thyroid gland, and although the function of the cells is unknown, Hürthle cell tumors possess TSH receptors and produce thyroglobulin. Only about 10 percent of these tumors trap radioiodine.

Hürthle cell carcinomas account for approximately 3 percent of thyroid malignancies. They differ from follicular carcinoma in that they are more often multifocal and bilateral, are more likely to metastasize to local nodes (25 percent), and usually do not take up ¹³¹I. Hürthle cell neoplasms usually are diagnosed by FNAC, and about 20 percent of these tumors are found to be malignant. There is controversy as to whether all Hürthle cell neoplasms are malignant. Grant and colleagues reported that the mortality rate was less than 1 percent of 642 patients with Hürthle cell adenoma. Adenomas, as in follicular neoplasms, have no vascular or capsular invasion, whereas cancers do have vascular or capsular invasion.

Management is similar to that of follicular neoplasms, with lobectomy, isthmectomy, and pyramidal lobe resection being sufficient surgical treatment for unilateral Hürthle cell adenomas. When Hürthle cell neoplasms are found to be invasive on intraoperative frozen-section or definitive paraffin-section histology, then total thyroidectomy should be performed. These patients should also undergo routine central neck node removal, similar to patients with medullary thyroid cancer, and modified radical neck dissection when lateral neck nodes are palpable. T₄ treatment is recommended postoperatively.

Although radioiodine scanning and ablation usually are ineffective, they probably should be considered to ablate any residual normal thyroid tissue and occasionally ablate tumor, since there is no other good therapy. ^{99m}Tc- sestamibi scanning has been reported to be useful for detecting persistent local or metastatic disease. The mortality rate from Hürthle cell cancer is higher than from follicular thyroid cancer.

Medullary Carcinoma

Medullary thyroid cancer (MTC) accounts for about 5 percent of thyroid carcinomas. Although MTC was originally classified as an anaplastic tumor, Hazard and associates recognized that it was a separate entity, because patients survived longer than they did with anaplastic carcinoma. Williams determined that medullary thyroid cancer arose from the parafollicular or C cells of the thyroid. These cells are neuroectodermal and originate from the ultimobranchial bodies of the fourth and fifth branchial pouches. They descend to join the thyroid gland proper and are concentrated mainly in the superior poles laterally. This also is where C-cell hyperplasia is most evident and where medullary thyroid cancers develop. Pearse demonstrated the ability of the C cells or parafollicular cells to secrete calcitonin, a 32- amino-acid polypeptide that has an opposing action to parathyroid hormone (PTH) in that it is an antihypercalcemic hormone and lowers serum calcium levels. In some animals, especially those that lay eggs with shells, calcitonin is a significant regulator of calcium metabolism, but in human beings it has only minimal physiologic effects.

In 1966 Sipple described a postmortem case in which thyroid cancer, hyperplastic parathyroid glands, and pheochromocytoma coexisted; additional review of 500 cases of pheochromocytoma showed thyroid cancer to be six to fourteen times more frequently associated with this condition than expected. Since these initial observations, families have been identified whose members have only MTC; have MTC, hyperparathyroidism, and pheochromocytoma (MEN IIA or Sipple's syndrome), and patients with MTC, pheochromocytoma, ganglioneuromatosis, and Marfan's syndrome (MEN IIB). Patients with MEN IIA also may have Hirschsprung's disease and lichen cutaneous amyloidosis. Patients with von Hippel–Lindau disease also appear to be prone to developing pheochromocytoma and medullary thyroid cancer.

Pathology

Medullary carcinomas are located in the middle to upper thyroid poles (in keeping with the embryological origin of the C cells). Tumors typically are unilateral (75 percent). With familial cases, C-cell hyperplasia and multicentricity are prevalent, and 90 percent of the patients have bilateral tumors.

Microscopically, tumors are composed of sheets of infiltrating neoplastic cells separated by collagen and amyloid. Marked heterogeneity is present; cells may be polygonal or spindle-shaped. The presence of amyloid is a diagnostic finding, but immunohistochemistry for calcitonin has superseded it as a diagnostic tumor marker. C-cell hyperplasia is a premalignant precursor of MTC. These tumors also stain positively for *carcinoembryonic antigen (CEA)*, histaminase, and calcitonin gene–related peptide (CGRP).

Tumors spread initially to local lymph nodes in the neck and the superior mediastinum. Distant blood-borne metastases to liver, bone (frequently osteoblastic), and lung usually arise later in the disease. Local invasion into the trachea and esophagus also occurs.

Clinical Manifestations

Patients with MTC usually present with a neck mass that may be associated with palpable cervical lymphadenopathy (15 to 20 percent). Local pain is more common in patients with these tumors, and local invasion may produce symptoms of dysphagia, dyspnea, or dysphonia. The female-to- male ratio is 1.5:1. About 70 percent of the patients have sporadic disease, and age at presentation typically is between 50 and 60 years, though patients with familial disease present at a younger age.

Tumors may secrete a variety of peptides including calcitonin, calcitonin gene–related peptide, CEA, histaminadases, prostaglandins E₂ and F₂- alpha, and serotonin. Patients with extensive metastatic disease frequently develop diarrhea, which can be debilitating. The reason for this is unclear, although increased intestinal motility, impaired intestinal water and electrolyte flux might be responsible.

Between 2 and 4 percent of patients with MTC develop Cushing's syndrome as a result of ectopic production of ACTH. Kidney stones occur in patients with primary hyperparathyroidism, and hypertension occurs in those with pheochromocytoma.

Diagnosis

Diagnosis of MTC is established by history, physical examination, raised serum calcitonin or CEA levels, and FNAC of the thyroid mass. Attention to family history is important. Since it is not possible to distinguish sporadic from familial disease at initial presentation, all new patients with MTC should be screened for RET point mutations and for pheochromocytoma by measuring 24-h urinary levels of vanillylmandelic acid (VMA), catecholamine, and metanephrine. Serum calcium level also should be observed. A coexisting pheochromocytoma should be ruled out, because operation on a patient with an undiagnosed pheochromocytoma can result in a hypertensive crisis and death.

Familial Disease

MTC is sporadic (70 percent of new cases) or familial (30 percent of new cases). Familial MTC occurs as MEN IIA, as MEN IIB, or as familial MTC without other endocrinopathies. It also may occur in association with papillary thyroid cancer.

MEN IIA

These patients have a syndrome characterized by MTC, pheochromocytoma, or adrenal medullary hyperplasia and hyperparathyroidism. C-cell hyperplasia is present in all of them and normally is detectable by screening before the development of pheochromocytoma. Bilateral pheochromocytomas are detectable in more than 50 percent of the patients with the syndrome and occasionally is the presenting feature. Hyperparathyroidism (25 percent), Hirschsprung's disease, and cutaneous amyloidosis are present in some patients.

MEN IIB

MTC, bilateral pheochromocytomas, and ganglioneuromas affecting mucosal surfaces are found in patients with this condition. Patients have a characteristic facies with a thickened tongue and lips (Fig. 36-27). Marfanoid features, slipped epiphysis, and pectus excavatum also may occur. Patients with MEN IIB have the most aggressive medullary thyroid cancers; patients with familial MTC without other endocrinopathies have the least aggressive thyroid cancers.

Screening of patients with familial MTC for RET point mutations on chromosome 10 has replaced using provocation testing with pentagastrin or calcium-stimulated calcitonin levels to make the diagnosis. Calcitonin and CEA are used to identify patients with persistent or recurrent MTC. The specific phenotypes of MTC are also associated with specific RET mutations (familial MTC, 768 and 804; MEN IIA, 609, 611, 618, 620, and 634; MEN IIB, 918).

Treatment

Most clinicians agree that total thyroidectomy is the treatment of choice for patients with MTC because of the high incidence of multicentricity and the more aggressive course. Because tumors are of C-cell origin, radioiodine therapy and levothyroxine sodium TSH suppression therapy usually are not helpful. The central compartment nodes from carotid sheath to trachea frequently are involved early in the disease process and always should be meticulously cleared along with paratracheal nodes (central neck node dissection). In patients with palpable cervical nodes or involved central neck nodes, ipsilateral or bilateral, bilateral modified radical neck dissection is recommended. Similarly, patients with MTCs larger than 2 cm should undergo ipsilateral prophylactic modified radical neck dissection, because more than 60 percent of these patients have nodal metastases.

If superior mediastinal lymph nodes are noted at operation they also should be removed, although it is rarely necessary to perform a median sternotomy. Surgeons should be prepared to sacrifice the recurrent laryngeal nerve when it is involved in the tumor mass. Though this does not occur frequently, the possibility should be discussed with the patient preoperatively.

Tumor debulking in cases of metastatic disease or local recurrence should be undertaken. This frequently ameliorates symptoms of flushing and diarrhea, and it helps in decreasing the risk of death resulting from recurrent central neck disease.

External-beam radiotherapy for patients with tumors at resection margins or unresectable tumors is controversial. It is recommended for patients with unresectable residual or recurrent tumor, although the results are debatable. There is no effective chemotherapy regimen.

When patients have associated conditions such as pheochromocytoma or hyperparathyroidism, these conditions also require careful evaluation. Pheochromocytomas should be operated on first, before thyroidectomy is performed. In most cases pheochromocytomas can be removed laparoscopically. In patients who have hypercalcemia at the time of thyroidectomy, the parathyroid glands should be identified and, when abnormal, selectively removed. In patients with normocalcemia, efforts should be made to preserve the parathyroid glands, which should be marked with a stitch or clip in patients with MEN IIA. When a normal parathyroid cannot be maintained on a vascular pedicle, it should be removed, biopsied to confirm that it is a parathyroid, and then autotransplanted to the forearm of the nondominant arm.

Postoperative Follow-up and Prognosis

Patients should be assessed at regular postoperative intervals, and serum calcitonin and CEA levels should be monitored regularly. Calcitonin level is more sensitive for detecting persistent or recurrent disease, and CEA for predicting outcome. In a study of 123 patients with MTC, Russell and colleagues reported that 67 percent of the patients whose MTC was confined to the thyroid were clinically and biochemically (calcitonin) free of disease (mean follow-up time 5.5 years), compared to only 8 percent of patients who had extrathyroidal spread. A later paper from the same group reported on 31 patients who, despite having adequate primary operations for MTC, had persistently raised calcitonin levels. The 5- and 10-year survival rates for these patients were 90 percent and 86 percent, respectively, with only two patients dying from MTC. Postoperatively raised calcitonin

levels are frequently encountered and are a cause for concern signaling the need for evaluation.

When recurrent or metastatic MTC is suspected and suggested by rising calcitonin levels, localization studies for occult or clinically apparent disease should be used. The investigative tools available include CT, MRI of the neck and mediastinum, ultrasound examination, and selective venous catheterization (with calcitonin assay), including hepatic vein and jugular sampling after pentagastrin stimulation. An increase in the hepatic veins before an increase in the cervical veins indicates hepatic metastases. Nuclear imaging studies with ¹³¹I metaiodobenzylguanidine (MIBG), dimercaptosuccinic acid (DMSA), and ^{99m}Tc-sestamibi all have been used but are only occasionally helpful. Positron-emission tomography (PET) scanning appears to be useful in selected patients. Some clinicians examine the liver laparoscopically to rule out liver metastases.

Ultimately prognosis is related to disease stage. The 10-year survival rate is approximately 80 percent, and only 45 percent in those patients with lymph node involvement. Survival also is significantly influenced by disease type. It is best in patients with familial non-MEN medullary thyroid cancer, followed by those with MEN IIA, then those with sporadic disease, and it is worst in patients with MEN IIB. Patients with tumors that stain poorly for calcitonin and with a heterogeneous antibody uptake do worse than patients in whom calcitonin staining is increased and homogeneous.

With the identification and localization to chromosome 10 of the gene that causes MTC, it is now possible to screen family members and identify children at risk of developing MTC. Performing thyroidectomy at the C-cell- hyperplasia stage or before, after the age of 5 years, will prevent MTC and improve survival rates. Patients with MEN IIA and MEN IIB are at risk for pheochromocytoma.

Anaplastic Carcinoma

This tumor is one of the most aggressive malignancies, with few patients surviving 6 months beyond diagnosis. The incidence of anaplastic carcinoma has dropped to about 1 percent of thyroid carcinomas in the United States. The decrease in incidence is probably a result of the decrease in iodine-deficient (endemic) goiter, which was a precursor for this condition. Most anaplastic carcinomas arise from differentiated thyroid cancers. There is a female-to-male ratio of 1.5:1, and the majority of tumors present in the seventh and eighth decade of life. Presentation before the age of 50 years is rare.

Pathology

Tumor growth is extremely rapid, with macroscopic invasion of surrounding tissues. Lymph node involvement frequently is present. Microscopically, sheets of cells are seen, with marked heterogeneity. Cells may be spindle- shaped, polygonal, or large, multinucleated cells. Foci of more differentiated cells may be seen, either follicular or papillary in pattern. Anaplastic tumors originally described as small-cell are recognized to be lymphomas or less commonly medullary carcinomas of the thyroid gland and should be treated as such.

Clinical Manifestations

Patients typically are elderly females with a history of a lump in the neck that has been present for some time before rapidly enlarging and becoming painful. Associated symptoms such as dysphonia, dysphagia, and dyspnea are common. The tumor is hard and may be fixed to surrounding structures or may be ulcerated (Fig. 36-28). Lymph nodes usually are palpable at presentation. Evidence of metastatic spread also may be present. Diagnosis is confirmed by FNAC revealing characteristic giant and multinucleated cells. Incisional biopsy to confirm the diagnosis rarely is necessary.

Treatment

All forms of treatment have been disappointing. Tennvall and colleagues reported on 33 patients treated prospectively over an 8-year period with initial hyperfractionated radiotherapy with doxorubicin, then debulking thyroidectomy, followed by completion of radiotherapy and chemotherapy. Using this regimen, only four patients (12 percent) survived more than 2 years, and few patients died of suffocation because of central neck disease. Although the overall survival rate remains very poor, this treatment is recommended until better treatment results are documented.

Lymphoma

Approximately 1 percent of thyroid malignancies are lymphomas, and most of the primary thyroid lesions are of the non-Hodgkin's B-cell type. Although the disease can arise as part of a generalized lymphomatous condition, most thyroid lymphomas develop in patients with chronic lymphocytic thyroiditis (Hashimoto's thyroiditis). Chronic antigenic lymphocyte stimulation has been suggested to result in lymphocyte transformation.

Patients usually present with symptoms similar to those of patients with anaplastic carcinoma, although the rapidly enlarging neck mass often is painless. Because there is frequently coexisting Hashimoto's thyroiditis, patients may be clinically hypothyroid or may already be receiving thyroxine treatment. Most patients presenting with thyroid lymphoma are elderly women. The diagnosis usually is suggested by FNAC, although needle-core or open biopsy may be necessary for definitive diagnosis. When nodal dissection is being performed, taking a biopsy sample of lymph node often helps to clarify the diagnosis.

Patients with thyroid lymphoma usually respond rapidly to chemotherapy (CHOP—cyclophosphamide, doxorubicin, vincristine, and prednisone), and combined treatment with radiotherapy and chemotherapy is often recommended. Thyroidectomy and nodal resection are used to alleviate symptoms of airway obstruction in patients who do not respond quickly to chemotherapy or chemotherapy and radiotherapy.

Prognosis depends on the histologic grade of the tumor and whether the lymphoma is confined to the thyroid gland or is disseminated. The overall 5-year survival rate is about 50 percent; patients with extrathyroidal disease have markedly lower survival rate compared to those with intrathyroidal disease—40 percent versus 85 percent. Although there are no prospective studies, similar remission rates have been reported for patients who underwent diagnostic biopsy plus adjuvant therapy alone (85 percent), compared to debulking surgery plus adjuvant therapy. These findings support the argument for chemotherapy and radiotherapy for most patients.

Metastatic Carcinoma

The thyroid gland is a rare site of metastases from other cancers, including kidney, breast, lung, and melanoma; the most common metastatic tumor to the thyroid is hypernephroma. Approximately 3 percent of bronchogenic carcinomas metastasize to the thyroid, and these metastases account for 20 percent of secondary metastases to the thyroid.

Clinical examination and the history often suggest the source of the metastatic disease, and FNAC usually provides definitive diagnosis. Resection of the thyroid, usually lobectomy, may alleviate symptoms in symptomatic cases.

SURGERY OF THE THYROID

Patients who have had any change in voice or any previous neck surgery should have a preoperative vocal cord assessment, by direct or indirect laryngoscopy. Patients should be euthyroid at operation. On the operating table, a sandbag should be placed between the patient's shoulder blades, the head supported by a ring, and the neck extended to provide optimal exposure.

Operative Technique

A Kocher transverse collar incision is sited in a skin crease about 1 cm below the cricoid cartilage and carried through the platysma muscle (Fig. 36-29). The length of the incision varies, typically around 5 cm, and it should be symmetrical. Longer incisions are necessary in patients with larger tumors, in patients with a short, fat neck or whose neck cannot be extended, and in patients with low-lying thyroid glands. The upper flap is dissected in an avascular subplatysmal plane anterior to the anterior jugular veins and deep to the platysma muscle, to the level of the thyroid cartilage. The lower flap is then mobilized in a similar manner, to the suprasternal notch. Skin towels and a self-retaining retractor are applied. Meticulous hemostasis in thyroidectomy facilitates the identification of the parathyroid glands and the recurrent laryngeal nerves, reducing the possibility of inadvertent damage.

The thyroid gland is exposed via a midline incision through the superficial layer of the deep cervical fascia between the sternohyoid and sternothyroid muscles (Fig. 36-30). Care should be taken not to disrupt the veins on the thyroid surface. On the side to be dissected first, the more superficial sternohyoid muscle is separated from the deeper sternothyroid muscle by blunt dissection. The dissection proceeds laterally until the ansa cervicalis is noted on the lateral edge of the sternohyoid muscle and the medial aspect of the internal jugular vein. The deeper sternothyroid muscle is dissected free from the thyroid gland by blunt or sharp dissection, exposing the middle thyroid vein or veins laterally. The thyroid lobe is retracted anteromedially by finger traction and the lateral tissues swept posterolaterally by a pledget. The middle thyroid veins are divided and ligated. In rare cases the strap muscles must be divided in order to gain access to a large thyroid tumor. When this maneuver is necessary, the muscles should be divided high so that innervation from the ansa hypoglossal nerve is preserved. Occasionally a thyroid cancer invades the strap muscles; in this situation the involved muscles should be removed en bloc with the thyroid.

The superior thyroid pole is identified by retracting the thyroid inferiorly and medially. The tissues lateral to the upper pole of the thyroid and medial to the carotid sheath can be mobilized quickly by blunt dissection, because there are no nerves in that region. The upper pole of the thyroid is then mobilized caudally and laterally. The dissection plane should be close to the thyroid to avoid possible injury to the external branch of the superior laryngeal nerve. An effort should be made to identify and preserve the branch of the external laryngeal nerve that supplies the cricothyroid muscle. The branch is visible in about 80 percent of patients on the cricothyroid muscle (Fig. 36-31). The superior pole vessels are then individually identified, skeletonized, and doubly ligated low on the thyroid gland.

When the superior pole vessels have been ligated and divided, the tissues posterolateral to the superior pole can be swept away from the thyroid gland in a posteromedial direction. The direction is important because it reduces the chances of damaging the vessels supplying the upper parathyroid glands. The recurrent laryngeal nerves enter the larynx at the level of the cricoid cartilage, passing under or through Berry's ligament and entering the larynx deep to the cricothyroid muscle (Fig. 36-32).

At this point it should be possible to identify the upper parathyroid gland. The upper gland usually is situated at the level of the cricoid cartilage. The lower parathyroid gland is found most commonly just below where the inferior thyroid artery and recurrent laryngeal nerve cross. In approximately 80 percent of patients the glands are located within 1 cm of the artery. The lower parathyroid gland usually is situated anterior to the recurrent laryngeal nerve, and when it is not, it is usually located in the thymus or in parathymic fat (Fig. 36-33).

The thyroid is completely mobilized by gently sweeping dorsally all tissue along the posterolateral border away from the thyroid. All vessels are ligated and divided on the thyroid capsule to minimize the possibility of devascularizing the parathyroids. No tissue should be clamped or divided that might be or contain a recurrent laryngeal nerve.

Parathyroids that are ischemic, or are situated very anteriorly on the thyroid gland, or have been removed along with the thyroid lobe, should be examined after removal. They should be biopsied and confirmed by frozen- section examination, minced into 1-mm cubed pieces, and implanted into a small pocket in the ipsilateral sternohyoid muscle, with a

silk suture or clip marking the site.

The recurrent laryngeal nerves should be identified; the nerve is located more medially on the left and runs more obliquely on the right. The most difficult part of the operation usually is during the dissection where the recurrent nerve passes through Berry's ligament. It is here that the nerve is in close proximity to the thyroid, tethered down by the ligament, through which runs a small artery, and it is here that the nerve is most commonly injured. If bleeding occurs at this site it should be controlled by gentle pressure before identification of the nerve to avoid injury, and then the vessels are all ligated. Use of electrocautery to control bleeding should be strictly avoided.

A pyramidal lobe is present in about 80 percent of patients and should be dissected free superiorly to the level of the thyroid cartilage or higher and removed in continuity with the thyroid lobe and the isthmus. One or more lymph nodes are often present just cephalad to the isthmus of the thyroid gland (Delphian node) and should be removed with the thyroid.

When lobectomy is performed, the isthmus is divided flush with the contralateral gland and oversewn. When total thyroidectomy is performed, the same procedure is followed on the contralateral side. If a subtotal thyroidectomy is being performed (for Graves' disease or multinodular goiter), normal practice is to perform a total lobectomy on one side and a subtotal lobectomy on the other side. Removing all thyroid tissue on one side has the advantage of obviating the need to reoperate on one side in cases in which reoperation may be necessary, and the possibility of damage to the recurrent laryngeal nerve and parathyroid on that side during future procedures is reduced.

Subtotal lobectomy is achieved in a similar fashion to lobectomy, but after division of the superior thyroid pole, arterial clamps are placed along the length of the lobe, anterior to the insertion of the inferior thyroid artery, usually leaving approximately 1 to 2 cm ³ (4 to 8 g) of thyroid tissue. The thyroid is oversewn and secured to the lateral tracheal fascia to encourage hemostasis. Suction drains rarely are necessary. They can be used when there is a large space, e.g., after removing a large substernal goiter. Strap muscles are reapproximated with absorbable sutures, as is the platysma, and the skin is closed with subcuticular sutures or clips.

Surgical Removal of Intrathoracic Goiter

In 99 percent of the patients, intrathoracic goiter can be removed via a collar incision in the neck rather than resorting to a sternotomy. Patients with substernal goiters who have had previous thyroid operations, some with invasive malignant tumors, and patients with no thyroid tissue in the neck may require a median sternotomy.

Vascular connections in the neck are identified, ligated, and divided. The superior pole vessels and the middle thyroid veins are identified and ligated first. Dividing the isthmus may help with subsequent mobilization of the substernal goiter from beneath the sternum.

Several large 1-0 or 00 sutures are placed deeply into the goiter, and through a combination of traction and blunt dissection the thyroid usually can be delivered into the neck from its substernal location. In patients with a very large colloid goiter, the colloid occasionally may be evacuated to permit delivery, but this maneuver usually is unnecessary, and it is contraindicated if cancer is present. Although 1 percent of goiters contain cancer, the frequency of cancer in substernal goiter is as high as 10 percent in some series.

In approximately 1 percent of patients, part or all of the thyroid gland is intrathoracic. This usually is a consequence of extension or migration of the thyroid into the mediastinum rather than aberrant glandular tissue. The blood supply usually originates in the neck rather than the mediastinum. These goiters are more difficult to remove because it is difficult to apply traction.

Substernal goiters occasionally are too large to deliver via a neck incision, and a sternotomy is required. In patients in whom thyroid cancer is suspected or demonstrated in an intrathoracic gland, attempts should be made to avoid rupture of the thyroid capsule.

Because the blood supply to the thyroid gland, thymus gland, and lower parathyroids primarily is derived from the inferior thyroid arteries in the neck, median sternotomy usually is unnecessary. Sternotomy is indicated in patients in whom the intrathoracic gland is cancerous and mediastinal node involvement or metastatic cancer deep in the mediastinum has been demonstrated by scanning. When sternotomy is indicated, the sternum usually should be divided to the level of the third intercostal space and then laterally on one side at the space between the third and fourth ribs (Fig. [36-34](#)). Median sternotomy provides excellent exposure of the upper mediastinum and lower neck.

Neck Dissection for Nodal Metastases

Lymph nodes in the central compartment (medial to the carotid sheath) are frequently involved in metastatic spread in patients with papillary, medullary, and Hürthle cell carcinomas. In these patients all central neck nodes should be removed at the time of the primary procedure, preserving the recurrent laryngeal nerves and, when possible, the parathyroid glands, autotransplanting them when necessary, after histologic confirmation, into the sternocleidomastoid muscle. Removal of all central nodes is important in patients with medullary and Hürthle cell carcinoma because of the high frequency of microscopic tumor spread and because these tumors do not take up radioiodine and cannot be ablated with ¹³¹I.

In patients with thyroid cancer with palpable cervical lymph nodes or in cases of medullary carcinoma when the thyroid lesion is larger than 2 cm, a modified radical neck dissection usually should be performed. This operation is possible via extension of the collar incision laterally to the anterior margin of the trapezius muscle (MacFee incision). The internal jugular vein, the spinal accessory nerve, and the sternocleidomastoid muscle are preserved unless they are invaded or adhered to tumor. During dissection all tissue in the anterior triangle of the neck from the hyoid bone to the clavicle is removed. Dissection along the spinal accessory nerve is most important because this is a frequent site of metastatic disease. The deep dissection plane is the anterior scalenus muscle, the brachial plexus, and the medial scalenus muscle. The phrenic nerve is preserved on the scalenus anterior muscle, as are the cervical sensory nerves in most patients.

Complications of Thyroid Surgery

The mortality rate from thyroidectomy is low. In two large series, Gould and colleagues reported no mortality in more than 1000 consecutive patients. Similarly, Colcock reported no deaths in his personal series after 1954. Morbidity is approximately 13 percent. Serious morbidity occurs in fewer than 2 percent of patients. Complications specifically attributable to thyroidectomy include nerve damage or respiratory problems, bleeding with hematoma formation, hypoparathyroidism, wound infection, keloid formation, and, rarely, thyroid storm.

Injury of the Recurrent Laryngeal Nerve

Permanent recurrent laryngeal nerve injury is a relatively uncommon event, with a rate of approximately 1 percent; the rate is probably lower in patients with benign disease and higher in patients with large, invasive, or recurrent thyroid carcinomas. Injuries may be temporary, caused by traction or stretching of the nerve, or permanent, caused by division or ligation of the nerve. Injuries may be unilateral or bilateral. Recurrent laryngeal nerve injury leads to vocal cord paralysis, with the vocal cord assuming a paramedian position. In cases of bilateral nerve damage, the cords may obstruct the airway and lead to acute respiratory distress immediately after extubation. Reintubation and, occasionally, subsequent tracheostomy are needed. Unilateral nerve damage leads to paralysis of the ipsilateral vocal cord, with the cord lying in a paramedian position, usually resulting in a hoarse or husky voice. Because the nerve also is sensory to the larynx and because the vocal cords do not approximate, patients also describe choking and coughing when drinking fluids.

When damage is temporary, vocal cord function usually returns within 6 months, but it may take 1 year. When function has not returned by 6 to 12 months, injection lateral to the cord with Teflon mobilizes the cord to the midline and improves the voice. Occasionally the contralateral cord compensates and vocal cord paralysis may be unnoticed, so all patients should have direct or indirect laryngoscopy to evaluate vocal cord function before reoperation. Speech therapy also helps some patients with vocal cord dysfunction and should be done before Teflon injection.

Identification of the recurrent laryngeal nerves during surgery has been shown to decrease the incidence of permanent damage, although transient paralysis is more common. When nerves are not identified, transient paralysis is reduced, but the incidence of permanent damage is three or four times higher. Disease-specific risk factors for permanent nerve damage, in order of frequency, include recurrent thyroid cancer or recurrent goiter, thyroid cancer, large substernal goiters, chronic lymphocytic thyroiditis, Graves' disease, and euthyroid nodular goiter.

A rare cause of injury to a laryngeal nerve is that it is a nonrecurrent laryngeal nerve. This anomaly occurs in about 0.5 percent of patients and is almost always encountered on the right side. Awareness of this possibility reduces the risk of nerve injury when the recurrent nerve cannot be identified at the usual site.

Meticulous hemostasis, precise dissection between the thyroid capsule and sheath, and care dissecting Berry's ligament, where injuries are most likely to occur, all help to reduce the possibility of damage. The right recurrent laryngeal nerve takes a more oblique course in the neck as it loops around the subclavian artery, whereas the left recurrent laryngeal nerve loops around the ligamentum arteriosus and assumes a more midline position in the tracheoesophageal groove. Either nerve may branch before it enters the larynx posterior to the cricothyroid muscle at the level of the cricoid cartilage.

Injury of the External Branch of the Superior Laryngeal Nerve

Injury to the external laryngeal nerve results in difficulty shouting or singing high notes (the nerve is also called the "high note nerve"). The risk of injuring the nerve can be greatly reduced by retracting the strap muscles laterally to provide adequate visualization of the superior thyroid pole. A plane is opened by blunt dissection between the thyroid pole and the cricothyroid muscle bed. In about 80 percent of patients the nerve can be seen on the cricothyroid muscle. The superior thyroid vessels are individually ligated and divided low on the thyroid gland, rather than taken all together in one large bloc to avoid injury to the external laryngeal nerve. If these steps are followed, injury to the nerve is uncommon (2 percent).

Hypoparathyroidism

The chances of permanent hypoparathyroidism after thyroidectomy vary with the size and degree of invasion of the tumor, the type of pathology, the extent of the procedure, and the experience of the surgeon. Total thyroidectomy and central neck compartment clearance in medullary thyroid carcinoma has a higher incidence of subsequent hypoparathyroidism than does subtotal thyroidectomy for multinodular goiter (2 percent versus 0.5 percent), because all of the central neck fibroadipose and lymphatic tissue should be removed in patients with medullary carcinoma. After most operations the serum calcium level falls by about 1 mg/dL. Symptomatic postoperative hypoparathyroidism after thyroidectomy usually is transient and resolves in most cases within a few days without treatment and with calcium supplementation when treatment is necessary. When the serum phosphorus level is low or normal, there is less concern in the hypocalcemic patient than when it is high, because the latter suggests hypoparathyroidism.

In most instances, postoperative hypoparathyroidism is a result of parathyroid ischemia from bruising and partial interruption of parathyroid blood supply. This situation can be avoided by dissection along the thyroid capsule and gently teasing the parathyroid gland on a broad plane of tissue away from the thyroid gland in a posterolateral direction. This decreases the risk of disruption of the parathyroid blood supply, derived from the inferior and superior thyroid arteries. Hypoparathyroid patients usually exhibit early tingling and numbness around the lips, followed by the same sensation in the fingers. A positive Chvostek's sign (twitching of the lips after tapping over the facial nerve) is usually present. When hypocalcemia is not treated, patients may progress to carpopedal spasm. Symptoms occur when the calcium level falls below 8 mg/dL. Hypocalcemia also increases anxiety

and respiratory alkalosis. Hyperventilation can cause tetany with or without associated hypocalcemia. Patients with postoperative hypocalcemia are treated initially with approximately 1 g of calcium every 4 h if symptomatic. When the serum calcium level remains low, intravenous calcium (1 to 10 ampoules of calcium gluconate or calcium chloride) can be given over several hours. Extravasation into the subcutaneous tissues can cause tissue necrosis. Treatment with calcitriol (Rocaltrol) 0.25 to 1.0 mg twice daily is occasionally necessary. In patients with persistent symptoms, the serum magnesium level should be evaluated.

Postoperative Management of Differentiated Thyroid Cancer

Thyroid Hormone

After thyroid surgery for carcinoma, patients should be placed on thyroxine. This is necessary as replacement therapy in patients who have undergone total thyroidectomy but has the additional effect of suppressing TSH and reducing the growth stimulus to any possible residual thyroid cancer cells. TSH suppression reduces tumor recurrence rates, particularly in patients with papillary cancer. Thyroxine should be administered to ensure that the patient remains euthyroid, with circulating TSH levels about 0.1 mU/L in low-risk patients or less than 0.1 mU/mL in high-risk patients.

Thyroglobulin Measurement

Thyroglobulin levels in patients who have undergone total thyroidectomy should be below 2 ng/mL when the patient is taking thyroxine, and below 3ng/ml when the patient is not taking thyroxine. A thyroglobulin level above 3 ng/mL is highly suggestive of metastatic disease or persistent normal thyroid tissue, especially if it increases when TSH levels increase when thyroid hormone treatment is discontinued in preparation for radioiodine scanning. In this situation, radioiodine scan should be performed.

About 95 percent of patients with persistent or recurrent thyroid cancer of follicular cell origin will have thyroglobulin levels higher than 3 ng/mL. High- risk patients should also have a CT or MRI scan of the neck and mediastinum for early detection of any persistent or recurrent disease.

Radioiodine Therapy

Metastatic differentiated thyroid carcinoma can be detected and treated by radioactive iodine in about 75 percent of patients. Screening and treatment are facilitated by the removal of all normal thyroid tissue, which effectively competes for iodine uptake and is the most compelling of arguments in favor of total thyroidectomy for differentiated thyroid carcinoma. Radioiodine is more effective in young patients, even with occult pulmonary metastases, and less effective in older patients with less well differentiated thyroid cancers.

Most follicular carcinomas concentrate iodine. One exception is the Hürthle cell carcinomas, of which only about 10 percent take up iodine. Screening with radioactive iodine is a more sensitive test of metastatic disease than chest x-ray or CT but less sensitive than serum thyroglobulin determination. Micrometastases in the chest are detectable by radioiodine scanning even when chest x-rays show no evidence of disease. Approximately 75 percent of these patients have been successfully treated after receiving ablative doses of ¹³¹I for micrometastases, especially when total thyroidectomy has been performed. The success rate of treating macro–pulmonary metastases with ¹³¹I is less than 10 percent. TSH suppression with thyroxine after treatment with ¹³¹I should be used.

Radioactive iodine scanning and treatment follow a standard protocol after initial or completion total thyroidectomy. Patients should have their levothyroxine therapy discontinued for approximately 8 weeks before the scanning with ¹³¹I. During the first 6 weeks of this time they are given a synthetic triiodothyronine (T₃); this decreases the period (and discomfort) of hypothyroidism, because T₃ has a half-life of about 1 day, whereas T₄ has a half-life of about 1 week. The T₃ is then discontinued for remaining 2 weeks, which allows TSH levels to rise. A low-iodine diet also is recommended during these 2 weeks immediately before scanning.

In most centers, a screening dose of about 2 mCi of ¹³¹I is administered and the uptake is measured at 24 h. The uptake in most patients should be less than 1 percent, with no “hot” spots in the neck or elsewhere (remnant normal thyroid or metastases). The most frequent cause of a hot spot in the neck after initial screening is residual thyroid tissue in the thyroid bed. If there is significant uptake (>1 percent), then a therapeutic dose of ¹³¹I, 30 to 50 mCi in low-risk patients and 100 to 200 mCi in high-risk patients, is recommended. Some physicians omit the scanning dose of ¹³¹I for patients who are thyroglobulin positive, especially if the thyroglobulin level increases when the patient is hypothyroid in preparation for scanning or treatment. For these patients a scan should be performed 5 to 7 days after the treatment dose; about one-third will become thyroglobulin and radioiodine-uptake negative, which documents a therapeutic benefit.

After scanning and treatment, patients are placed again on levothyroxine (normal dose is about 125 mg/day) and observed over follow-up with serum thyroglobulin determinations and physical examination at regular intervals. Patients with previously positive scans and patients with serum thyroglobulin levels over 3 ng/mL usually need another ¹³¹I treatment after 6 to 12 months. The maximum dose of radioiodine that can be administered at one time is approximately 200 mCi with a cumulative dose of 1000 mCi.

External Beam Radiotherapy and Chemotherapy

External beam radiotherapy is required occasionally to control unresectable locally invasive or recurrent disease. It also is of value for the treatment and control of pain from bony metastases when there is no appreciable radioiodine uptake.

Multidrug chemotherapy and adriamycin have been used with little success in disseminated thyroid cancer; remission sometimes occurs, but cure is rare. Taxol has been reported recently to be of some value in patients with incurable disease.

Parathyroid

Jon A. van Heerden and David R. Farley

Historical Background Anatomy Pathology Physiology of Calcium Homeostasis Primary Hyperparathyroidism Persistent and Recurrent Hyperparathyroidism Secondary Hyperparathyroidism Tertiary Hyperparathyroidism Multiple Endocrine Neoplasia and Hyperparathyroidism Hypercalcemic Crisis Parathyroid Carcinoma

HISTORICAL BACKGROUND

In 1849 Sir Richard Owen, curator of the London Zoological Gardens, performed an autopsy on a rhinoceros that had been given to the London Zoo by the government of India (Fig. [36-35](#)). Owen's original description of the mysterious structures he found within the neck of the rhinoceros remains an accurate assessment of the normal parathyroid gland—“a small, yellow, glandular body ... attached to the thyroid at the point where the veins emerge.” In 1879 Anton Wölfler documented the relationship between total thyroidectomy and the ensuing tetany in the first such patient operated on by C.A. Theodor Billroth. The first description of the parathyroid glands in human beings was that by Ivar Sandström, a medical student in Uppsala, Sweden, in 1880 (Fig. [36-36](#)). He suggested that these glands be named the *glandulae parathyroideae*. The function of these structures was unknown at that time.

Friedrich von Recklinghausen, in 1891, described the fibrocystic disease of bone produced by hyperparathyroidism. The first association of hyperparathyroidism with osteitis fibrosa cystica (von Recklinghausen's disease) was made by Askanazy in 1903. In the same year, Erdheim recorded the coexistence of a parathyroid tumor and a pituitary tumor, foreshadowing the description of multiple endocrine neoplasia almost 50 years later. In 1909 serum calcium level determination became possible, and the association between serum calcium level and the parathyroid glands was established.

The first patient to be operated on for a parathyroid abnormality was a 38- year-old male who underwent a successful parathyroidectomy in 1925 by Felix Mandl. The patient had severe bone pain secondary to advanced osteitis fibrosa cystica. His condition dramatically improved after the operation, and he lived for another 7 years before dying of recurrent hyperparathyroidism or renal failure. In 1926, at the Massachusetts General Hospital, Edward Churchill, assisted by an intern named Oliver Cope, operated for the seventh time on the famous sea captain Charles Martell (Fig. [36-37](#)) for severe primary hyperparathyroidism. No parathyroid abnormalities were identified in any of Captain Martell's previous six cervical explorations. It was only at the time of the seventh operation that an ectopic adenoma was found in the substernal position. Captain Martell died 6 weeks after the successful seventh operation, most likely from laryngeal spasm after a subsequent surgical procedure to relieve ureteral obstruction secondary to stones.

The first successful parathyroidectomy for hyperparathyroidism based on an accurate preoperative biochemical diagnosis was performed in 1928 by Isaac Y. Olch at the Barnes Hospital in St. Louis, Missouri, on a 56-year-old female. At operation, a 3 × 3 cm adenoma was found attached to the left lower lobe of the thyroid gland. Postoperatively the patient developed life- threatening tetany, with her serum calcium dropping as low as 4.5 mg/dL. The patient recovered and lived for many years, although she required lifelong supplemental oral calcium lactate.

ANATOMY

The superior parathyroid glands arise from the fourth branchial pouch in conjunction with the ultimobranchial bodies. Because of this association, the superior parathyroid glands remain in close proximity to the posterior portion of the upper thyroid lobes. This position is fairly constant, and superior parathyroid glands located in ectopic sites are less common in comparison to the inferior glands. When the superior glands enlarge, they tend to “descend by gravity” into or along the tracheoesophageal groove and may be inferior to the inferior parathyroid glands (Fig. [36-38](#)). When true embryologic descent of the superior glands does occur, the descent is in a posterior plane (Fig. [36-39](#)). When such glands descend, they can be located in the posterior or middle mediastinum, with the aortopulmonary window being the most common site. The superior parathyroid glands usually are found in close association with the posterior aspect of the middle and upper thirds of the thyroid gland (Fig. [36-40](#)). In this position, they are in close proximity to the cricothyroid membrane entrance of the recurrent laryngeal nerve and usually cephalad to the superior thyroid artery.

The inferior parathyroid glands arise from the third branchial pouches in conjunction with the thymus. Given this important association, and since the embryologic journey is a longer one, ectopic sites for the inferior glands are more common and more widely distributed than ectopic sites for superior glands (Fig. [36-41](#)). With rare exceptions, thymic tissue is intimately related to inferior glands in ectopic locations. Ectopic sites range from an intrathymic parathyroid gland located in the anterior-superior mediastinum to an undescended inferior parathyroid gland (parathymus) located superior to the superior parathyroid gland. Undescended inferior glands are most often found at the angle of the mandible anterior to the carotid artery but may be seen as high as the base of the skull.

The normal inferior glands are more variable in location than superior glands (see Fig. [36-40 B](#)). The most common site (approximately 60 percent) for inferior parathyroids is within a circle 2 cm in diameter centered on a point that is on the posterolateral aspect of the lower pole of the thyroid gland. The majority of the remaining normal inferior parathyroid glands will be found in the thyrothymic tongue of tissue (thyrothymic ligament), which extends inferiorly from the lower pole of each thyroid lobe to the thymus gland, usually located in the anterior mediastinum (Fig. [36-42](#)).

A normal parathyroid gland usually weighs less than 50 mg, measures approximately 3 × 3 × 3 mm, and may be difficult to differentiate from adjacent fatty tissue. Parathyroid glands are more yellow-brown than normal fatty tissue, and close observation will invariably reveal small “grains of salt” on their surface, an observation that aids in differentiation

from surrounding fatty tissues. Normal glands have more clearly defined margins, which aids in this differentiation and which has led to their being described as looking like “the tongue of a jaundiced hummingbird.”

PATHOLOGY

Primary hyperparathyroidism may be due to parathyroid adenoma(s), hyperplasia, or, rarely, carcinoma. Of these causes, a parathyroid adenoma is by far the most common, accounting for 90 percent of cases. In almost 2 percent of patients with adenomatous disease, double adenomas are present. Parathyroid hyperplasia (multiglandular disease) may be found in nearly 8 percent of patients with sporadic hyperparathyroidism. The incidence of hyperplasia increases markedly and may be universal in patients with hyperparathyroidism that occurs in families with multiple endocrine neoplasia (MEN) type I or II, and in patients with non-MEN familial hyperparathyroidism. Parathyroid hyperplasia is seldom equal or even symmetrical. This “unequal hyperplasia,” if misinterpreted by the surgeon, may lead to a mistaken diagnosis of single versus multiglandular disease. Parathyroid carcinoma, as a cause for hyperparathyroidism, is exceedingly rare, occurring in less than 1 percent of patients with this disease.

Differentiation between the causes of hyperparathyroidism is of paramount importance, because the selection of appropriate surgical treatment depends on it. This delineation is made primarily by the operating surgeon identifying the pathology encountered at the time of surgical exploration. A structure that looks like a little kidney in the neck or mediastinum is virtually pathognomonic for single-gland disease. Finding multiple enlarged glands somewhat “paler than a normal kidney” suggests multiglandular disease. The pathologist can be of great help to the surgeon, and vice versa. Unequivocal histologic identification of the causes of primary hyperparathyroidism can be difficult without the surgeon's description of the gross characteristics of all glands. With small biopsy specimens alone of parathyroid glands, it can be extremely difficult for the pathologist to accurately differentiate histologically between a truly normal gland, a parathyroid adenoma, or a hyperplastic parathyroid gland. Our own preference is to avoid taking biopsy specimens of normal-appearing glands and to rely on visual inspection of the remaining glands and the histology of the excised abnormality to determine the diagnosis.

Normal parathyroid glands contain primarily chief cells, with occasional oxyphil cells. An adenoma is typically defined as an enlarged parathyroid gland that is made up of solid sheets of chief cells, oxyphil cells, or varying combinations of both with a rim of compressed normal parathyroid tissue (Fig. 36-43). This compressed rim of normal tissue is present in only 20 to 30 percent of parathyroid adenomas. The histology of parathyroid hyperplasia may be difficult to discern unless the nature of the other parathyroid glands is placed in context (Fig. 36-44). Histologically, the presence or absence of parathyroid fat cells has been believed by some authorities to be of help in the diagnosis of parathyroid hyperplasia; the higher the percentage of fat cells, as demonstrated by fat staining, the lower the chances of hyperplasia, and vice versa.

The macroscopic appearance of the parathyroid glands, as visualized by an experienced parathyroid surgeon, continues to be the most accurate means of identifying the various types of parathyroid pathology.

PHYSIOLOGY OF CALCIUM HOMEOSTASIS

Parathyroid Hormone (PTH)

In patients with primary and tertiary hyperparathyroidism, the secretion of parathyroid hormone is abnormal and excessive. In secondary hyperparathyroidism, PTH secretion is increased, but appropriately so. The effect of PTH in human beings is to increase serum calcium concentration (Fig. 36-45). Hypercalcemia occurs because the principal actions of PTH involve the following:

- (1) Increase of bone osteoclast and osteoblast activity.
- (2) Increase in rate of conversion of 25-hydroxycholecalciferol (25- hydroxyvitamin D 3) to 1,25-dihydroxycholecalciferol (1,25- dihydroxyvitamin D 3) in the kidneys by stimulation of 1 a-hydroxylase activity.
- (3) Increase of gastrointestinal absorption of calcium by enhancing vitamin D synthesis.
- (4) Increased excretion of bicarbonate by the kidney.
- (5) Decrease in serum phosphate level by increasing the excretion of urinary phosphate.

PTH is a single-chain polypeptide of 84 amino acids. The entire polypeptide is referred to as *intact PTH*, a hormone that has a half-life of 2 to 5 min. The *N*-terminal PTH portion (1–34), also called the amino (*N*) terminal, is biologically active, with a half-life of 2 min. The *C*-terminal PTH portion (35–84) or carboxyl (*C*) terminal, is biologically inactive, with a half-life of approximately 30 min. Among other effects, PTH acts to promote bone resorption and to increase calcium absorption from filtered urine in the renal tubule.

Calcium

Calcium is the principal regulator of parathyroid hormone release. This regulation is mediated by calcium receptors or other receptors on the parathyroid cell surface. A membrane protein that has been proposed to constitute such receptors (calcium ion–sensing receptor [CaR]) demonstrates reduced expression in hyperparathyroidism and seems causally related to the relative insensitivity of the secretion to external calcium in pathologic parathyroid cells. This protein consequently may be an important cause for the hypercalcemia of this disorder. Modulating the functions of such proteins comprises interesting means of future therapy in several disorders of mineral metabolism. Increased knowledge of these proteins and parathyroid pathophysiology may clarify the presence of relevant mutations, secretory derangements, and propensity for cell growth, from which accurate classification of parathyroid diseases and improved treatment strategies might evolve.

Calcium is the most abundant cation in human beings, and its distribution within cells is essential for virtually all physiologic functions. Calcium balance in a normal individual is tightly regulated. The body contains approximately 1000 g of calcium, with minimal variations from day to day in serum calcium concentration (Fig. 36-46). Total serum concentrations, as tested by routine laboratory analysis, are accurate as long as serum protein concentrations are normal. Total calcium level must always be considered in its relationship to plasma protein levels, especially serum albumin (for each gram per deciliter of alteration of serum albumin above or below 4.0 mg/dL, there is a 0.8 mg/dL increase or decrease in protein-bound calcium and thus in total serum calcium).

Vitamin D

Vitamin D₃ (cholecalciferol) is vital to calcium homeostasis. It is absorbed through the gastrointestinal tract and synthesized in the skin and is converted by the liver to 25-hydroxycholecalciferol, which is subsequently converted to its active form, 1,25-dihydroxycholecalciferol in the kidneys by the enzyme 1a-hydroxylase. Vitamin D₃ increases the absorption of calcium from the gastrointestinal tract and promotes phosphate retention; it elevates serum calcium and phosphate levels and in turn enhances the mineralization of bone.

A fall in serum calcium concentration stimulates secretion of PTH, which, among other effects, results in a rise in vitamin D₃ synthesis. Vitamin D₃ then acts peripherally to raise the serum calcium concentration toward normal. Vitamin D₃ promotes bone resorption and enhances absorption of calcium and phosphorus from the gut. An elevation of the serum calcium level reduces PTH secretion and the formation of vitamin D₃. Both effects tend to lower serum calcium concentration.

PRIMARY HYPERPARATHYROIDISM

Primary hyperparathyroidism is the hypercalcemic condition generated as a result of overproduction of PTH by one or more parathyroid glands. PTH enhances gastrointestinal absorption of calcium, stimulates the production of vitamin D₃, and inhibits renal calcium excretion. PTH secretion from the parathyroid glands is inversely related to serum calcium levels and is tightly regulated. In primary hyperparathyroidism, the normal negative feedback loop is altered, and inappropriately elevated PTH levels occur in the face of hypercalcemia.

The incidence of primary hyperparathyroidism in the United States is estimated to be 1:700, with a female-to-male ratio of nearly 3:1. Primary hyperparathyroidism is the most common cause of hypercalcemia observed. Only advanced malignancy is a more common cause of hypercalcemia in hospitalized patients in the United States. Of the nearly 100,000 cases diagnosed yearly in the United States, most occur in postmenopausal women, in whom the incidence is nearly 1:200. The incidence of this disease appears to have decreased over the past decade, for unknown reasons.

Etiology

While the exact cause or stimulus that brings about the autonomous production of PTH remains unclear, pathologically the disease is manifested by three separate entities. Most commonly, primary hyperparathyroidism is caused by the benign enlargement of one (90 percent) or, occasionally, two (2 percent) parathyroid glands, a condition referred to as benign parathyroid adenoma(s). Less common is the occurrence of multiple gland enlargement (8 percent), which is referred to as multiglandular hyperplasia. Parathyroid carcinoma is a rare (less than 1 percent), often lethal, cause of hyperparathyroidism.

Clinical Manifestations

Before calcium levels were routinely included in serum chemistry analysis in the late 1960s, and before their routine use in the 1970s as an adjunct for general medical examinations, the diagnosis of primary hyperparathyroidism was rarely considered, typically only in patients with renal stones, bony abnormalities, or severe mental changes. Patients with primary hyperparathyroidism had variable symptoms: renal stones (64 percent), bone disease (20 percent), peptic ulcer disease (12 percent), and hypertension (4 percent). Walter St. Goar, in 1957, first used the mnemonic “Hyperparathyroidism is a disease of stones, bones, and abdominal groans” to remind physicians to consider this diagnosis in patients with varied complaints. With time, the triad of St. Goar has been extended to include other nonspecific regional symptoms of hypercalcemia pertaining to the abdomen, kidney, bone, constitutional, and neurologic manifestations (Fig. 36-47).

Over half of all patients diagnosed with primary hyperparathyroidism are asymptomatic, and the disease is discovered by routine serum chemistry analysis showing hypercalcemia. After successful surgical management of primary hyperparathyroidism, nearly all patients realize that they had in fact been symptomatic.

Evaluation

The diagnosis of primary hyperparathyroidism is most commonly one of exclusion. It is most commonly considered when serum chemistry analysis reveals hypercalcemia. A focused history in symptomatic and asymptomatic patients is important to rule out medicinal use that might cause hypercalcemia and obscure the diagnosis—most commonly thiazide diuretics, but also lithium and excessive ingestion of vitamin A or vitamin D. Consumption of extraordinary amounts of milk or antacids might alert the clinician to the

diagnosis of the milk-alkali syndrome. Questions regarding renal failure, dehydration, immobilization, and Paget's disease can lead to identifying the cause of the hypercalcemia. Endocrine abnormalities such as adrenal insufficiency, and either hyperthyroidism or hypothyroidism are similarly rare causes of hypercalcemia. Granulomatous diseases such as sarcoidosis, tuberculosis, coccidioidomycosis, histoplasmosis, and berylliosis can elevate serum calcium levels. It is critical to evaluate hypercalcemic patients for previous or current malignancy with possible bony metastases. The malignant causes of hypercalcemia include renal cell carcinoma, multiple myeloma, and squamous or small cell lung cancer. Over 90 percent of patients with hypercalcemia have primary hyperparathyroidism or malignancy, most commonly metastatic from the prostate or breast, as the source of their elevated serum calcium levels.

A family history of multiple endocrine neoplasia or benign familial hypocalciuric hypercalcemia (BFHH) is critical. MEN and BFHH usually occur in younger patients with a known genetic predisposition. BFHH involves sufficient retention of calcium to cause hypercalcemia. Unlike patients with primary hyperparathyroidism, BFHH patients have unusually low urinary calcium excretion. Cervical exploration in these patients is of no benefit and is contraindicated.

Physical examination of patients with primary hyperparathyroidism is rarely of diagnostic aid, but it is important to assess for neck masses, voice abnormalities or hoarseness, cervical lymphadenopathy, and mobility of the cervical spine. Neck masses most commonly represent benign thyroid nodules but could be the rare parathyroid cancer. Hypercalcemia coupled with a firm cervical mass with hoarseness or a recent change in voice would indicate involvement of a recurrent laryngeal nerve secondary to this rare malignancy. As a corollary, parathyroid adenomas, regardless of their size, are rarely palpable. While cervical lymphadenopathy can be observed with parathyroid carcinoma, this finding would more likely represent a recent upper respiratory infection or a more common malignant process such as metastatic lung, thyroid, or oropharyngeal cancer. Assessing cervical mobility is important if the patient eventually becomes a surgical candidate with the need for neck extension to facilitate surgical exposure.

While band keratopathy is easily detectable in most patients undergoing ophthalmologic examination with hyperparathyroidism, keratopathy and other ocular changes are not specific; all patients with hypercalcemia show ocular changes. Most patients with primary hyperparathyroidism have no other signs or symptoms; the only likely physical findings of interest would be those to help exclude the diagnosis or to create the suspicion of malignancy as the source of hypercalcemia.

Formerly, a solitary test result showing hypercalcemia was repeated to verify the hypercalcemic state. Today, given the accuracy of serum calcium determination, a single analysis is cost effective and sufficient. After a thorough history and physical examination of an otherwise healthy patient with no risk factors for hypercalcemia, the diagnosis of primary hyperparathyroidism is virtually certain if hypercalcemia exists and the serum PTH level is inappropriately high. "Inappropriately high" does not necessarily mean outside the normal serum range (usually 1 to 5 pmol/L). Given a normally functioning feedback loop, serum PTH levels in an otherwise healthy patient with hypercalcemia should be virtually zero. Nearly 20 percent of patients with primary hyperparathyroidism have an inappropriately normal level of PTH in the presence of hypercalcemia. Serologic testing of patients with primary hyperparathyroidism usually generates consistent results: hypophosphatemia, hypercalcemia, and elevated PTH levels (Fig. [36-48](#)). In any patient in whom the diagnosis of primary hyperparathyroidism is not secure, additional testing is mandatory. Some examples include:

- (1) Intravenous pyelography or ultrasonography to rule out renal cell carcinoma (a rare cause of hypercalcemia resulting from the secretion of a parathyroid hormone related protein PTHrP)
- (2) Protein gel electrophoresis to rule out multiple myeloma and sarcoidosis
- (3) 24-h urine collection for calcium determination to eliminate BFHH
- (4) Chest x-ray to exclude sarcoid, fungal diseases, or a malignancy of the lungs or mediastinum
- (5) Mammography in postmenopausal women

Imaging

Radiologic studies are unnecessary unless previous neck operations have been undertaken. Some advocate the use of preoperative ultrasonography in the hopes of avoiding bilateral exploration and minimizing operative time and morbidity, but surgical exploration remains the most accurate localization modality. Use of ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI), venous sampling, and nuclear scintigraphy is not cost effective for initial cervical exploration in patients with primary hyperparathyroidism.

In patients who are undergoing reoperation, in whom dissection is more hazardous, ultrasonography of the neck is helpful to the operating surgeon. Similarly, technetium-99m-sestamibi (Cardiolite) scanning has great sensitivity in identifying missed parathyroid tissue in the neck as well as in the mediastinum. Venous sampling, CT, and MRI are helpful in identifying aberrantly located glands, but they are expensive and less accurate. Our preference when performing a reoperation is to use ultrasonography with fine-needle aspiration of any suspicious cervical masses. If ultrasonography is equivocal, we proceed with technetium-99m-sestamibi scanning.

Treatment

With success rates higher than 98 percent for initial operations, the indication for surgical intervention is the diagnosis of primary hyperparathyroidism. Given the negligible operative mortality and minimal morbidity associated with first-time parathyroidectomy, virtually all patients who have no prior neck operations should be offered exploration. Given the safety and success of primary cervical exploration, concomitant surgical procedures ranging from minor (breast biopsy, dilatation and curettage) to major (cardiopulmonary bypass and cholecystectomy) may be considered. With success rates of roughly 90 percent in reoperative cases, repeat exploration for recurrent or persistent hyperparathyroidism is generally indicated if the offending gland has been localized in imaging studies.

Solitary parathyroidectomy, without biopsy of other glands, is sufficient treatment for solitary adenomas. Instances of double adenomas require both offending glands to be removed. Therapy for four-gland or multiglandular hyperplasia is controversial; our preference is three and one-half gland parathyroidectomy. Total parathyroidectomy with reimplantation of approximately 50 mg in the neck or forearm is another viable option. The use of the rapid (15 min) intact immunoreactive parathyroid hormone (iPTH) assay allows intraoperative assurance of removal of the offending (or sufficient) tissue. The indications for such intraoperative serological testing are rare.

Postoperative laboratory analysis reveals normocalcemia within 24 to 72 h in nearly all cases. Transient hypocalcemia is common with removal of large adenomas, resections involving four-gland hyperplasia, and in patients who have had prolonged hyperparathyroidism. The cause of postoperative hypocalcemia is invariably bone hunger or hypoparathyroidism. Bone hunger and hypoparathyroidism are differentiated by analysis of the serum phosphate and PTH levels. Bone hunger is identified by low phosphate levels associated with transient hypocalcemia after parathyroidectomy. PTH levels are within the normal range. The aparathyroid state is diagnosed by elevated phosphate levels and abnormally low PTH levels.

Hospitalization of patients without complications usually is an overnight stay. Serum calcium levels should be checked to demonstrate the equilibration of calcium before discharge. Obtaining a follow-up serum calcium level is appropriate in 1 or 2 months and yearly thereafter to assess for persistent or recurrent hyperparathyroidism.

The goals of cervical exploration in patients with primary hyperparathyroidism are as follows:

- (1) *Identification of the pathology.* In practice, this most commonly consists of the identification of a solitary adenoma (90 percent). A less common occurrence is multiglandular disease (8 percent) involving diffuse hyperplasia, which may be unequal. Identification of dual adenomas (2 percent) or parathyroid carcinoma (1 percent) is rare.
- (2) *Removal of sufficient pathologic tissue to restore normocalcemia.* The ideal postoperative result in these patients is a normocalcemic state with no need for supplemental calcium or vitamin D therapy. Transient hypocalcemia is sometimes unavoidable if severe osteopenia is present preoperatively, with ensuing postoperative bone hunger. Postoperative hypocalcemia or hypercalcemia may be caused by errors in cervical exploration: not finding the offending pathology (most often in single-gland disease); finding the diseased tissue but excising insufficient amounts of it (multiglandular hyperplasia), with resultant hypercalcemia; overly aggressive exploration with biopsies and possible injury to normal glands in single-gland disease; or excising too much hyperplastic tissue in multiglandular disease, with resulting hypocalcemia.
- (3) *Minimal postoperative morbidity and mortality.* The operative mortality in most reported series is less than 1 percent, which attests to the safety of cervical exploration. Complications and morbidity involve wound infection, cervical hematoma, injury to the inferior or superior laryngeal nerves, postoperative hypocalcemia, and, perhaps most commonly, poor cosmesis. All such complications should occur in less than 1 percent of patients.

To achieve these goals, cervical exploration should be undertaken with the conviction that parathyroid pathology exists and requires identification and removal. Cervical exploration should not be undertaken to make or confirm the diagnosis of hyperparathyroidism. A gentle technique that delivers a bloodless operative field is imperative. Poor visualization may result in disruption or fracture of parathyroid tissue with subsequent functional regrowth, the so-called parathyromatosis. The surgeon who is unfamiliar with parathyroid embryology should not undertake cervical exploration for hyperparathyroidism. It is important to be unhurried when performing this procedure. The surgical adage "The slower one operates, the sooner the operation is over" applies to this operation.

Technique

Although cervical exploration for primary hyperparathyroidism can be performed under regional blockade supplemented by intravenous sedation and local anesthesia, general anesthesia with endotracheal intubation is preferable in most patients. After induction of general anesthesia, both arms are tucked at the patient's sides with careful attention to any interference with the previously placed intravenous lines and with particular attention to possible pressure points on the patient's elbows and wrists. The midline placement of a folded surgical towel beneath the patient's scapulae and a 5- to 10-degree downward tilt of the head of the operating table facilitates exposure of the lower portion of the neck. Cervical extension should be minimized in older patients to avoid undue postoperative neck pain. It is prudent to ask all patients to demonstrate the degree of extension they can achieve voluntarily without discomfort. Sandbags are placed alongside the patient's head to avoid any side-to-side movement during the operation (Fig. [36-49](#)).

It is important for cosmesis that the patient's chin, thyroid cartilage, and suprasternal notch be aligned vertically before an incision is made. Failure to establish this alignment may result in an asymmetric and unsightly incision. A thyroid surgical drape is placed with tie-tapes passing behind each earlobe and anchored at the head of the table with a single hemoclip (see Fig. [36-49](#)).

A collar incision is made approximately two fingerbreadths above the suprasternal notch, which usually means the incision is situated over the cricoid cartilage. The surgeon should

look carefully for obvious skin creases slightly above or below this site; if there are any, they should be used for improved cosmesis. The length of the incision is approximately 10 cm. The planned site should be marked in a linear fashion by applying pressure to the imaginary line with a suture. The course of the incision extending laterally above both clavicles should be carefully inspected and equalized, ensuring cosmesis. An incision that is higher on one side than the other is noticeable and troublesome to the patient. The table is positioned with the patient's head and chest elevated approximately 15 to 20 degrees to the horizontal.

The skin incision is made and, with use of electrocautery, the platysma muscle is identified and divided. This muscle is extremely thin, usually measuring no more than 1 to 2 mm in thickness. The avascular plane, just deep to the platysma, should be searched for. If the dissection at this stage is carried too deep, there will be unnecessary bleeding because of injury to the anterior jugular veins. By placing sharp rake retractors and tensing the platysma anteriorly, the avascular plane immediately beneath the muscle can be easily developed superiorly to the level of the thyroid cartilage. There should be no bleeding during this stage of the operation. Minimal dissection is required inferiorly, usually just enough to allow placement of a self- retaining retractor (Fig. 36-50).

The midline is readily identified by finding the thin, avascular fascial plane connecting right and left strap muscles. This plane is divided with electrocautery. This dissection is carried posteriorly until the thyroid isthmus is clearly identified. Occasionally, small veins cross this midline space; if these are encountered, they should be isolated, ligated, and divided. A meticulous and bloodless operative technique is mandatory. Even minimal bleeding will render identification of parathyroid glands, particularly normal glands, extremely difficult. Once the thyroid isthmus has been identified, traction is applied laterally to the strap muscles with retractors that work against digital medial retraction of the thyroid lobe.

Anterior and medial displacement of the thyroid lobe brings the middle thyroid vein into view. The vein is isolated, ligated, and divided. The carotid sheath should appear in the posterior aspect of the dissection. Clamps are placed on the inferior and superior aspects of the thyroid lobe, and the thyroid gland is elevated anteriorly, superiorly, and eventually medially. This elevation of the thyroid lobe is required for accurate identification of the parathyroid glands (Fig. 36-51).

The superior parathyroid glands usually are found in close association with the posterolateral aspect of the superior pole of the thyroid lobe. The superior thyroid pole seldom needs to be taken down to visualize the superior parathyroid gland as long as there is anterior and medial displacement of the thyroid lobe. The inferior parathyroid gland, in contrast, is intimately involved with the inferior aspect of the thyroid gland and the thyrothymic tongue of fat (thyrothymic ligament) that extends inferiorly toward the mediastinum. Many inferior parathyroid glands are subcapsular, i.e., attached to the inferior aspect of the thyroid lobe and covered by a thin layer of fibrous tissue. This does not represent an intrathyroid parathyroid gland, as is often erroneously assumed. To be truly intrathyroid, the parathyroid gland should be circumferentially surrounded by thyroid parenchyma.

If the superior parathyroid gland is not visualized, the tracheoesophageal groove should be explored digitally. Enlarging superior parathyroid adenomas often descend into this groove and can be palpated accurately by inserting a finger into the avascular space immediately superior to the inferior thyroid artery (Fig. 36-52). Such glands are often located inferior to the inferior parathyroid glands. The parathyroid glands can be found from as high as the angle of the mandible or the base of the skull (undescended parathymus) to as low as an intrathymic location in the anterior mediastinum or in the aortopulmonary window in the posterior mediastinum.

In the hands of an experienced surgeon, a negative exploration usually is the result of an unusual variation in embryologic descent. The following rule is useful in this situation: If the superior parathyroid gland is not found in its conventional position, the surgeon should search inferior to the inferior parathyroid gland for the superior parathyroid gland (tracheoesophageal groove); conversely, if the inferior parathyroid gland is not found inferior to the superior parathyroid gland, the surgeon should search superior to the superior parathyroid gland for the inferior parathyroid gland (undescended parathymus). If these simple maneuvers are unsuccessful, the surgeon should become suspicious that the offending parathyroid gland might be in one of the following locations:

- (1) In the thymus. A fairly complete transcervical thymectomy should be performed to search for the missing inferior gland.
- (2) Within the thyroid. Truly intrathyroid glands are exceedingly rare. Most so-called intrathyroid parathyroid glands are inferior parathyroid glands located beneath the thin veil of the thyroid capsule and can be easily teased off the thyroid.
- (3) In the carotid sheath. This is extraordinarily rare, but opening the carotid sheath should be considered when all other possibilities have been ruled out.
- (4) Lateral to the carotid sheath. This location is even more unusual than in the carotid sheath, but the parathyroid gland can hide laterally.

Routine bilateral exploration is recommended in all patients. If a single adenoma is identified, excision of this adenoma, without biopsy of normal glands, is suggested. If hyperplasia is documented, all but approximately 50 mg of clearly viable parathyroid tissue is excised and routine transcervical thymectomy is performed. With the approach outlined, surgical success should approach 99 percent. Permanent laryngeal nerve palsy or hypoparathyroidism occurs in less than 1 percent of patients.

Persistent and Recurrent Hyperparathyroidism

Almost twenty years ago, Wexler stated, "Persistent hypercalcemia in suspected hyperparathyroidism is due to an improperly performed primary operation." This statement is as true today as it was then. A faulty operation is the cause for the vast majority of patients with persistent or recurrent hyperparathyroidism.

Persistent hyperparathyroidism is defined as elevated serum calcium levels that do not return to normal after an operation for biochemically proved hyperparathyroidism; this situation should be differentiated from the less common situation of recurrent hyperparathyroidism in which at least 6 months of well-documented normocalcemia occurs after cervical exploration for hyperparathyroidism before hypercalcemia is documented again.

The results of reoperation for recurrent or persistent hyperparathyroidism do not parallel those achieved for the initial operation. Regardless of the surgical expertise involved, reoperative cure rates are uniformly 10 to 20 percent lower (i.e., 80 to 90 percent success rates) than those obtained in first-time operations (95 to 99 percent). In addition, postoperative morbidity involving recurrent laryngeal nerve palsies, postoperative hematomas, and permanent hypocalcemia are three to five times higher than the rates anticipated after first-time explorations (<1 percent).

The principal factors contributing to persistent or recurrent hyperparathyroidism are varied (Table 36-4). A systematic, logical approach to management is required (Fig. 36-53), with a concerted effort by all members of a multidisciplinary team.

The initial step is to reconfirm the diagnosis of hyperparathyroidism. In particular, the possibility of benign familial hypocalciuria hypocalcemia must be excluded, because cervical exploration is of no benefit to patients with BFHH. This diagnostic trap may be avoided by obtaining a 24-h urine collection for calcium determination. The clinician must be concerned with the hypercalcemic patient who has 24-h urine calcium levels of less than 100 mg. If diagnostic uncertainty persists, the urine calcium-to-creatinine ratio should be calculated. In BFHH patients, this ratio invariably is less than 0.01. The formula for calculating this ratio, the fractional excretion of calcium, is:

$$FE\ Ca_2 + = \text{Urinary } Ca_2 + (\text{mg/total volume}) \times \text{serum creatinine (mg/dL)} / \text{Serum } Ca_2 + (\text{mg/dL}) \times \text{urinary creatinine (mg/total volume)}$$

Although the indication for operation for primary hyperparathyroidism is the diagnosis itself, this is not the case in patients who are undergoing reoperation. Risk assessment in this situation is of great importance. Objective data that are helpful in evaluating the anesthetic risks can be obtained. The criteria most often used are those of the American Society of Anesthesiologists physical status classification (Table 36-5) and of the Goldman multivariate index of cardiac risk (Table 36-6).

While cervical exploration invariably is well-tolerated, if the operative risk is unacceptably high, consideration should be given to nonsurgical means of parathyroid tissue ablation. The two most commonly used modalities are angiographic embolization and ultrasound-guided alcohol ablation; the latter technique is somewhat safer in the high-risk patient and is gaining worldwide acceptance.

When repeat exploration is warranted, the surgeon should review the prior operative report. It is important to know what tissue was removed and from which site. Additional clues may be obtained regarding the thoroughness of the previous operation. Statements in the report may lead to finding pathologic tissue in more common locations rather than in unusual hiding places.

The pathologist should review previously excised tissue to ascertain whether it was in fact parathyroid tissue and not thyroid or lymphatic tissue and whether any histologic evidence is present to suggest single- gland versus multiglandular disease or parathyroid carcinoma.

The radiologist should be consulted regarding the most appropriate localizing modality for the patient under consideration. In contrast to primary operations for hyperparathyroidism, in which preoperative localization is not indicated, localization in the reoperative setting is crucial. A wide variety of localizing modalities are in use (Table 36-7). The most popular modalities are radionuclide scanning with technetium-99m- sestamibi and ultrasonography (Fig. 36-54). Sestamibi scanning reaches sensitivity rates of over 90 percent and is particularly useful for parathyroid tissue located in the mediastinum (Fig. 36-55). A powerful adjunct to ultrasonography is liberal use of ultrasound-guided fine-needle aspiration (Fig. 36-56), not only for cytologic review but also for the measurement of iPTH. A high level confirms that the image seen is parathyroid tissue.

If all localizing modalities are negative, blind exploration should not be performed. Clinical and biochemical monitoring with interval reevaluation is the more prudent course to follow.

Although a repeat cervical or mediastinal exploration is considerably more difficult than a primary exploration, accurate preoperative localization enables the surgeon to perform a more focused exploration (e.g., unilateral cervical exploration, lateral cervical approach, thoracoscopic mediastinal exploration, mediastinal exploration, or a Chamberlain parasternal approach). Imaging studies of the abnormally located parathyroid gland minimize surgical morbidity and increase success rates. A recent adjunct in these difficult scenarios is the rapid iPTH assay, which can be performed and completed before finishing the operation. The assay is especially helpful for reexplorations in patients with multiglandular disease.

SECONDARY HYPERPARATHYROIDISM

Secondary hyperparathyroidism is an uncommon clinical disorder that develops in patients with chronic renal failure or intestinal malabsorption. PTH secretion appropriately increases because of hypocalcemia caused by hyperphosphatemia (renal failure) or dysfunctional calcium and vitamin D absorption (malabsorption).

Etiology

Calcium and phosphorus levels are inversely related. As chronic renal failure worsens and hypocalcemia occurs secondary to rising phosphate levels, the parathyroid glands secrete PTH in an attempt to counteract the hypocalcemic effects of renal failure. Compounding the problem is abnormal renal function that causes diminished 1 α -hydroxylase activity, which is necessary for proper vitamin D production. Because of the high content of aluminum in renal dialysate and the need to take phosphate-binding medications, patients on hemodialysis or oral phosphate binders can develop aluminum accumulation in bone, (aluminum intoxication), which contributes to osteomalacia and further PTH stimulation.

The cause of PTH stimulation with malabsorption syndromes is that calcium and vitamin D absorption is hindered to the point of hypocalcemia. Although PTH elevation is significant in renal failure and malabsorption, calcium regulation remains tight, and patients with either disease remain normocalcemic.

Clinical Manifestations

The diagnosis of secondary hyperparathyroidism often is made serologically before any symptoms are identified by the patient. When symptomatic, the most problematic feature for patients with secondary hyperparathyroidism usually is severe bone pain related to renal osteodystrophy. Less commonly, patients complain of soft-tissue calcifications (calcium tachyphylaxis) or pruritus. Others may develop bony fractures secondary to bone reabsorption.

Treatment

Initial therapy of secondary hyperparathyroidism is nonsurgical. In renal failure, dietary restriction of phosphate along with consumption of oral phosphate binders counteracts the underlying cause, hyperphosphatemia. In renal failure and malabsorption, oral calcium supplementation along with vitamin D ingestion is helpful. Altering the dialysate to minimize aluminum content and maximize calcium content is beneficial in renal failure patients.

Removing three and one-half parathyroid glands or total parathyroidectomy with reimplantation of a portion of tissue is indicated in secondary hyperparathyroidism for: (1) uncontrollable bone pain, (2) bone fractures, (3) intractable pruritus, (4) symptomatic ectopic calcifications, or (5) intractable disease that cannot be controlled medically. The preference is to excise three and one-half parathyroid glands and leave in situ approximately 50 mg of clearly viable parathyroid tissue.

TERTIARY HYPERPARATHYROIDISM

Tertiary hyperparathyroidism represents the continuation of secondary hyperparathyroidism; parathyroid tissue under the constant stimulation of hyperphosphatemia, and subsequent hypocalcemia, autonomously produces PTH and creates hypercalcemia. This rare situation is seen most commonly in patients with long-standing renal dysfunction who undergo successful renal transplantation.

Surgical removal of parathyroid tissue in tertiary hyperparathyroidism rarely is necessary. Most instances of tertiary hyperparathyroidism are short-lived; PTH levels eventually return to normal under the control of a functioning transplant kidney that corrects or improves hyperphosphatemia. Should persistent, autonomous overproduction of PTH occur with a well-functioning transplanted kidney, operative intervention is indicated.

MULTIPLE ENDOCRINE NEOPLASIA AND HYPERPARATHYROIDISM

Multiple endocrine neoplasia (MEN) is a collection of syndromes with an autosomal dominant pattern of inheritance. Hyperparathyroidism is one factor among many to consider in patients with endocrine abnormalities (Fig. 36-57). The hallmark of the pathology in all MEN syndromes, regardless of the organ involved, are multicentricity and bilaterality.

MEN I

Formerly known as Wermer's syndrome, MEN I includes pituitary, parathyroid, or pancreatic neoplasms. Given enough time, all patients with MEN I develop primary hyperparathyroidism. Tumors of the pituitary (15 to 50 percent) and pancreas (30 to 80 percent) are less common.

MEN IIA & IIB

All patients with Sipple's syndrome (MEN IIA) develop C-cell hyperplasia and, subsequently, medullary thyroid carcinoma if total thyroidectomy is not performed prophylactically. Adrenal medullary hyperplasia/pheochromocytoma (approximately 50 percent) and parathyroid abnormalities (10 to 25 percent) are less common but important to rule out. Patients with MEN IIB also can develop medullary thyroid carcinoma and adrenal neoplasms along with a marfanoid habitus, a typical facies (Fig. 36-58), and mucosal neuromas (Fig. 36-59). Chief-cell hyperplasia of the parathyroid glands is uncommon. MEN IIB is inherently virulent, and malignancy shortens the patient's life span if prophylaxis is not undertaken.

Clinical Manifestations

Over 75 percent of patients with MEN have a long and classic family history of endocrine abnormalities. The genetic abnormalities appear linked to changes in the long arm of chromosome 11 in MEN I and in the centromeric region of chromosome 10 in MEN II. Medical surveillance and genetic testing can identify abnormalities before they become symptomatic, but with time, patients with MEN present with complaints of visual changes, kidney stones, ulcer pain, or diabetes (MEN I); neck masses and hypertension (MEN IIA); and buccal and lingual nodules, hypertension, neck masses, and a marfanoid habitus (MEN IIB).

Evaluation

Patients with suspected MEN should undergo serologic surveillance of the appropriate markers of disease (Table 36-8). A classic family history for medullary thyroid carcinoma/hyperplasia warrants genetic testing for the ret-proto-oncogene. Such testing is important in children with normal serum calcitonin levels. Prophylactic thyroidectomy in the face of a positive ret-proto-oncogene and normal calcitonin levels is aggressive but appropriate therapy for these children, who otherwise develop a medullary thyroid malignancy.

Treatment

After a thorough work-up, operative intervention is suggested in order to minimize the ravages of primary hyperparathyroidism. Therapy for this syndrome is often less important than for the associated thyroid, pancreas, or adrenal disorders.

Treatment of primary hyperparathyroidism in MEN patients involves removing all but 50 mg of parathyroid tissue. A true solitary adenoma is unusual in these patients because of their propensity for multiple tumors. Given classic chief-cell hyperplasia, surgical options include a three and one-half gland parathyroidectomy or total parathyroidectomy with autotransplantation of heterotopic tissue. The concept of removing all but 50 mg of viable tissue is most important. Given the propensity of MEN patients to have more than four functioning glands, a thorough exploration and a transcervical thymectomy should be a routine part of the operation. It is not uncommon for these patients to have five or six parathyroid glands.

HYPERCALCEMIC CRISIS

Hypercalcemic crisis represents the life-threatening systemic condition brought on by elevated serum calcium levels (generally 13.0 mg/dL or higher).

Clinical Manifestations

Symptoms are varied and range from neuromuscular changes with mild fatigue and irritability to coma. Signs of dehydration are prominent. Gastrointestinal manifestations include anorexia, nausea, vomiting, and weight loss. Cardiac dysrhythmias, most commonly secondary to a shortened Q-T interval, may occur and can be lethal. Cancer cachexia may be evident in patients presenting with skeletal metastases. A palpable neck mass in the face of hypercalcemic crisis is considered a parathyroid carcinoma until proved otherwise. Therapy is initiated before attempting to differentiate the cause of hypercalcemia.

The list for the differential diagnosis is theoretically long, with any cause of hypercalcemia included, but in the vast majority (more than 90 percent) of cases it comes from advanced malignancy or primary hyperparathyroidism.

Treatment

Intravenous saline resuscitation is started and advanced to achieve a diuresis of higher than 100 mL/h. Dehydration is prominent with hypercalcemic crisis, and volume resuscitation (often 4 to 5 L) is the cornerstone of therapy. Once the patient appears adequately hydrated and urine output is more than 1 mL/kg/h, loop diuretics are given to stimulate a natriuresis and subsequent calciuresis. Cardiac dysrhythmias typically resolve with hydration and the subsequent decrease in the serum calcium level. Should dysrhythmias persist or become symptomatic, they are treated with standard agents for dysrhythmia. When hypercalcemia persists in the face of aggressive resuscitation and diuresis, the addition of mithramycin, phosphate binders (e.g., pamidronate), vitamin D, estrogen, calcitonin, or steroids may be useful. Hypercalcemia in patients with advanced malignancy and skeletal metastases can be difficult to normalize.

Patients with hyperparathyroidism usually can be stabilized with simple hydration and diuresis. Identification of a cervical mass on physical examination in the face of hypercalcemic crisis foreshadows a diagnosis of parathyroid carcinoma. Vocal cord function should be checked and the diagnosis of malignancy discussed with the patient. Ultrasonography may be useful while resuscitation is under way, or in the rare patient who is critically ill from persistent hypercalcemia. If a large adenoma is identified, hypercalcemia can be resolved quickly with a focused exploration and parathyroidectomy. With hemodynamic stability, regardless of the results of ultrasonography, operative intervention is undertaken early in patients with primary hyperparathyroidism. If the diagnosis of primary hyperparathyroidism is secure, even in comatose patients or in those with hemodynamic instability despite volume resuscitation, operative intervention should not be delayed. Postoperative transient hypocalcemia is common, and frequent calcium assessment is prudent because of the previous instability of these patients.

PARATHYROID CARCINOMA

Carcinoma of the parathyroid glands accounts for less than 1 percent of reported cases of hyperparathyroidism. In one study, 38 of 5010 patients with hyperparathyroidism were found to have parathyroid carcinoma (0.8 percent). Somewhat higher incidences have been reported in Japan, although the reason for this is unknown. The criteria that should alert the clinician to the possibility of this diagnosis can be divided into preoperative, intraoperative, and histopathologic findings.

Preoperative Findings

1. Palpable neck mass

An identifiable mass is present in 40 to 50 percent of patients with parathyroid carcinoma; this is in striking contrast to the incidence of less than 1 percent of this sign in patients with nonmalignant primary hyperparathyroidism. When found on physical examination, a mass secondary to parathyroid carcinoma has the expected characteristics of

malignancy—firmness and fixation to surrounding tissue.

2. Markedly abnormal biochemistry

In patients with parathyroid carcinoma, the biochemical parameters used for the diagnosis of hyperparathyroidism are highly abnormal. Mean serum calcium levels usually are greater than 13.0 mg/dL (normal range is 8.9 to 10.1 mg/dL), parathyroid hormone levels are elevated tenfold, and alkaline phosphatase levels are elevated threefold. In our practice, any patient found to have a serum calcium level above 13.0 mg/dL is presumed to have a parathyroid carcinoma until proved otherwise.

3. Complications of hyperparathyroidism

Clear-cut symptoms of hyperparathyroidism should alert clinicians to the possibility of the diagnosis of parathyroid carcinoma. In one study, bone disease, renal insufficiency, and nephrolithiasis occurred in 91 percent, 84 percent, and 56 percent, respectively, of patients with parathyroid carcinoma, which are considerably higher than the rates found in nonmalignant hyperparathyroidism.

Intraoperative Findings

The surgeon should suspect the presence of a parathyroid carcinoma after finding a parathyroid mass that is firmer than the usual soft, friable, nonmalignant parathyroid gland. When a gland has been transformed from the normal bean shape and kidney color to an irregular, pale white and is adherent to any of the surrounding structures, the diagnosis of malignancy is almost certain.

Histopathologic Findings

Most parathyroid carcinomas weigh more than 2000 mg, but weight does not aid in the diagnosis. The hallmark of malignancy histologically are the presence of a thick, fibrous capsule, fibrous septa interdigitating throughout the tumor, and the enlargement, hyperchromasia, and variation of nuclear size. Clear-cut invasion through the capsule to involve adipose tissue, nerves, thyroid gland, or surrounding muscle usually is present (Fig. 36-60).

Treatment

The treatment of choice is en bloc surgical resection of the tumor and the involved surrounding structures. This treatment usually involves a total thyroid lobectomy on the affected side. Regional nodal metastases are uncommon, but lymph nodes should be assessed and, if involved, surgically treated by appropriate resection, which may include a modified radical neck dissection. Adjuvant external beam radiotherapy or multimodality chemotherapy have been used with limited success. Symptomatic hypercalcemia is controlled with mithramycin and biphosphonates. Malignant recurrence after initial operative resection occurs in at least 66 percent of cases. When recurrent disease is seemingly localized, repeat surgical resection—repeated as often as can be safely performed—is optimal. Nonsurgical treatment modalities are inadequate. Most patients with parathyroid cancer will undergo one or more explorations ranging from cervical exploration to thoracotomy or craniotomy. In a study of 43 patients, the 3-year and 5-year survival rates were 84 and 69 percent, respectively. The cause of death often was attributed to the metabolic consequences of hypercalcemia and not to the local effects of malignancy.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 37 - Pediatric Surgery

Philip C. Guzzetta
Kathryn D. Anderson
R. Peter Altman
Kurt D. Newman
Jay J. Schnitzer

INTRODUCTION

Pediatric surgery combines the technical challenge of performing complex operations in neonates, the appeal of solving a wide array of problems in many anatomic locations, and the immense satisfaction of caring for children and offering them a long normal life. The approach to surgical problems in children should consider the long-term effects of the therapy as well as its immediate success rate. The resiliency of children to major surgical stress must always be balanced against their intolerance to inaccurate fluid and medication administration. The gratitude of a parent whose child has been saved by surgery is an additional benefit of this unique specialty.

GENERAL CONSIDERATIONS

Fluid and Electrolyte Balance

In an infant or child, the margin between dehydration and fluid overload is small. The infant is born with a surplus of body water, but within a few days this is excreted. At birth and for the first 10 days of life, fluid requirements are 65 to 100 mL/kg (750 to 1000 mL/m²). Daily maintenance fluids can be calculated by the formula: 100 mL/kg up to 10 kg, add 50 mL/kg for 11 to 20 kg, and add 25 mL/kg for each kilogram of weight thereafter. Because I.V. fluid orders are written as milliliters per hour, this can be conveniently converted to 4 mL/kg/h up to 10 kg, add 2 mL/kg/h for 11 to 20 kg, and add 1 mL/kg/h for each additional kilogram. For example, a 26-kg child has a maintenance fluid requirement of (10 × 4) + (10 × 2) + (6 × 1) = 66 mL/h. Fluid for maintenance is generally provided as 5% dextrose in ¼ or [frac13] normal saline. For short-term intravenous therapy, sodium 5 meq/kg/day and potassium 2 meq/kg/day will satisfy the daily need. Fluid and electrolyte losses secondary to protracted vomiting or diarrhea are corrected by modifying this formula according to the measured losses. In the infant the normal serum osmolarity is between 280 and 290 mO/L.

Newborns have the ability to concentrate their urine well by the fifth day of life; thus urine concentration as well as output must be considered when ordering I.V. fluids postoperatively. Daily electrolyte determinations are performed in infants on I.V. alimentation to obviate the serious sequelae of hyposmolarity or hyperosmolarity. If the child has a significant ongoing fluid loss (e.g., from a nasogastric tube), it is best to properly replace that loss with I.V. fluids at least every 8h.

Whatever the formula used to calculate fluid replacement for the infant or small child, there is no substitute for collecting and analyzing fluid losses and replacing the depleted constituents precisely.

Acid-Base Equilibrium

Acidosis in children is an ominous sign unless it is the result of a chronic bicarbonate loss from the gastrointestinal tract or an acid accumulation as in chronic renal failure. Acute acidosis usually implies inadequate tissue perfusion, and the cause of acidosis must be sought. Treatment of acute acidosis with a base deficit of greater than 5 is to administer sodium bicarbonate in the following dose: base deficit × weight in kilograms × 0.5 (in newborns). In small children, the last factor in the equation should be 0.4, and it should be 0.3 in older children. The dose should be diluted to 0.5 meq/mL because full-strength sodium bicarbonate is hyperosmolar. One-half the corrective dose is given, and the serum pH is measured again. During cardiopulmonary resuscitation (CPR), one-half the corrective dose can be given as an I.V. bolus and the other half given slowly I.V.

Alkalosis is usually caused by hyperventilation or stomach acid loss. In the ventilated child, hyperventilation can be corrected easily by decreasing the respiratory rate or tidal volume. In the child with gastric fluid loss, I.V. fluids of 5% dextrose, 0.5% normal saline, and 20 mEq KCl/L usually correct the alkalosis. This is the preferred replacement fluid in children with pyloric stenosis and alkalosis.

Blood Volume and Blood Replacement

A useful guideline for estimation of blood volume for the infant is 85 mL/kg of body weight. When packed red blood cells are utilized, the transfusion requirement is calculated as 10 mL/kg, which roughly is equivalent to a 500-mL transfusion for a 70-kg adult. In the child, coagulation deficiencies rapidly may assume clinical significance after extensive blood transfusion. It is advisable to have fresh frozen plasma and platelets available if more than 30 mL/kg have been transfused. Plasma is given in a dose of 20 mL/kg and platelets are given in a dose of 1 unit/5 kg. Each unit of platelets consists of 40 to 60 mL of fluid and the platelets can be spun down to a platelet “button” in infants who require restricted fluid administration.

Hyperalimentation and Nutrition

The physiologic nutritional demands imposed on the growing infant are well recognized. When these are compounded by illness and the need to repair tissue and heal surgical wounds, the risks of protein-calorie malnutrition are considerable. Thus parenteral nutritional support has been integrated into the management of infants and children with surgical illnesses. When the gastrointestinal tract is not usable because of mechanical, ischemic, or inflammatory disorders, several options for nutritional support are available. Techniques for delivering calories by a central or a peripheral venous route have been refined to the point where the caloric needs of all patients can be satisfied. Central venous catheters continue to be the most common route for administering I.V. hyperalimentation for an extended period of time. Peripheral I.V. alimentation, utilizing less concentrated but greater volumes of solutions, in combination with I.V. lipid supplements has eliminated the need for central alimentation in some patients. The infusion of all solutions utilized for alimentation, indeed any I.V. solution in an infant or small child, always should be controlled by a properly alarmed constant infusion pump. To prevent the development of trace metal deficiencies, supplementary copper, zinc, and iron are provided to patients receiving long-term parenteral nutritional support.

By using these techniques, positive nitrogen balance can be accomplished for all infants and children. Refinements and advances in the techniques of parenteral nutrition have had an enormous impact on the survival of pediatric surgical patients. Despite the remarkable advances in I.V. alimentation, enteral feeding is always the preferred route, and every effort is made to use the gastrointestinal tract by mouth or by tube feeding as quickly as possible.

Venous Access

Obtaining reliable vascular access in an infant or child is a major responsibility of the pediatric surgeon. The goal always must be to place the catheter in the least invasive, least risky, and least painful manner and in a location that is most likely to allow use of the catheter without complications for as long as needed. Regardless of the size of the patient, most infants requiring short-term I.V. medications and nutrition can be managed by peripheral percutaneous I.V. catheter placement. For long-term I.V. access in small infants, use of Silastic central venous catheters is best. These are usually placed by cutdown in infants and by a percutaneous technique in older children. When a central venous catheter is placed by cutdown, the preferred sites are external jugular vein, facial vein, or proximal saphenous vein. Secondly, the internal jugular vein may be used with a purse-string suture placed at the venotomy site to prevent venous occlusion. The catheters are tunneled to an exit site separate from the venotomy site. Currently, most central venous catheter placement by cutdown is done in the operating room under general anesthesia.

The percutaneous placement of central venous catheters in infants can be done with success in experienced hands. The risk of doing this procedure blindly in an infant always must be weighed against the benefit of avoiding a venous cutdown. In the infant with multiple central venous lines placed by cutdowns, the percutaneous approach may be the only reasonable option. Regardless of whether the catheter is placed by cutdown or percutaneously, a chest x-ray to confirm central location of the catheter tip and no chest complication from the placement is mandatory. We do this routinely within the operating room before the child is awakened.

Despite the technologic advances in the design of small catheters, the complication rates for central venous lines in infants are high. Sepsis from the catheter approaches a 10 percent incidence in many series, and superior or inferior vena caval occlusion is a significant risk, particularly in the smallest premature patients. We make every effort to feed the infants by the enteral route as soon and as completely as possible to minimize the length of time they depend on I.V. nutrition.

Thermoregulation

Infants or children compromised by disease are extremely thermolabile. Premature infants are particularly susceptible to changes in environmental temperature. Because they are unable to shiver and lack stores of fat, their potential for thermogenesis is impaired. Since these patients lack adaptive mechanisms to cope with the environment, the environment must be regulated. Attention to heat conservation during transport of the infant to and from the operating room is essential. Transport units incorporating heating units are necessary for premature infants. In the operating room, the infant is kept warm by the use of overhead heating lamps, a heating blanket, warming of inspired gases, and coverage of the extremities and head with occlusive materials. During abdominal surgery, it is best to use an adhesive plastic sheet to avoid sheets that are wet and cold. All fluids used to irrigate the chest or abdomen must be warmed to body temperature. Constant monitoring of the child's temperature is critical in a lengthy procedure.

Pain Control

The various physiologic responses of both neonates and fetuses to painful stimuli have been studied, and the results refute a commonly held belief that such patients do not have mature enough nervous systems to sense pain in a manner similar to adults. Despite concern that respiratory depression might result from administration of narcotic analgesics in infants, if the medications are given in the proper dose and with a proper time interval, they have an acceptable safety margin. In our hospital, virtually all parenteral analgesics are given intravenously. For all patients with an adequate understanding of the administration system, patient-controlled analgesia (PCA) is an excellent method of pain control. There is no fixed age below which PCA is not offered; however, patients below the age of 7 years generally are poor candidates for PCA.

Some attempts at parent-controlled analgesia have been made, but the emotional involvement of the parents and their desire to see the child pain-free rather than comfortable have limited the usefulness of that approach. Whenever possible, the use of anesthetic blocks, such as caudal tetracaine for hernias or thoracic epidural blocks for chest surgery, are encouraged. Two to three days of excellent pain relief can be obtained by epidural catheter administration of narcotics and/or local anesthetics, and such patients are monitored closely postoperatively. All pediatric patients having surgery, including circumcision, have pain, and should be given appropriate analgesia by the appropriate route to control that pain.

Minimally Invasive Surgery

New technology has allowed many procedures that traditionally were done as open procedures to be done using video-assisted, minimally invasive techniques in infants and children, regardless of size. Laparoscopic cholecystectomy, splenectomy, Nissen fundoplication, appendectomy, and placement of peritoneal dialysis and ventriculoperitoneal shunts are performed routinely with less morbidity than their open counterparts. Laparoscopic assessment of the asymptomatic side in inguinal hernia and localization of the nonpalpable testicle are also becoming commonplace.

Laparoscopic pyloromyotomy is comparable with the open technique, but not demonstrably superior. Thoracoscopy for evacuation of empyema, lung biopsy and thoracic mass excision or biopsy also has gained wide acceptance in pediatric surgery.

LESIONS OF THE NECK

Thyroglossal Duct Remnants

Pathology and Clinical Manifestations

The thyroid gland buds off the foregut diverticulum at the base of the tongue in the region of the future foramen cecum at 3 weeks of embryonic life. As the fetal neck develops, the thyroid tissue becomes more anterior and caudad until it rests in its normal position. The “descent” of the thyroid is intimately connected with the development of the hyoid bone. Residual thyroid tissue left behind in the migration may persist and present in the midline of the neck as a thyroglossal duct cyst. The mass is most commonly appreciated in the 2- to 4-year-old child when the baby fat subsides and irregularities in the neck are more readily apparent. Usually the cyst is encountered in the midline at or below the level of the hyoid bone and moves up and down with swallowing or with protrusion of the tongue. Occasionally it presents as an intrathyroidal mass. Most thyroglossal duct cysts are asymptomatic. If the duct retains its connection with the pharynx, infection may occur, and the resulting abscess will need drainage, occasionally resulting in a salivary fistula. Submental lymphadenopathy and midline dermoid cysts can be confused with a thyroglossal duct cyst. Rarely, midline ectopic thyroid tissue masquerades as a thyroglossal duct cyst and may represent the patient's only thyroid tissue. Therefore, if there is any question regarding the diagnosis or if the thyroid gland cannot be palpated in its normal anatomic position, it is advisable to obtain a nuclear scan to confirm the presence of a normal thyroid gland. Although rarely the case in children, in adults the thyroglossal duct might contain thyroid tissue that undergoes malignant degeneration.

Treatment

If the cyst presents with an abscess, treatment should consist of drainage and antibiotics. Following resolution of the inflammation, resection of the cyst in continuity with the central portion of the hyoid, and the tract into the pharynx, plus ligation at the foramen cecum (the Sistrunk operation), is curative. Lesser operations result in unacceptably high recurrence rates, and recurrence is more frequent following infection.

Branchial Cleft Anomalies

Paired branchial clefts and arches develop early in the fourth gestational week. The first cleft and the first, second, third and fourth pouches give rise to adult organs. The embryologic communication between the pharynx and the external surface may persist as a fistula (Fig. 37-1). A fistula is seen most commonly with the second branchial cleft, which normally disappears, and extends from the anterior border of the sternocleidomastoid muscle superiorly, inward through the bifurcation of the carotid artery, and enters the posterolateral pharynx just below the tonsillar fossa. The branchial cleft remnants may contain small pieces of cartilage and cysts, but internal fistulas are rare. A second branchial cleft sinus is suspected when clear fluid is noted draining from the external opening of the tract at the anterior border of the lower third of the sternomastoid muscle. The treatment is surgical, and complete removal of the cyst and tract is necessary for cure. Dissection of the sinus tract is facilitated by passing a fine lacrimal duct probe through the external opening into the tract and utilizing this as a guide for dissection. Injection of a small amount of methylene blue dye into the tract also may be useful. A series of two or sometimes three small transverse incisions in a “stepladder” fashion is preferred to a long oblique incision in the neck, which is cosmetically undesirable (Fig. 37-2).

Cystic Hygroma

Etiology and Pathology

Cystic hygroma (lymphangioma) occurs as a result of sequestration or obstruction of developing lymph vessels in approximately 1 in 12,000 births. Although the lesion can occur anywhere, the most common sites are in the posterior triangle of the neck, axilla, groin, and mediastinum. The cysts are lined by endothelium and filled with lymph. Occasionally unilocular cysts occur, but more often there are multiple cysts “infiltrating” the surrounding structures and distorting the local anatomy. A particularly troublesome variant of cystic hygroma is that which involves the tongue, floor of the mouth, and structures deep in the neck (Fig. 37-3). Adjacent connective tissue may show extensive lymphocytic infiltration. The mass may be apparent at birth or may appear and enlarge rapidly in the early weeks or months of life as lymph accumulates; most present by age 2 years. Extension of the lesion into the axilla or mediastinum occurs about 10 percent of the time and can be demonstrated preoperatively by chest x- ray, ultrasound (US) or computed tomographic (CT) scan. Occasionally cystic hygromas contain nests of vascular tissue. These poorly supported vessels may bleed and produce rapid enlargement and discoloration of the hygroma. The diagnosis of cystic hygroma by prenatal US, before 30 weeks' gestation, has detected a “hidden mortality” as well as a high incidence of associated anomalies, including abnormal karyotypes and hydrops fetalis.

Infection within the cysts usually caused by *Streptococcus* or *Staphylococcus* may occur. In the neck this can cause rapid enlargement, which may result in airway compromise. Rarely, it may be necessary to carry out percutaneous aspiration of an infected cyst to relieve respiratory distress.

Treatment

Surgical excision is the treatment of choice. Total removal may not be possible because of the extent of hygroma and its proximity to and intimate relationship with adjacent nerves, muscles, and blood vessels. Radical ablative surgery is not indicated for this benign lesion. Conservative excision and unroofing of remaining cysts is advised, with repeated partial excision of residual hygroma if necessary, preserving all adjacent crucial structures. Postoperative wound drainage is important and is best accomplished by closed-suction technique. Despite this, fluid may accumulate beneath the surgically created flaps in the area from which the hygroma was excised, requiring multiple needle aspirations. Injection of sclerosing agents (OK-432 or bleomycin) with favorable results has been reported, but widespread use of these agents has not been seen in this country.

Cervical Adenitis

Enlarged tender lymph nodes are usually the result of bacterial infection (*Staphylococcus* or *Streptococcus*). Treatment of the primary cause (e.g., otitis media or pharyngitis) with antibiotics often is all that is necessary. However, when the involved nodes become fluctuant, incision and drainage are indicated, followed by appropriate wound care and dressing changes. More chronic forms of lymphadenitis, including infections with tuberculosis and atypical mycobacteria, as well as cat-scratch fever, require further work- up to make the diagnosis. In certain cases, extensive excision of the involved nodes will be necessary.

Torticollis

Fibrosis of the sternocleidomastoid muscle in infancy may produce shortening of the muscle and rotation of the head towards the opposite side. In two-thirds of cases, a mass may be palpated in the affected muscle. Histologically, the lesion is characterized by the deposition of collagen and fibroblasts around atrophied muscle cells. For up to 80 percent of patients, operation may not be needed, and physical therapy may be of some benefit. When necessary, surgical transection of the affected muscle is curative.

RESPIRATORY SYSTEM

Congenital Diaphragmatic Hernia (Bochdalek)

Pathology

During formation of the diaphragm, the pleural and coelomic cavities remain in continuity by means of the pleuroperitoneal canal. The posterolateral communication is the last to be closed by the developing diaphragm. Failure of diaphragmatic development leaves a posterolateral defect known as Bochdalek hernia. This anomaly is encountered more commonly on the left.

Incomplete development of the posterior diaphragm allows the viscera to fill the chest cavity. The abdominal cavity is small and undeveloped and remains scaphoid after birth. Both lungs are hypoplastic, with decreased bronchial and pulmonary artery branching; lung weight, lung volume, and DNA content are decreased, all more severely evident on the ipsilateral side. On rare occasions lung development is less adversely affected, signaling a greater chance of survival of the infant.

Prenatal US is successful in making the diagnosis of congenital diaphragmatic hernia (CDH) in the unborn child as early as 15 weeks' gestation. US findings include herniated abdominal viscera, abnormal anatomy of the upper abdomen, and mediastinal shift away from the herniated viscera. Herniation appears to be a dynamic process in which the viscera move in and out of the chest in some fetuses.

Prenatal criteria for a high-risk fetus include polyhydramnios, early diagnosis, a dilated stomach in the chest, and low lung-thorax ratio. Amniocentesis with karyotypes may show chromosomal defects, especially trisomy 18 and 21. Associated anomalies, once thought to be uncommon, are identified in 40 percent of these infants, most often of the heart, brain, genitourinary system, craniofacial structures, or limbs. Prenatal repair of CDH has had some success, and recent work has suggested that prenatal obstruction of the trachea may exert a salutary effect on lung growth.

Following delivery, the bowel fills with air, and the mobile mediastinum shifts to the opposite side of the chest, compromising air exchange in the contralateral lung. Respiratory failure occurs as a combination of the effect of hypoplasia of the lungs, persistence of fetal circulation as a result of pulmonary hypertension, and the mechanical compression of the lungs from the herniated viscera. Chest x-ray is diagnostic (Fig. 37-4).

Treatment

Many infants are symptomatic at birth due to hypoxia, hypercarbia, and metabolic acidosis, and prompt cardiorespiratory stabilization is mandatory. The goal of preoperative therapy is to reverse the persistent pulmonary hypertension that results in right-to-left shunting across the open foramen ovale and the ductus arteriosus. Improvement in the infant's status is apparent by improved oxygenation and ventilation. Immediate surgical repair is no longer performed. Instead the infant is stabilized by mechanical ventilation with 100% oxygen. Hyperventilation is no longer widely practiced and the infants are often ventilated by high-frequency oscillation to avoid the pulmonary ravages of high inflating pressures. The goal of preoperative therapy is to prevent or reverse pulmonary hypertension and metabolic acidosis and provide adequate oxygen delivery. Levels of hypercarbia once considered unthinkable are tolerated. Newer techniques, such as intratracheal pulmonary ventilation and liquid ventilation, may offer some improved survival. The use of nitric oxide has been disappointing in correcting the pulmonary hypertension associated with CDH.

Infants with CDH who fail to respond, and remain severely hypoxic despite optimal care are candidates for treatment of their respiratory failure by extracorporeal membrane oxygenation (ECMO). Venovenous or venoarterial bypass is used. Venovenous bypass is established with a single cannula through the internal jugular vein, with blood removed from and infused into the right atrium by separate ports. Venoarterial bypass is used preferentially by some centers because it provides the cardiac support that is often needed. The right atrium is cannulated by means of the internal jugular vein and the aortic arch through the right common carotid artery (Fig. 37-5). As much of the cardiac output is shunted through the membrane oxygenator as needed to provide oxygenated blood to the infant and remove carbon dioxide. The infant is maintained on bypass until the pulmonary hypertension is reversed and lung function, as measured by compliance, is improved. This is usually seen within 7 to 10 days but in some infants may take up to 3 weeks to occur.

Newborns who do not demonstrate significant improvement over this time have pulmonary hypoplasia that will not benefit from further extracorporeal life support.

The timing of diaphragm repair is controversial. Some surgeons perform early repair on bypass; others wait until the infant's lungs are fully recovered, repair the diaphragm and discontinue bypass within hours of surgery. Still others repair the diaphragm only after the infant is off bypass.

Operative repair of the diaphragmatic hernia is best accomplished by an abdominal approach. Through a subcostal incision the abdominal viscera are withdrawn from the chest, exposing the defect in the diaphragm. The anterior margin is often apparent, while the posterior muscular rim is attenuated. Most infants who need ECMO support prior to hernia repair have large defects, often lacking the medial and posterior margins. Prior to the availability of ECMO therapy, most of these infants died.

If the infant is heparinized on bypass, minimal dissection of the muscular margins is performed. Electrocautery is used liberally to minimize postoperative bleeding. About three-fourths of infants repaired on bypass require prosthetic material to patch the defect, suturing it to the diaphragm remnant or around ribs or costal cartilages for the large defects. If there is adequate muscle for closure, a single layer of nonabsorbable horizontal mattress suture is used to close the defect. Just before the repair is complete, a chest tube is positioned in the thoracic cavity and connected to 4 to 5 cmH₂O of suction.

Anatomic closure of the abdominal wall may be impossible after reduction of the viscera. Skin flaps are elevated and closed, creating a ventral hernia that may be closed using a Goretex patch. Rarely, a Silastic "silo" must be placed because the skin cannot be approximated.

If the diaphragm has been repaired on ECMO, weaning and decannulation are accomplished as soon as possible. All infants are ventilated postoperatively to maintain postductal arterial oxygenation of 80 to 100 torr. Very slow weaning from the ventilator is necessary to avoid recurrent pulmonary hypertension. Oscillation ventilation may be switched to conventional ventilation as part of the process of weaning.

Currently there are no absolute criteria that successfully exclude the use of ECMO for these infants. Severe hypoxia or hypercarbia does not select infants who uniformly are nonsurvivors. At present, over 60 percent of infants with congenital diaphragmatic hernia who are supported by ECMO can be expected to survive.

Congenital Lobar Emphysema

Congenital lobar emphysema (CLE) is a rare condition manifested during the first few months of life as a progressive hyperexpansion of one or more lobes of the lung. It can be life-threatening in the newborn period but in the older infant causes less respiratory distress. Air entering during inspiration is trapped in the lobe; on expiration, the lobe cannot deflate and progressively overexpands, causing atelectasis of the adjacent lobe or lobes. This hyperexpansion eventually shifts the mediastinum to the opposite side and compromises the other lung. CLE usually occurs in the upper lobes of the lung (left greater than right), followed next in frequency by the right middle lobe, but it also can occur in the lower lobes. It is caused by intrinsic bronchial obstruction from poor bronchial cartilage development or extrinsic compression. Approximately 14 percent of children with this condition have cardiac defects, with an enlarged left atrium or a major vessel causing compression of the ipsilateral bronchus.

Symptoms range from mild respiratory distress to full-fledged respiratory failure with tachypnea, dyspnea, cough, and late cyanosis. These symptoms may be stationary or they may progress rapidly. Diagnosis is made by chest x-ray which shows a hyperlucent affected lobe with adjacent lobar compression and atelectasis with varying degrees of shift of the mediastinum to the opposite side and compression of the contralateral lung (Fig. 37-6). If definitive diagnosis is unclear by chest x-ray, CT scan may be helpful. Unless foreign body or mucous plugging is suspected as a cause of hyperinflation, bronchoscopy is not advisable because it can produce more air trapping and cause life-threatening respiratory distress in a stable infant. Treatment is resection of the affected lobe, which may need to be done in urgent circumstances. The prognosis is excellent.

Congenital Cystic Adenomatoid Malformation

The clinical presentation of this lesion can be quite variable and lead to errors in diagnosis. The malformation consists of cystic proliferation of the terminal airway, producing cysts lined by mucus-producing respiratory epithelium, and elastic tissue in the cyst walls without cartilage formation. There may be a single cyst with a wall of connective tissue containing smooth muscle. Cysts may be large and multiple (type I), smaller and more numerous (type II), or resemble fetal lung without macroscopic cysts (type III). Most frequently located in the left lower lobe, this lesion can occur in any lobe and may occur in both lungs simultaneously. In the left lower lobe, type I may be confused at birth with a congenital diaphragmatic hernia. Clinical symptoms may range from none at all to severe respiratory failure. The cyst(s), whether single or multiple, can produce air trapping and thus be confused with congenital lobar emphysema. They also can be involved with repeated infections and produce fever and cough in older infants and children, and be confused with pulmonary sequestration or pneumatoceles.

The diagnosis often can be made by chest x-ray. In certain cases, US or CT scan may be definitive. US evaluation of the fetus may suggest the diagnosis prenatally, and in the newborn US may be useful, especially to distinguish between adenomatous malformation and congenital diaphragmatic hernia. In older children who present with repeated infections, it may be wise to rule out sequestration by Doppler US or aortography. Resection is curative and may need to be performed urgently in the infant with severe respiratory distress. Lobectomy is usually required and the children do well following this procedure.

Pulmonary Sequestration

Pulmonary sequestration is uncommon and consists of a mass of lung tissue, usually in the left lower chest, occurring without the usual connections to the pulmonary artery or tracheobronchial tree. There are two kinds of sequestration. Extralobar sequestration is usually a small area of nonaerated lung separated from the main lung mass, with a systemic blood supply, located immediately above the left diaphragm. It is commonly found in cases of CDH. Intralobar sequestration more commonly occurs within the parenchyma of the left lower lobe but can occur on the right. There is no major connection to the tracheobronchial tree, but a secondary connection may be established, perhaps through infection or via adjacent intrapulmonary shunts. The blood supply is systemic from the aorta, often multiple, and frequently originates below the diaphragm (Fig. 37-7). Venous drainage of both types can be systemic or pulmonary. The cause of sequestration is unknown but most probably involves an abnormal budding of the developing lung that picks up a systemic blood supply and never becomes connected with the bronchus or pulmonary vessels.

Extralobar sequestration is asymptomatic and is usually discovered incidentally on chest x-ray. If the diagnosis can be confirmed, e.g. by CT scan, resection is not necessary. Diagnosis of intralobar sequestration, on the other hand, is usually made after repeated infections manifested by cough, fever, and consolidation in the posterior basal segment of the left lower lobe. Increasingly the diagnosis is being made in the early months of life by US, and color Doppler often can be helpful to delineate the systemic arterial supply. An arteriogram may be helpful in the older child in whom US of the chest is difficult, to help delineate the origin of the systemic blood supply and make surgical excision safer. Removal of the entire left lower lobe is usually necessary since the diagnosis often is made late after multiple infections. Occasionally the sequestered part of the lung can be removed segmentally. Prognosis is excellent.

Bronchogenic Cyst

Bronchogenic cysts can occur anywhere along the respiratory tract from the neck out into the lung parenchyma. They can present at any age. Histologically, they are hamartomatous and usually consist of a single cyst lined with respiratory epithelium and containing cartilage and smooth muscle. They are probably embryonic rests of foregut origin that have been pinched off from the main portion of the developing tracheobronchial tree and are closely associated in causation with other foregut duplication cysts associated with the esophagus. Bronchogenic cysts may be seen on prenatal US but are discovered most often postnatally incidentally on chest x-ray and may be completely asymptomatic. They may, however, produce symptoms, depending on their anatomic location. In the paratracheal region of the neck they can produce airway compression and respiratory distress. In the lung parenchyma, they may become infected and present with fever and cough. In addition they may cause obstruction of the bronchial lumen with distal atelectasis and infection. Chest x-ray usually will show a dense mass, and CT scan (Fig. 37-8) or magnetic resonance imaging (MRI) will delineate the precise anatomic location of the lesion. Treatment is resection of the cyst. Resection may need to be undertaken in emergency circumstances for airway compression. The prognosis is excellent.

Bronchiectasis

Bronchiectasis is an abnormal dilatation of the bronchi and bronchioles associated with chronic suppurative disease of the airways. Usually the children have an underlying congenital pulmonary anomaly, cystic fibrosis, or immunologic deficiency. Bronchiectasis also can result from infection secondary to a neglected bronchial foreign body. The symptoms include a chronic cough, often productive of purulent secretions, recurrent pulmonary infection, and hemoptysis. The diagnosis is suggested by a chest x-ray that shows increased bronchovascular markings in the affected lobe. Historically, a bronchogram was required to demonstrate the saccular or fusiform distortion of the peripheral bronchi. Chest CT delineates these lesions of bronchiectasis with excellent resolution. The preferred treatment for bronchiectasis is medical, consisting of antibiotics, postural drainage, and avoidance of inhaled pollutants. Lobectomy or segmental resection is indicated for localized disease that has not responded appropriately to medical therapy.

Foreign Bodies

Airway

Aspiration of foreign bodies most commonly occurs in the toddler age group. The material aspirated is often the ubiquitous peanut, but there is a plethora of small objects that youngsters put into their mouths that can be aspirated. A solid foreign body often will cause air trapping, with hyperlucency of the affected lobe or lung seen especially on expiration (Fig. 37-9). Oil from the peanut is very irritating and may cause pneumonia. Delay in diagnosis can lead to atelectasis and infection. The most common anatomic location for a foreign body is the right main stem bronchus or the right lower lobe. The child usually will cough or choke during eating but may then become asymptomatic. Total respiratory obstruction with tracheal foreign body may occur, but usually respiratory distress is mild if present at all. A unilateral wheeze is often heard on auscultation. This wheeze often leads to an inappropriate diagnosis of "asthma" and may delay the correct diagnosis for some time. Chest x-ray will show a radiopaque foreign body, but in the case of nuts, seeds, or plastic toy parts, the only clue may be nondeflation of the affected lobe on an expiratory film or fluoroscopy. Bronchoscopy confirms the diagnosis and allows removal of the foreign body. This can be a very simple procedure or it may be extremely difficult, especially with a smooth foreign body that cannot be grasped easily or one that has been retained for some time. Bronchiectasis may be seen as an extremely late phenomenon after repeated infections of the poorly aerated lung and may require partial or total resection of the affected lobe.

Esophagus

The most common foreign body in the esophagus is a coin. Again, toddlers are most commonly affected. The coin is retained in the esophagus at one of three locations: the cricopharyngeus, the area of the aortic arch, or the gastroesophageal junction, all areas of normal anatomic narrowing. Symptoms are variable depending on the anatomic position of the foreign body and the degree of obstruction, but occasionally include respiratory distress from compression of the trachea or inability to swallow saliva and drooling. Many children are totally asymptomatic. The chest x-ray is diagnostic in the case of a coin. A contrast swallow may be required for nonradiopaque foreign bodies. Coins lodged within the upper esophagus for less than 24 hours usually can be removed safely by an experienced physician in the emergency room. For all other situations, the treatment is by esophagoscopy, rigid or flexible, and removal of the foreign body. In the case of sharp foreign bodies such as open safety pins, extreme care is required on extraction to avoid injury of the esophagus itself. Diligent follow-up is required after removal of foreign bodies such as batteries, which can cause strictures, and occasionally, a foreign body calls

attention to a pre-existing stricture.

ESOPHAGUS

Esophageal Atresia and Tracheoesophageal Fistula

Perhaps no pediatric surgical condition has had as gratifying progress as the treatment of esophageal atresia and tracheoesophageal fistula. In the early decades of the twentieth century, all babies born with this anomaly died. In 1939 Ladd and Leven achieved the first success with a complicated management plan that involved gastrostomy and fistula ligation, followed by esophageal reconstruction with a skin tube. Over 50 years ago, Cameron Haight in Ann Arbor, Michigan, performed the first successful primary anastomosis for esophageal atresia. This landmark achievement has been followed by steady improvement in survival for babies born with this condition. Currently most infants survive unless an associated major anomaly supervenes. This progress has resulted not only from refinements in surgical technique, but also from advances in nutrition, respiratory care, antibiotics, anesthesia, cardiology, genetics, and imaging technology.

Etiology and Pathology

The cause of this malformation is uncertain; however, it almost certainly relates to the common embryologic origin of the esophagus and trachea. Normally they divide into separate tubes by 34 to 36 days' gestation, at which time the submucosal and muscular layers of both esophagus and trachea are apparent. When this separation is incomplete, a wide spectrum of malformations may result. The most common esophageal malformation (85 percent) occurs as a blind-ending upper esophageal segment. The lower portion of the esophagus connects to the trachea (Gross-Vogt type C) (Fig. 37-10). The lower esophageal fistula usually joins the trachea at or just above the tracheal bifurcation, admitting inspired air into the stomach and, in a retrograde fashion, gastric juice into the lungs.

Associated congenital anomalies are common and are the most significant factor influencing survival. Frequently esophageal atresia occurs as part of a collection of defects known by the acronym VATER or VACTERRL (vertebral, anorectal, cardiac, tracheoesophageal, renal, radial, limb). In nearly 20 percent of the infants born with esophageal atresia, some variant of congenital heart disease occurs.

Clinical Manifestations

The earliest and most obvious clinical sign of esophageal atresia is regurgitation or drooling. Feedings are followed by choking or coughing, indicating aspiration. Abdominal distention is a prominent feature, occurring when inspired air is transmitted through the fistula into the stomach. This sign differentiates "pure" esophageal atresia, in which no air can pass into the gastrointestinal tract, and the abdomen is therefore scaphoid.

Regurgitated gastric juice passes through the fistula and into the trachea and lungs, leading to chemical pneumonia. The pulmonary problems are compounded by atelectasis, which derives from diaphragmatic elevation secondary to gastric distention. The diagnosis may not be confirmed at the initial newborn examination by a physician, unless an attempt is made to pass a tube into the stomach. Nurses who are feeding the baby and observing behavioral patterns such as the accumulation of mucus or saliva often make the early diagnosis. In recent years, the diagnosis has been made prenatally with US examination. Polyhydramnios arouses suspicion, and failure to visualize a fetal stomach with US is suggestive of the diagnosis of esophageal atresia.

For many surgeons, the demonstration of a tube coiled in an air-filled upper pouch on a plain film is adequate to diagnose esophageal atresia. For others, a contrast x-ray study is required to confirm the diagnosis. Contrast medium instilled into the pouch outlines the blind upper esophagus (Fig. 37-11). It also may document the existence of an upper pouch fistula. Tracheal air enters the stomach via the fistula and establishes the presence of a tracheoesophageal communication on the x-ray. The condition of the lungs and the existence of pneumonia or atelectasis are also essential information. Examination of the heart and great vessels with echocardiography is important to exclude cardiac defects and verify the location of the aortic arch. In a stable infant, bronchoscopy at the time of repair allows localization of the level of the fistula, exclusion of upper pouch fistulas, and identification of a laryngeotracheoesophageal cleft.

Treatment

With the diagnosis secure, the following measures should be instituted immediately: (1) infant warmer; (2) 30-degree head-up position; (3) route for I.V. therapy; (4) antibiotic treatment (even if pneumonia is not yet clinically manifest); and (5) sump catheter suction in the upper pouch. It has become apparent that the major determinant of poor survival is the presence of other severe anomalies; low birth weight and pneumonia are seldom risk factors for death. The optimal management strategy has evolved and is based on the physiologic status of the infant. If the infant is stable, a primary repair is undertaken. If unstable, the repair is delayed until the problems are resolved and the infant can safely undergo a major operation.

Primary Surgical Correction

The operative technique for primary repair is depicted in Fig. 37-12. The keys are careful division of the fistula with closure of the tracheal opening and adequate mobilization of the upper esophagus to permit an anastomosis without tension. The operation for primary repair of esophageal atresia and tracheoesophageal fistula has changed little since its description by Cameron Haight, although most surgeons now perform a single-layer anastomosis rather than the original two-layer. The transpleural approach may be easier, but the retropleural approach offers advantages. Exposure with this method is perfectly adequate, and protection of the lung by maintaining its pleural envelope has a salutary postoperative effect. More important, an anastomotic leak will not communicate with the pleural cavity. A leak can be drained directly from the mediastinum posteriorly with decreased morbidity. Although once routine, gastrostomy is rarely performed in the stable group of infants having primary repair.

Care must be taken when mobilizing the distal esophagus to avoid devascularization, since the blood supply is segmental from the aorta. Most of the esophageal length is obtained from mobilizing the upper pouch, since the blood supply travels via the submucosa from above. Livaditis showed that incising the muscle of the upper pouch circumferentially produced remarkable lengthening without compromising blood supply. This approach is useful if the distance between the two pouches precludes anastomosis without producing undue tension.

Delayed or Staged Repair

For infants with serious coexisting anomalies or extreme prematurity, survival is limited. For these infants, a gastrostomy is placed and repair delayed. Nutritional support, coupled with suction of the upper pouch and gastrostomy drainage, makes it possible to maintain infants with esophageal atresia and tracheoesophageal fistula indefinitely. Growth and weight gain are achieved, pulmonary status is cleared, and the congenital anomalies are studied and corrected. Some infants, particularly premature ones with noncompliant lungs, will not stabilize and require urgent division of the fistula. Anastomosis of the esophagus is accomplished at a later date.

Surgical correction of esophageal atresia with tracheoesophageal fistula leads to a satisfactory outcome with nearly normal esophageal function in most patients. Overall survival rates of greater than 90 percent have been achieved in patients classified as stable. Unstable infants have an increased mortality (40 to 60 percent survival) because of potentially fatal associated cardiac and chromosomal anomalies. However, the use of the staged procedure also has increased the survival in these high-risk infants.

Postoperative Course and Complications

The infant is maintained on parenteral nutrition for several days after repair. The chest tube is assessed daily for the presence of saliva indicating a leak. Many surgeons obtain a contrast swallow 1 week after repair to assess the caliber of the anastomosis and to look for a leak. If there is no leak, the infant is begun on feedings. A leak at the anastomosis usually heals without intervention, particularly if the pleural envelope has been maintained and drainage accomplished from the mediastinum via the retropleural route. Improved management of infection, nutrition, and respiratory support in infants has contributed to success in handling anastomotic leaks.

Strictures are not infrequent (10 to 20 percent), particularly if a leak has occurred. A stricture may become apparent at any time, from the early postoperative period to months or years later. It may present as choking, gagging, or failure to thrive, but often becomes clinically apparent with the transition to eating solid food. A contrast swallow or esophagoscopy is confirmatory, and simple dilatation is usually corrective. Occasionally, repeated dilatations are required.

"Recurrent" tracheoesophageal fistula may represent a missed upper pouch fistula or a true recurrence. The mediastinal infection seen with an anastomotic disruption may cause recurrence of the tracheoesophageal fistula. This complication may be difficult to identify and requires another operation to divide the tracheoesophageal fistula.

It has become apparent that even after repair, esophageal motility and the anatomy of the gastroesophageal junction are altered, potentially leading to gastroesophageal reflux. The clinical manifestations of this reflux are similar to those seen in other infants with primary gastroesophageal reflux. An antireflux procedure, such as a Nissen fundoplication, is used to prevent further reflux, but the child may have feeding problems after antireflux surgery as a result of the innate dysmotility of the distal esophagus.

Isolated Esophageal Atresia

Among those infants born with esophageal anomalies, 8 percent have isolated esophageal atresia. Characteristically, infants with isolated esophageal atresia present with a scaphoid abdomen, since the gastrointestinal tract is devoid of air. The x-ray finding of a blind upper pouch and the absence of air below the diaphragm is pathognomonic of isolated esophageal atresia without fistula.

The ideal surgical management is based on conservation of the native esophagus. Many surgeons perform a gastrostomy for feeding and wait 6 to 10 weeks for the upper pouch to dilate and elongate. On imaging studies, if it looks as though the two ends can be approximated, primary anastomosis is attempted. If the ends do not reach, however, esophagostomy with the upper esophageal pouch brought to the skin of the left side of the neck allows drainage of saliva and prevents aspiration. Esophageal replacement with a colon segment or reversed gastric tube is then recommended at 1 year of age.

Isolated (H-Type) Tracheoesophageal Fistula

In rare instances (4 percent), an isolated congenital fistula connecting the trachea to the esophagus may exist. In this anomaly, both the trachea and the esophagus are otherwise normal, with no narrowing or obstruction. Infants with this condition seem to swallow normally. The clinical features are subtle; weeks or months may elapse before a correct diagnosis is made. The presence of an H-type tracheoesophageal fistula is suggested by the following triad of symptoms: (1) choking when feeding, (2) gaseous distention of the bowel, and (3) recurrent aspiration pneumonia. Diagnosis usually can be confirmed by cine contrast x-ray studies or by bronchoesophagoscopy. Definitive treatment consists of dividing the fistula. Surgical closure of the esophagus and trachea must be meticulous, and encroachment on the lumen of the trachea must be avoided. The fistula is usually accessible to surgical repair through an incision just above the clavicle.

Laryngotracheoesophageal Cleft

Another variant of the failure of the trachea and esophagus to separate results in a communication that may be confined to the larynx or extend over the entire length of the trachea

to the carina. The diagnosis is suggested by recurrent aspiration and is confirmed with bronchoscopy. The cleft is frequently present in association with esophageal atresia. The early management consists of tracheostomy, gastrostomy, and antireflux operation. At a later time the cleft is closed. Although once always fatal, most of these defects can now be repaired successfully.

Corrosive Injury of the Esophagus

The toddler age group is the most vulnerable to injury by ingestion of corrosive substances. Curiosity may lead a child to ingest crystals or an unknown liquid especially if it is placed in a drinking container. Since children gulp rather than taste, a single mouthful can cause severe injury. Both strong alkalis and strong acids produce injury by liquefaction or coagulation necrosis, and since all corrosive agents are extremely hygroscopic, the caustic will cling to the esophageal epithelium. Injury and therefore subsequent strictures occur at the anatomic narrowed areas of the esophagus, cricopharyngeus, midesophagus, and gastroesophageal junction. A child who has swallowed an injurious substance may be symptom-free but usually will be drooling and unable to swallow saliva. Most corrosives are swallowed accidentally by children and the volume ingested is limited to a single swallow because of the noxious taste and pain. Distal injuries to the stomach are therefore less common than in adults. However, the child must be observed carefully for signs of peritoneal irritation.

There is no effective immediate antidote. Diagnosis is by careful physical examination of the mouth and endoscopy with a flexible or a rigid esophagoscope. It is important to endoscope only to the first level of the burn in order to avoid perforation. Early barium swallow may delineate the extent of the mucosal injury, and it is important to realize that the esophagus may be burned without evidence of a mouth injury. Steroids have not been shown to alter the development of stricture or modify the extent of injury and therefore are no longer used, but antibiotics are continued for 3 weeks after injury. Circumferential esophageal injuries with necrosis will almost certainly stricture and therefore will be dealt with most safely by a gastrostomy and a string inserted either through the esophagus immediately or at a repeat esophagoscopy 3 weeks later. Dilatation is attempted after a minimum of 3 weeks to avoid perforation and is restricted to established strictures. For more severe injuries, retrograde dilatations are safest, using graduated dilators brought through the gastrostomy and advanced into the esophagus via the transesophageal string. For less severe injuries, dilatation may be attempted in antegrade fashion by either graded bougies or balloons. Perforation during dilation is extremely rare in experienced hands. Treatment of an iatrogenic perforation is controversial and ranges from nonoperative treatment with closed chest drainage to immediate operative closure. Abandonment of the esophagus is usually not necessary if recognition is prompt and treatment timely. Dilatations are continued as often as necessary, with the goal of keeping the child out of the hospital and eating a normal diet until the stricture is resolved or it is clear that esophageal substitution is required. Strictures have resolved as late as 6 months to a year after injury. Every attempt should be made to preserve the native esophagus.

Esophageal Substitution

Esophageal substitution is required less frequently than heretofore. The incidence of caustic strictures has been markedly diminished by the introduction of child-proof containers, and there have been successes with end-to-end anastomosis after a period of growth in infants with isolated esophageal atresia or wide gap esophageal atresia with fistula. However, these conditions are still the major indications for esophageal substitution in children. The options for esophageal substitution are the colon (right colon or transverse/left colon), reversed gastric tube, gastric pull-up or pedicled or free, grafts of the jejunum. The right colon is based on a pedicle of the midcolic artery, the left colon, on a pedicle of the midcolic or left colic artery (Fig. 37-13). Gastric tubes are made from the greater curvature of the stomach based on the pedicle of the left gastroepiploic artery. When the whole stomach is used, the vascularity is supplied by the right gastric artery. The route of the neoesophagus to the neck is at the surgeon's discretion. The substernal and left transthoracic route are still the most popular. Bringing the substitute through the posterior mediastinum, either following blunt esophagectomy in the case of caustic stricture or by creating a new tunnel in the case of esophageal atresia, avoids a thoracotomy and ensures the straightest path from neck to abdomen. Long-term follow-up has shown that all methods of substitution can support normal growth and development, and the children enjoy reasonably normal eating habits. Because of the potential for late complications such as ulceration and stricture, follow-up into adulthood is mandatory, but complications appear to diminish with time.

Gastroesophageal Reflux

Gastroesophageal reflux (GER) occurs to some degree in all children. GER is considered to be pathologic when it interferes with nutrition, causing failure to thrive; damages the esophagus, causing bleeding or stricture formation; or it leads to pulmonary problems such as reactive airway disease, aspiration pneumonia or apnea. Failure to thrive and pulmonary problems are particularly common in infants with GER, whereas strictures and esophagitis are more common in older children and adolescents. The majority of children with severe GER that is unresponsive to medical therapy are neurologically impaired.

Clinical Manifestations

A history of repeated episodes of vomiting in an infant is the clearest indication of GER after obvious anatomic obstruction at or beyond the pylorus has been excluded. When the vomiting is associated with failure of normal development or chronic respiratory symptoms, the likelihood of pernicious GER is increased.

A child suspected of having GER is evaluated with a barium swallow to rule out an anatomic obstruction in the stomach or duodenum and a 24-h pH probe study to determine the severity of the reflux. The radioisotope “milk scan” is used to evaluate gastric emptying but is not obtained routinely if the upper gastrointestinal study shows normal gastric emptying. Endoscopy with biopsies to prove esophagitis is utilized only in selected patients.

Treatment

Most patients with GER are treated initially by conservative means. In the infant, propping and thickening the formula with rice cereal are generally recommended, although some authors prefer a prone, head-up position. In the infant unresponsive to position and formula changes and the older child with severe GER, medical therapy is based on gastric acid reduction with an H₂ blocking agent (i.e. ranitidine) and a prokinetic agent (i.e. cisapride). Surgery is indicated if medical therapy has failed to prevent complications of GER or there is an anatomic abnormality such as a large hiatal hernia or esophageal stricture. In children less than 1 year old or those unable to swallow because of neuromotor problems, a gastrostomy is performed routinely with the antireflux procedure.

Antireflux surgery is either by a complete (i.e. Nissen) or partial (i.e. Thal) fundic wrap around the distal esophagus. These procedures can be done safely in children laparoscopically. Postoperative complications of gagging, pulmonary problems and wrap breakdown with recurrent GER are quite high in children with neurologic impairment.

GASTROINTESTINAL TRACT

Pyloric Stenosis

Clinical Manifestations

Hypertrophic pyloric stenosis occurs in approximately 1 in 300 live births and is the most common surgical cause of vomiting in infancy. The typical infant is 4 to 6 weeks of age, male, and the first born child. Nonbilious vomiting, becoming increasingly projectile, occurs over several days to weeks. Eventually the infant will develop a nearly complete obstruction by the second to fourth week of life, and will not be able to hold down even clear liquids. This invariably proceeds to severe dehydration if not treated. These infants develop a metabolic alkalosis with severe depletion of potassium and chloride ions. The serum pH level is high, whereas the urine pH level is high initially but eventually drops as the severe potassium deficit leaves only hydrogen ions to exchange with sodium ions in the distal tubule of the kidney.

The diagnosis of pyloric stenosis usually can be made on physical examination by palpation of the typical “olive” in the right upper quadrant and the presence of visible gastric waves on the abdomen. When the olive cannot be palpated, US in experienced hands will diagnose the condition accurately in 95 percent of patients.

Treatment

Pyloric stenosis is not a surgical emergency. Fluid resuscitation with correction of electrolyte abnormalities and metabolic alkalosis is essential before induction of general anesthesia for operation. For most infants, I.V. fluid containing 5% dextrose and 0.45% saline with added potassium of 2 to 4meq/kg over 24h at a rate of approximately 150 to 175mL/kg for 24h will correct the underlying deficit. After this resuscitation, a Fredet-Ramstedt pyloromyotomy is performed (Fig. 37-14). Postoperatively, I.V. fluids are continued for several hours after which small frequent feedings of dilute formula or breast milk are offered and are gradually increased to full strength and full volume over the next 24 to 48h. Most infants can be discharged home within 48h following surgery.

Intestinal Obstruction in the Newborn

Bilious (green) vomiting is the most common presenting symptom in a newborn with intestinal obstruction. If the abdomen is not distended and the abdominal plain films show complete obstruction with no distal air, then the obstruction is most likely at the level of the duodenum or the pylorus. If there is incomplete obstruction with some gas more distally in the bowel, then an upper gastrointestinal series is mandated to diagnose malrotation with possible midgut volvulus. If, however, the abdomen is distended, the obstruction is most likely more distal. The physical examination will determine whether or not the anus is patent. Calcifications on the abdominal plain film may indicate meconium peritonitis; pneumatosis and/or free abdominal air indicates necrotizing enterocolitis (NEC) with or without intestinal perforation. Otherwise a contrast enema will show whether there is a microcolon indicative of jejunal-ileal atresia or meconium. If a microcolon is not present, then the diagnoses of Hirschsprung's disease, small left colon syndrome, and meconium plug will be entertained.

Duodenal Malformations

Duodenal obstruction may be complete as in duodenal atresia or partial as in duodenal web or stenosis, annular pancreas, and malrotation of the midgut. A prenatal US may show duodenal obstruction by demonstrating fluid filled cystic structures in the upper abdomen. Associated polyhydramnios is common and presents in the third trimester. In 85 percent of infants with duodenal obstruction, the entry of the bile duct is proximal to the level of obstruction, such that vomiting is bilious. Abdominal distention is not present because of the proximal level of obstruction. In those infants with obstruction proximal to the bile duct entry, the vomiting is nonbilious. Approximately one third of newborns with duodenal atresia have associated Down's syndrome (trisomy 21). The “double bubble” seen on the air contrast upper gastrointestinal series is characteristic of duodenal atresia, with small amounts of air seen distally if obstruction is incomplete (Fig. 37-15).

A naso- or orogastric tube is passed to decompress the stomach and duodenum and the infant is given I.V. fluids to maintain adequate urine output. The infant should be evaluated thoroughly for other associated anomalies, especially cardiac, since surgery is non-emergent. However, surgery should not be delayed if malrotation is the cause of duodenal obstruction or if malrotation is present with intrinsic duodenal obstruction.

Once the work-up is complete and the infant is stable, he or she is taken to the operating room and the abdomen entered through a transverse right upper quadrant supraumbilical incision under general endotracheal anesthesia. Associated anomalies found at the time of the operation include annular pancreas, malrotation, anterior portal vein, second distal web, and biliary atresia. Most surgeons will perform a duodenoduodenostomy to correct the anomaly using either a standard side-to-side or proximal transverse-to-distal longitudinal (diamond shaped) anastomosis with appropriate suture. An annular pancreas is never divided. Webs can be excised through a vertical duodenal incision, the mucosal oversewn, and the duodenotomy closed horizontally. Gastrostomy tubes are not placed routinely. Recently reported survival rates exceed 90 percent. Late complications from repair of

duodenal atresia occur in approximately 12 to 15 percent of patients and include megaduodenum, intestinal motility disorders, duodenogastric reflux, gastritis, peptic ulcer, gastroesophageal reflux, choledochal cyst, and cholelithiasis with cholecystitis.

Intestinal Atresia

Fetal mesenteric vascular accidents are likely to be the cause of most cases of intestinal atresia, based on a large amount of experimental and clinical data. The incidence of intestinal atresia has been estimated to be between 1 in 2000 to 1 in 5000 live births, with equal representation of the sexes. Clinically the infants with jejunal or ileal atresia present with bilious vomiting and progressive abdominal distention. The degree to which the abdomen becomes distended and the number of obstructed loops on upright abdominal films correlates roughly with the level of the obstruction (Fig. [37-16](#)).

In 12 percent of patients, calcification can be observed on the plain abdominal films diagnosing prenatal perforation and meconium peritonitis. Work-up of the newborn with suspected jejunal-ileal atresia includes a barium enema. The information from this procedure can distinguish small versus large intestinal distention, whether or not a microcolon is present, the level of obstruction, and the position of the cecum.

Surgical correction of the small intestinal atresia is urgent. At laparotomy, one of several types of atresia will be encountered. In type I there is a mucosal atresia with intact muscularis. In type 2 the atretic ends are connected by a fibrous band. In type 3A the two ends of the atresia are separated by a V-shaped defect in the mesentery. Type 3B is an "apple- peel" deformity or "Christmas tree" deformity in which the bowel distal to atresia receives a retrograde blood supply from the ileal colic or right colic artery. In type 4 atresia, there are multiple atresias with a "string of sausage" or "string of beads" appearance. Disparity in lumen size (Fig. [37-17](#)) between the proximal distended bowel and the small diameter of collapsed bowel distal to the atresia has lead to a number of innovative techniques of anastomosis. These include 1) end-to-back technique, fish- mouthing the antimesenteric border of the distal loop; 2) tapering of the proximal distended loop to correspond to the distal loop with end-to-end anastomosis; 3) Bishop-Koop end-to-side union with exteriorization of the distal lumen, and rarely, 4) Mikulicz exteriorization or 5) end ileostomies with delayed anastomosis (Fig. [37-18](#)).

Malrotation and Midgut Volvulus

During fetal development, the midgut supplied by the superior mesenteric artery grows too rapidly to be accommodated in the abdominal cavity. Prolapse into the umbilical cord occurs around the sixth week. Between the tenth and twelfth week, the midgut returns to the abdominal cavity, undergoing a 270 degree counterclockwise rotation around the superior mesenteric artery. The duodenum rotates caudal to the artery, and its C- loop traces this path. The transverse and ascending colon demonstrate the path of rotation of the cecum cephalad to the artery. The duodenum becomes fixed retroperitoneally in its third portion, emerging at the ligament of Treitz, and the cecum becomes fixed to the lateral abdominal wall by peritoneal bands. The takeoff of the branches of the superior mesenteric artery elongates and becomes fixed along a line extending from its emergence from the aorta to the cecum in the right lower quadrant. If rotation is incomplete, the cecum remains in the epigastrium, but the bands fixing the duodenum to the retroperitoneum and cecum continue to form. This results in bands (Ladd's bands) extending from the cecum to the lateral abdominal wall and crossing the duodenum with the potential for obstruction. The mesenteric takeoff remains confined to the epigastrium, resulting in a narrow pedicle suspending all the branches of the superior mesenteric artery and the entire midgut. A volvulus may occur around the mesentery, obstructing the jejunum and also cutting off the blood supply to the midgut. Intestinal obstruction and the potential for total vascular compromise of the midgut supervene unless the condition is corrected.

Midgut volvulus can occur at any age, though it is seen most often in the first few weeks of life. Bilious vomiting is usually the first sign of volvulus and all infants with bilious vomiting must be evaluated rapidly to insure that they do not have intestinal malrotation with volvulus. Chronic obstructive symptoms from Ladd's bands across the duodenum may manifest as intermittent vomiting, but a volvulus presents early with irritability and bile vomitus and later with manifestations of vascular compromise such as bloody stools and eventually circulatory collapse. The index of suspicion for this condition must be high, since abdominal signs are minimal in the early stages. Advanced ischemia of the intestine results in erythema and edema of the abdominal wall, shock and rapidly, death. Abdominal films show a paucity of gas throughout the intestine with a few scattered air-fluid levels. The diagnosis of malrotation is best made by an upper gastrointestinal series that shows incomplete rotation with the duodenojejunal junction displaced to the right. The duodenum may show a corkscrew effect diagnosing volvulus, or complete duodenal obstruction, with the small bowel loops entirely in the right side of the abdomen. Barium enema may show a displaced cecum, but this sign is unreliable, especially in the small infant in whom the cecum is normally in a somewhat higher position than in the older child.

Early surgical intervention is mandatory if the ischemic process is to be avoided or reversed. Volvulus occurs clockwise and it is therefore untwisted counterclockwise. The bands between the cecum and the abdominal wall and between the duodenum and terminal ileum are divided sharply to splay out the superior mesenteric artery and its branches. This brings the straightened duodenum into the right lower quadrant and the cecum into the left lower quadrant. The appendix is removed to avoid diagnostic errors in later life (Fig. [37-19](#)). If the bands have been lysed completely no suturing of the cecum or duodenum is necessary. With advanced ischemia, reduction of the volvulus without the Ladd procedure is accomplished, and a "second look" 24 hours to 36h later often will show vascular recovery. Frankly necrotic bowel can then be resected conservatively. With early diagnosis and correction the prognosis is excellent. With delays in diagnosis, the infant may succumb or be condemned to the ravages of shortcut syndrome until intestinal transplantation becomes routinely possible.

Meconium Ileus

Meconium ileus produces intestinal obstruction from impaction of meconium in the distal ileum. The ideology of this problem is cystic fibrosis. These infants have deficits of pancreatic enzymes in the intestine producing viscous meconium. Bilious vomiting appears late, following progressive abdominal distention, failure to pass meconium, and impressively dilated loops of small bowel on upright abdominal film without air fluid levels. Small bubbles of gas entrapped in the inspissated meconium in the terminal ileum may produce a characteristic "ground glass" appearance. A contrast enema will demonstrate a microcolon and the terminal ileum filled with pellets of meconium (Fig. [37-20](#)). In patients with prenatal intestinal perforation, intraperitoneal calcifications form, producing an eggshell pattern on plain abdominal x-ray. Patients with uncomplicated meconium-ileus, i.e. patients without intestinal perforation or vascular compromise of the distended ileum, can be treated nonoperatively by a method originally described by Noblett.

Dilute Gastrografin is advanced through the colon under fluoroscopic control into the dilated portion of the ileum. Since Gastrografin acts partially by absorbing fluid from the bowel wall into the intestinal lumen, maintaining adequate hydration of the infant during this maneuver is extremely important. The enema may be repeated at 12-h intervals over several days until all the meconium is evacuated per rectum. If surgical intervention is required because of failure of Gastrografin enemas to relieve obstruction, operative irrigation with dilute Gastrografin or acetylcysteine (Mucomyst) through a purse-string suture may be successful. Alternatively, resection of the distended terminal ileum is performed, the meconium pellets are flushed from the distal small bowel by acetylcysteine, and ileostomies, a Bishop-Koop anastomosis, or an end-to-end anastomosis is performed.

Necrotizing Enterocolitis Clinical Manifestations

Necrotizing enterocolitis (NEC) is a disease of multifactorial origin that almost exclusively affects stressed premature infants. Common factors found in most infants with NEC include intestinal ischemia, bacterial colonization of the gut, and enteral feedings of synthetic formulas. Hypoalbuminemia may predispose neonates to NEC. Experimental evidence suggests that both ischemia and reperfusion may lead to the intestinal damage seen in NEC. Clinically the first sign of NEC is formula intolerance, evident by vomiting or a large residual volume of a previous feeding in the stomach at the time of the next feeding. Abdominal distention and hematochezia are the next signs of NEC and suggest ileus and mucosal ischemia, respectively. Invasion of the ischemic mucosa by gas-forming organisms causes pneumatosis intestinalis, which is the pathognomonic radiographic finding in NEC (Fig. [37-21](#)). Development of hepatoportal venous gas suggests a particularly severe form of NEC. Some children, despite appropriate medical therapy, will develop a fulminant course with progressive peritonitis, acidosis, sepsis, disseminated intravascular coagulopathy, and death.

Treatment

In all infants suspected of having NEC, feedings are discontinued, a nasogastric tube is placed, and parenteral antibiotics are given. Free intraperitoneal air and signs of diffuse peritonitis are obvious indications for operation. Children with significant acidosis (pH < 7.20) after volume resuscitation and those with hepatoportal venous gas who do not improve promptly on medical therapy are also considered surgical candidates. In equivocal cases, tapping the peritoneal cavity may aid in the decision to operate if the fluid withdrawn contains bacteria or intestinal contents. Resection of frankly gangrenous bowel should be carried out, and in the vast majority of cases, the intestinal ends are brought out as stomas. When there is massive intestinal involvement, marginally viable bowel is retained and a "second-look" procedure is carried out in 24h. In the selected neonate weighing less than 1500 grams, drainage of the peritoneal cavity under local anesthesia is an effective method of improving the desperately ill infant, with approximately one-third surviving without additional operations.

Total parenteral nutrition is maintained for at least 2 weeks postoperatively, after which oral feedings of small volumes of dilute formula are gradually introduced. Strictures develop in 20 percent of medically or surgically treated patients, and a contrast enema is mandatory before reestablishing intestinal continuity. If all other factors are favorable, the ileostomy is closed when the child is between 2 and 2.5 kg. Despite the severity of the illness in this premature population, survival rates for NEC approach 80 percent in many series.

Intussusception

Intussusception is a common cause of intestinal obstruction in the infant. It is observed most often in infants between 8 and 12 months of age and is slightly more common in males. The most common cause is hypertrophy of the Peyer's patches in the terminal ileum from an antecedent viral infection. The hypertrophied lymphatic patch becomes drawn into the lumen of the terminal ileum and is moved progressively into the ascending and transverse colon. Polyps, malignant tumors, such as lymphoma, and Meckel's diverticulum may act as lead points for intussusception; such intussusceptions are rarely reduced by air or contrast enema, and thus the lead point is identified when operative reduction of the intussusception is performed.

Clinical Manifestations

Since intussusception is frequently preceded by a gastrointestinal viral illness, the onset may not be determined easily. Typically, the infant develops paroxysms of crampy abdominal pain and intermittent vomiting. Between attacks the infant may act completely well, but as symptoms progress, increasing lethargy develops. Bloody mucus ("currant-jelly" stool) may be passed per rectum. Ultimately, if reduction is not accomplished, gangrene of the intussusceptum occurs.

The pathognomonic physical finding is an elongated mass in the right upper quadrant or epigastrium with an absence of bowel in the right lower quadrant (Dance's sign). The mass may be seen on plain abdominal x-ray but is more easily demonstrated on air or contrast enema.

Treatment

The air enema is the preferred method of diagnosis and nonoperative treatment of intussusception. Air is introduced with a manometer and pressure should not exceed 120 mmHg. Peritonitis constitutes a contraindication to air enema. The air may push the intussusception before it and accomplish reduction. Free reflux into multiple small bowel loops and an infant who abruptly becomes well are characteristic of successful reduction. Unless both these signs are observed, it cannot be assumed that the intussusception is reduced, and preparations for exploration are made. Despite reports of successful reduction in more than 90 percent of children with intussusception using air enema, most centers are achieving about 60 percent successful reduction with this technique.

If hydrostatic reduction is successful, the infant may be given oral fluids immediately. The incidence of recurrent intussusception is 5 percent whether the intussusception is reduced by operation or by hydrostatic pressure. Failure to reduce the intussusception mandates surgery. Exploration is carried out through a right lower quadrant incision, delivering the intussuscepted mass into the wound. Reduction usually can be accomplished by gentle distal pressure, milking the bowel out of the intussusciens, never pulling it out. The blood supply to the appendix is often compromised, and appendectomy is performed. Resection of frankly gangrenous bowel is carried out without attempting reduction of the intussusception. As a rule, primary ileocolic anastomosis can be performed after resection.

I.V. fluids are continued until peristalsis returns. If resection is necessary, prophylactic antibiotics are also administered for 72 h.

Appendicitis

Clinical Manifestations

Appendicitis in children may differ in presentation from the same disease in adults. The incidence of perforation rises dramatically as the age of the patient decreases so that perforated appendicitis occurs in approximately 25 percent of adolescents but 80 percent of children less than 5 years old. The frequency of gastroenteritis and the poor localization of pain in the young child both increase the risk of ascribing the pain to a viral etiology, delaying a visit to the doctor and once seen by the doctor, the low index of suspicion that the problem is appendicitis further delays the correct diagnosis, increasing the risk of perforation.

Once the appendix has perforated, the symptoms may be most consistent with severe gastroenteritis with pain, vomiting and diarrhea, or the child may present with a pattern of intestinal obstruction. An abdominal mass in the lower abdomen, in an ill-appearing child may well be an appendiceal abscess rather than a malignancy.

Sonography has been disappointing in assisting with the diagnosis of appendicitis. Contrast enema and abdominal CT with enteric contrast material are the best radiographic studies to diagnose late appendicitis in the unclear case. In the male, right lower quadrant peritoneal tenderness and leukocytosis should prompt exploration regardless of other signs and symptoms. In the female, ovarian or uterine pathology also must be considered. In girls over 10 years of age, laparoscopy is very helpful in making the correct diagnosis of low abdominal pain and is almost always a successful method of treating that problem. Management of an appendiceal abscess “mass” by drainage percutaneously and I.V. antibiotics has been advocated by some, but most pediatric surgeons prefer early operative drainage of the abscess and appendectomy.

Duplications, Meckel's Diverticulum, and Mesenteric Cysts

Duplications

Duplications can occur at any level in the gastrointestinal tract but are found most commonly in the ileum. They may be long and tubular, but usually are cystic masses lying within the leaves of the mesentery and sharing a common wall with the intestine. Symptoms can include a palpable mass and/or frank intestinal obstruction. Torsion may produce gangrene and perforation, and subtle or massive bleeding may occur. This bleeding comes from ulceration in the duplication or adjacent intestine if the duplication contains ectopic gastric mucosa.

The ability to make a preoperative diagnosis of duplication usually depends on the presentation. Sonography and technetium pertechnetate scanning are the two most helpful diagnostic tests. In the case of short duplications, resection of the cyst and adjacent intestine with end-to-end anastomosis can be performed in a straightforward fashion. If resection of long duplications would compromise intestinal length, multiple enterotomies and mucosal stripping in the duplicated segment will allow the walls to collapse and become adherent. An alternative method is to divide the common wall using the GIA stapler, forming a common lumen. This should not be done in duplications that contain ectopic gastric mucosa. In the patient with a very long duplication or multiple duplications, intraoperative nuclear medicine scanning will ensure complete excision of the ectopic gastric mucosa.

Meckel's Diverticulum

A Meckel's diverticulum is a remnant of a portion of the embryonic omphalomesenteric (vitelline) duct. It is located on the antimesenteric border of the ileum, usually within 2 ft of the ileocecal valve. It may be found incidentally at surgery or may present with inflammation masquerading as appendicitis. Similar to duplications, ectopic gastric mucosa may produce ileal ulcerations that bleed and lead to the passage of maroon-colored stools. Diagnosis may be made by technetium pertechnetate scans when the patient presents with bleeding. Treatment is surgical, and a wedge resection of the diverticulum with transverse closure of the ileum can be performed if the base is narrow. If wedge excision would compromise the ileal lumen, sleeve excision and end-to-end ileoileostomy are performed.

Mesenteric Cysts

Mesenteric cysts are similar to duplications in their location within the mesentery. However, they do not contain any mucosa or muscular wall. Chylous cysts may result from congenital lymphatic obstruction (Fig. 37- 22). Mesenteric cysts can cause intestinal obstruction or may present as an abdominal mass. The diagnosis may be made by US or abdominal CT. At operation, surgical excision is preferred sometimes requiring resection of the adjacent intestine. Partial excision or marsupialization should be reserved for cysts involving a large portion of the mesentery.

Hirschsprung's Disease

Hirschsprung's disease results from the absence of ganglion cells in the myenteric plexus of the intestine. Recent findings suggest that mutations in the RET protooncogene may be responsible for many cases of Hirschsprung's disease. The precursors of the ganglion cells are neural crest cells that migrate into the intestine from cephalad to caudad. The process is completed by the twelfth week of embryonic life, but the migration from midtransverse colon to anus takes 4 weeks. This increases the time period of vulnerability for failure of migration and accounts for the fact that most cases of aganglionosis involve the rectum and rectosigmoid. Longer segments of absent ganglion cells also may occur, and total colonic aganglionosis, although rare, is also seen.

Aganglionic colon does not permit normal peristalsis to occur. Functional obstruction therefore supervenes, and the infant may present with complete colon obstruction or with a devastating enterocolitis. The presentation may, however, be much more subtle, with constipation and abdominal distention as the chief findings. Occasionally failure to thrive is the initial sign.

Diagnosis

Infants with Hirschsprung's disease usually will fail to pass meconium in the first 24 h of life, although this history is often difficult to obtain. Barium enema may be unreliable in diagnosing Hirschsprung's disease in the newborn infant because the colon is not dilated enough to show a transition zone. In older infants and children, barium enema will show the size difference between the dilated ganglionic colon and the distal constricted aganglionic rectal segment. The barium enema in total colonic aganglionic usually shows a markedly shortened colon.

Rectal biopsy makes the definitive diagnosis of Hirschsprung's disease. Suction rectal biopsy provides a small piece of mucosa and submucosa without the requirement for anesthesia. Occasionally the suction biopsy is not diagnostic and a full-thickness biopsy is required. The histopathology of Hirschsprung's disease is the absence of ganglion cells in the myenteric plexuses, increased staining of a cholinesterase stain and the presence of hypertrophied nerve bundles. Some surgeons have found the use of rectal manometry helpful, particularly in older children, but it is not as accurate a diagnostic tool as is rectal biopsy.

Treatment

Treatment is surgical in all cases. Some pediatric surgeons prefer to create a colostomy in the newborn period and wait until the child weighs 10 kg before performing a definitive pull-through operation. When a colostomy is done for Hirschsprung's disease, a frozen-section confirmation that ganglion cells are present at the colostomy site is mandatory. Some surgeons advocate primary pull-through operations in newborn infants without colostomy. The intraabdominal dissection can be performed using the laparoscope, even in the newborn. In the older infant and child who have been diagnosed belatedly, it is important to allow the distended hypertrophied colon to return to a normal size before performing a pullthrough. This is usually accomplished by waiting 3 to 6 months after a colostomy is performed.

Three pull-through procedures are currently in use for treating Hirschsprung's disease. The first of these is the original Swenson procedure, in which the aganglionic rectum is carefully dissected in the pelvis and removed down to the anus. The ganglionic colon is then anastomosed to the anus via a perineal approach (Fig. 37-23). In the Duhamel procedure (see Fig. 37-23), dissection outside the rectum is confined to the retrorectal space, and the ganglionic colon is anastomosed posteriorly just above the anus. The anterior wall of the ganglionic colon and the posterior wall of the aganglionic rectum are anastomosed, using a stapling device. In Soave's operation, dissection is entirely within the rectum (Fig. 37-24). The rectal mucosa is stripped from the muscular sleeve, and the ganglionic colon is brought through this sleeve and anastomosed to the anus. Complications with all procedures include enterocolitis, constipation and anastomotic stricture, but long-term results with the three procedures are comparable and generally excellent in experienced hands. These three procedures also can be adapted for total colonic aganglionosis; the ileum is used for the pull-through or anastomosed to the aganglionic segment of distal colon and rectum to improve absorption.

Imperforate Anus

Imperforate anus occurs in approximately 1 in 5000 live births, affecting males and females almost equally. Failure of descent of the urorectal septum in embryonic life produces a variety of anorectal and cloacal anomalies. The level to which this septum descends determines the separation of the urinary and hindgut systems. Broadly classified, the imperforate anus is characterized as “high” or “low” depending on whether the rectum ends above the levator ani muscle complex or partially descends through this muscle. In high imperforate anus in males the rectum usually ends as a fistula into the membranous urethra. In females, the high imperforate anus often is part of a cloacal anomaly.

The low lesions have a fistula to the perineum. In males this is seen in the median raphe of the scrotum or penis, and in females the most common perineal fistula ends at the posterior fourchette. Since the rectum has descended through the levator complex in low lesions, only a perineal operation is required; this situation occurs in 40 percent of males and 70 percent of females with imperforate anus. Such children will be expected to be continent, since the “muscle of continence,” the levator ani muscle, and the rectum are in a normal relationship to each other.

Infants having high imperforate anus require a colostomy in the newborn period, with some kind of pull-through procedure performed at about 2 months of age. Careful assessment of the genital tracts in females and the genitourinary system in all patients with imperforate anus is imperative because of the very high incidence of associated anomalies, particularly in children with a high imperforate anus. Many have associated spinal cord anomalies, particularly tethered cord, and all should be evaluated with spinal US and magnetic resonance imaging MRI. These infants require evaluation for signs of the VACTERRL association (see [Esophageal Atresia and Tracheoesophageal Fistula](#), above), including careful preoperative cardiac assessment.

Diagnosis of imperforate anus is not difficult. The location of the fistula site may sometimes be a problem. Beading of mucus or meconium along the median raphe of the perineum and scrotum denotes a low imperforate anus in the male. Air in the bladder, urinating gas or meconium, and a retrograde urethrogram with contrast material may demonstrate the

urinary fistula of the high imperforate anus in the male. In the female, since most of the lesions are low, careful examination of the perineum, fourchette, and vestibule of the vagina will locate the fistula in most cases. A high fistula in the female may terminate in the vagina and will be harder to demonstrate. If a single perineal opening is seen in the female, a cloacal abnormality is present with urethra, vagina, and rectum opening into a common urogenital sinus. Colostomy is usually delayed for 24h if no perineal fistula is seen. Cross-table lateral x-ray of the perineum with the buttock elevated may be done to determine the distance between the end of the rectum and the skin. If there is any doubt of the level, it is much safer to perform a colostomy than ruin any chance for continence by an injudicious perineal operation.

Pena and DeVries have devised a posterior sagittal approach for repair of high imperforate anus, dividing the levator ani and external sphincter complex in the midline posteriorly and bringing down the rectum after sufficient length is achieved. The muscles are then reconstructed and sutured to the rectum. Initial reports of results with the posterior sagittal anorectoplasty have been encouraging for both the primary and secondary procedures for high imperforate anus as well as cloacal anomalies. Pena also recommends this approach in the female with a rectovestibular fistula.

JAUNDICE

Biliary Atresia

Pathologic jaundice implies persistent elevation of the direct bilirubin level (> 2 weeks, > 2 mg/dL) and may be consequent to cholestasis or obstruction. The incidence of biliary atresia is approximately 1 in 20,000 births, and the condition affects not only the extrahepatic bile ducts but also the liver, which accounts for the biochemical overlap with many of the cholestatic causes of jaundice. Originally, biliary atresia was thought to be a developmental anomaly. There is recent suggestive evidence, however, that the condition may have an immunologic basis. Infectious etiologies have been proposed but never substantiated.

The obliterative process involves the common duct, cystic duct, one or both hepatic ducts, and the gallbladder, in a variety of combinations. Approximately 25 percent of the patients have coincidental malformations, often associated with polysplenia and may include intestinal malrotation, preduodenal portal vein, and intrahepatic vena cava.

Abdominal exploration in an infant with biliary atresia should commence with inspection of the left upper quadrant to determine the status of the spleen and thereby alert the surgeon to the possibility of coexisting anomalies should polysplenia be encountered.

Clinical Manifestations

Jaundice, a constant finding, is usually present at birth or shortly thereafter but may go undetected or may be regarded as “physiologic” until the child is 2 or 3 weeks old.

Initially, the growth and weight gain are normal, but in later stages retarded growth is apparent. As the liver becomes progressively more fibrotic, the patients develop stigmata of portal hypertension, particularly splenomegaly.

Diagnosis

No single test or combination of studies is absolutely reliable but with properly selected tests, it is usually possible to discriminate between cholestatic and obstructive jaundice. Initially, the synthetic functions of the liver are unimpaired. Thus, the serum albumin level and clotting mechanism are normal. In many centers the nuclear scan using technetium 99m IDA (DISIDA), performed after pretreatment of the patient with phenobarbital, has proven to be an accurate and reliable study. If radionuclide appears in the intestine, extrahepatic bile duct patency is ensured. If radiopharmaceutical is normally concentrated by the liver but not excreted despite treatment with phenobarbital, and the metabolic screen, particularly alpha-1-antitrypsin determination, is normal, the presumptive diagnosis is biliary atresia. This examination, particularly when complemented by percutaneous needle biopsy of the liver, will establish a diagnosis with a high degree of certainty.

Complementary studies include abdominal US. This examination is reliable only when performed in the fasting patient. It should be emphasized that the presence of a gallbladder does not exclude the diagnosis of biliary atresia; in approximately 15 percent of these patients the distal biliary tract is patent although the proximal ducts are atretic. It is worth noting that the intrahepatic bile ducts are never dilated in the patient with biliary atresia. The sonographer must be alert to the possibility that parallel vascular structures in the porta may be mistakenly interpreted as patent bile ducts. Doppler US will usually resolve confusion.

Neonatal jaundice also may be caused by the *inspissated bile syndrome*. This term is applied to patients with normal biliary tracts who have persistent obstructive jaundice. Increased viscosity of bile and obstruction of the canaliculi are implicated as causes. The condition has been seen in infants receiving parenteral nutrition and is also encountered secondary to hemolysis or cystic fibrosis. In some instances, no etiologic factors can be defined.

Treatment

Historically, a variety of innovative operations have been attempted to promote bile flow from the liver in infants with biliary atresia. Only the hepatoportoenterostomy, developed by Kasai, has stood the test of time and demonstrated promise as a procedure that not only will provide relief of jaundice but also may result in cure. The procedure is based on Kasai's observation that the fibrous tissue at the porta hepatis invests microscopically patent biliary ductules that, in turn, communicate with the intrahepatic ductal system. Transecting this fibrous tissue, invariably encountered cephalad to the bifurcating portal vein, opens these channels and establishes bile flow into a surgically constructed intestinal conduit, usually a Roux en Y limb of jejunum (Fig. 37-25). The likelihood of surgical success is increased if the procedure is accomplished before the infant attains the age of 8 weeks. The outlook is poor for patients after the twelfth week. Bile drainage is anticipated when the operation is carried out early; unfortunately, bile flow does not necessarily imply cure. Prognostic indicators include the rate at which the serum bilirubin level returns toward normal. It has been observed that in patients operated upon before 60 days, jaundice resolves in approximately 2 months, whereas infants having corrective surgery after 60 days rarely become anicteric or do so only after 6 months. Further, in the latter group the transaminases often remain elevated. The outcome is most favorable when the serum bilirubin reaches a level less than 1 mg./dL. Hepatic fibrosis may progress even when bile drainage is achieved. Postoperative cholangitis continues to plague infants having successful portoenterostomy procedures. The incidence and severity of cholangitis may be reduced by external venting of the surgically created biliary conduit and by the long-term administration of antibiotics. Despite these measures, the problem of ascending infection has not been eliminated. Reoperation, with resectional debridement of the portoenterostomy anastomosis, has rescued selected children with intractable cholangitis. Portal hypertension and its sequelae also jeopardize the outcome for some patients even when they have been relieved of jaundice. Gastrointestinal hemorrhage from esophageal varices usually can be managed successfully by endoscopic sclerotherapy or variceal ligation. For some patients portosystemic shunting has been successful in controlling hemorrhage from esophageal varices.

While “cure” may be attained in only 25 to 30 percent of patients after hepatoportoenterostomy, significant palliation and prolongation of life with the native liver has been achieved routinely with the probability of 5 year survival exceeding 60 percent. For the child afflicted with biliary atresia in whom the Kasai operation fails, or succeeds only temporarily, liver transplantation remains the final hope for salvage. It is generally acknowledged that a prior Kasai operation does not adversely affect on the results of subsequent transplantation. In the era of orthotopic and living related transplantation there is little justification for repeated operations attempting to salvage a failed Kasai operation. Biliary atresia has been the most common indication for transplantation in pediatric recipients, and the 1- year survival rate exceeds 80 percent.

Biliary Hypoplasia

In some infants coming to surgery with a presumptive diagnosis of biliary atresia, an operative cholangiogram may demonstrate the gallbladder and extrahepatic biliary system to be patent, albeit diminutive hence, *biliary hypoplasia*. Hypoplasia of the extrahepatic biliary system is associated with hepatic parenchymal disorders that cause severe intrahepatic cholestasis. Included among these are alpha 1-antitrypsin deficiency and arteriohepatic dysplasia (Alagille's syndrome). The primary pathology resides within the liver and not the bile ducts; therefore, portal dissection and portoenterostomy are not indicated in these patients. Rather, a generous liver biopsy is obtained and the operation terminated.

Choledochal Cyst

There have been numerous descriptions and classifications based on the location and anatomy of the choledochal cyst. Among the most useful is the classification proposed by Alonso-Lej. The type I cyst characterized by fusiform dilatation of the bile duct into which the cystic duct enters is the most common and clinically relevant. Choledochal cyst is most appropriately considered the predominant feature in a constellation of pathologic abnormalities within the pancreatic-biliary system. Frequently associated with choledochal cyst are anomalous junction of the pancreatic duct and common bile duct, distal bile duct stenosis, intrahepatic ductal dilatation, abnormal histology of the common bile duct, and hepatic histology ranging from normal to cirrhotic. These features are encountered in varying degrees and combinations and constitute the anatomic spectrum of the malformation.

The causation of choledochal cyst is controversial. Babbitt proposed an abnormal pancreatic biliary duct junction, with the formation of a “common channel” into which pancreatic enzyme secretions are discharged, with resulting weakening of the bile duct wall by gradual enzymatic destruction, leading to dilatation, inflammation, and finally cyst formation. It should be noted, however, that not all patients with choledochal cyst demonstrate an anatomic common channel.

Choledochal cyst is more common in females than in males (4:1). The so- called classic symptom complex of pain, mass, and jaundice is actually encountered in fewer than half the patients. The more usual presentation is that of episodic abdominal pain, often recurrent over months or years, and generally associated with only minimal jaundice that may escape detection. If the condition persists unrecognized, sequelae including cholangitis, cirrhosis, and portal hypertension are almost inevitable.

Diagnosis

With the increasing use of antenatal US, choledochal cyst is frequently diagnosed in the fetus. For the older child or adolescent, US and CT scanning are the diagnostic mainstays. These studies will demonstrate the dimensions of the cyst and define its relationship to the vascular structures in the porta, as well as the intrahepatic ductal configuration. Endoscopic retrograde cholangiopancreatography (ERCP) is reserved for patients in whom confusion remains after evaluation by less invasive imaging modalities.

Treatment

The surgical options include internal drainage by cystenterostomy and surgical excision. The morbidity from the former is excessive. The cyst wall is composed of fibrous tissue and is devoid of mucosal lining. Anastomotic obstruction from scarring is inevitable. Further, the thick-walled fibrous cyst does not contract after drainage but rather persists as a receptacle for stagnant bile, and complications related to biliary stasis militate against internal drainage procedures. An additional, and perhaps the most serious, consequence of cyst retention is the development of malignancy arising within the cyst wall. The incidence of this aggressive and highly lethal neoplasm further supports the recommendation for cyst resection.

In most circumstances, the caliber of the common hepatic duct cephalad to the choledochal cyst is normal. Successful resection of the cyst requires circumferential dissection, entering the posterior plane between the cyst and portal vein to accomplish removal. The pancreatic duct, which may enter the distal cyst, is vulnerable to injury during distal cyst excision but can be avoided by not dissecting into the pancreatic parenchyma.

For the patient in whom the anatomy of the porta is obscured and distorted by pericystic inflammation, an alternative technique is proposed. An arbitrary plane is entered within the posterior wall of the cyst that allows the inner lining of the back wall to be dissected free from the outer layer that directly overlies the portal vascular structures. The lateral and anterior cyst, as well as the internal aspect of the back wall, is removed. The outer posterior wall remains behind. In either circumstance, cyst excision is accomplished, with reconstruction using normal or near-normal proximal bile duct and a mucosal union between the biliary system and intestinal tract. The likelihood of postoperative anastomotic stricture with attendant sequelae is minimized. Further, the risk of malignant tumor is reduced although not completely eliminated.

DEFORMITIES OF THE ABDOMINAL WALL

Embryology

The abdominal wall is formed by four separate embryologic folds—cephalic, caudal, and right and left lateral folds—each of which is composed of somatic and splanchnic layers. Each of the folds develops toward the anterior center portion of the coelomic cavity, joining to form a large umbilical ring that surrounds the two umbilical arteries, the vein, and the yolk sac or omphalomesenteric duct. These structures are covered by an outer layer of amnion, and the entire unit composes the umbilical cord. Between the fifth and tenth weeks of fetal development the intestinal tract undergoes a rapid growth outside the abdominal cavity within the proximal portion of the umbilical cord. As development is completed, the intestine gradually returns to the abdominal cavity. Contraction of the umbilical ring completes the process of abdominal wall formation. Failure of the cephalic fold to close results in sternal defects (such as congenital absence of the sternum or the pentalogy of Cantrell. Failure of the caudal fold to close results in exstrophy of the bladder and, in more extreme cases, exstrophy of the cloaca. Interruption of central migration of the lateral folds results in omphalocele. Gastroschisis, originally thought to be a variant of omphalocele, probably results from a fetal accident in the form of intrauterine rupture of a hernia of the umbilical cord.

Umbilical Hernia

Failure of timely closure of the umbilical ring leaves a central defect in the linea alba. The resulting umbilical hernia is covered by normal umbilical skin and subcutaneous tissue, but the fascial defect allows protrusion of abdominal content (Fig. 37-26). Hernias less than a cm in size at the time of birth usually will close spontaneously by three years of life. Sometimes the hernia is large enough that the protrusion is disfiguring and disturbing to both the child and the family. In such circumstances early repair may be advisable. When the defect is small and spontaneous closure likely, delay of surgical correction until 4 or 5 years of age is appropriate. Incarceration is rarely seen in an umbilical hernia. Unlike treatment for inguinal hernia of infants and young children, attempts at reduction of an incarcerated umbilical hernia are unwise. Repair of uncomplicated umbilical hernia is performed through a small curving infraumbilical incision that fits into the skin crease of the umbilicus. The fascial defect is repaired with permanent or long-lasting absorbable, interrupted sutures in a transverse plane. Fascial flaps or other complicated umbilical hernia repairs that have been recommended for adult patients are unnecessary in children. The umbilicus should never be excised in the repair of umbilical hernias in the childhood period.

Patent Urachus

During the development of the coelomic cavity, there is free communication between the urinary bladder and the abdominal wall through the urachus, which exits adjacent the omphalomesenteric duct. Persistence of this tract results in a communication between the bladder and the umbilicus (Fig. 37-27). The first sign of a patent urachus is moisture or urine flow from the umbilicus; consequently, recurrent urinary tract infection can result. The urachus may be partly obliterated, with a remnant remaining beneath the umbilicus in the extraperitoneal position as an isolated cyst, which may be identified by US. Such a cyst usually presents as an inflammatory mass inferior to the umbilicus. Initial treatment is drainage of the abscess with cyst excision as a separate procedure once the inflammation has resolved. Diagnosis of patent urachus is most reliably made by a cystogram in the lateral projection. Surgical correction is carried out via extraperitoneal exposure of the infraumbilical area. After identification, excision of the urachal tract with closure of the bladder is curative.

Omphalocele

Prenatal US will identify many of these defects permitting genetic evaluation and counselling. Postnatally, an omphalocele (Fig. 37-28) presents as a mass of bowel and solid viscera in the central abdomen, covered by translucent membrane. The size varies from about 1 cm in diameter to huge defects containing much or all of the abdominal viscera. In the latter forms, the bowel has lost its right of domain in the abdominal cavity, and closure is difficult. Frequently the liver occupies much of the defect and does not tolerate compression. The diagnosis is made by inspection. Infants born with omphaloceles are prone to other anomalies. There is a 60 to 70 percent incidence of associated anomalies, especially cardiac and chromosomal abnormalities. Special syndromes such as exstrophy of the cloaca (vesicointestinal fissure) and the Beckwith-Wiedemann constellation of anomalies (macroglossia, macrosomia, hypoglycemia, and visceromegaly) include omphalocele.

Immediate treatment after delivery of an infant with omphalocele consists of covering the lesion with saline-soaked gauze and wrapping the trunk circumferentially. No pressure should be placed on the omphalocele sac in an effort to reduce its contents, because pressure can lead to rupture of the sac or may interfere with abdominal venous return, or impede respiratory effort. The omphalocele sac can be treated with desiccating substances such as Mercurochrome, but most infants in the United States are treated surgically. Indications for nonoperative therapy of an omphalocele are (1) a newborn with a giant omphalocele and other life-threatening anomalies whose correction takes precedence over repair of the omphalocele, (2) the neonate with other anomalies that complicate a surgical repair of the omphalocele, and (3) a newborn with severe associated anomalies that may not be consistent with survival.

When the abdomen cannot be closed primarily without undue tension, a temporary measure is placement of a “silo” of Silastic material sutured to the fascia around the circumference of the defect as an exterior cover without skin. This technique also can be used for gastroschisis when necessary. The main principle in the use of the plastic material is steady pressure with gradual reduction of the plastic envelope within 7 days to close the fascia without excessive tension. Today most infants, even those with large omphaloceles, can be salvaged. The persistent mortality rate of 20 to 30 percent from pediatric centers reflects the serious cardiac and chromosomal anomalies associated with this malformation.

Gastroschisis

Shaw first suggested that gastroschisis is simply a hernia of the umbilical cord that ruptures after the complete development of the abdominal wall. The defect that permits escape of the intestines is located at the junction of the umbilicus and normal skin. The defect is almost always to the right of the umbilicus. Infants with gastroschisis have a large amount of intestine on the surface of the abdominal wall (Fig. 37-29). The umbilicus becomes partly detached, allowing free communication with the abdominal cavity. The intestine lying free outside of the abdominal cavity may be glistening, moist, and normal in appearance, suggesting that the rupture occurred immediately before or during delivery of the infant. More commonly, the intestine is thick, edematous, discolored, and covered with an exudate. Unlike infants born with omphalocele, associated anomalies seen with gastroschisis are confined to bowel atresia, further substantiating the mechanics of this abnormality. Currently, prenatal US may detect gastroschisis, permitting counselling and elective controlled delivery (Fig. 37-30).

Treatment

All infants born with gastroschisis require urgent surgical treatment. The intestine can be returned to the abdominal cavity, and a secure, primary surgical closure of the abdominal wall is possible in many instances. Mechanical stretching of the abdominal wall aids in successful primary closure; however, for those infants whose intestine has become thickened and edematous, the construction of an extraabdominal compartment from Silastic sheeting is beneficial (Fig. 37-31). As in the surgical correction of omphalocele, the prosthetic pouch allows gradual enlargement of the abdominal cavity and accommodation of the intestines. Care must be taken so that increased abdominal pressure does not cause inferior vena cava compression that may impede venous return to the heart or diaphragmatic pressure preventing normal respiratory excursion. Intestinal function may not return for several weeks, and is especially delayed if the bowel is thickened and edematous. In these infants, the advent of I.V. alimentation has been lifesaving. The survival rate for infants with gastroschisis exceeds 90 percent in many series.

Exstrophy of the Cloaca (Vesicointestinal Fissure)

Exstrophy of the cloaca represents one of the severest forms of embryologic derangement. In infants with cloacal exstrophy, the normal ventral closure of the pelvis and the abdominal wall is imperfect. Major components of cloacal exstrophy are (1) omphalocele, (2) exstrophy of the bladder, (3) external intestinal fistula through the bladder (enterovesical fistula), (4) epispadias in males, (5) imperforate anus, and (6) foreshortened colon. In addition, there is often an associated orthopedic deformity of the lower limbs.

Treatment of these complicated anomalies should be done in conjunction with a pediatric urologist and pediatric orthopedist to gain maximal function with a minimal number of procedures. An ileostomy and primary closure of the omphalocele usually are possible. The small piece of distal colon should be preserved for future anal and vaginal reconstruction. Timing of bladder and urinary reconstruction needs to be individualized, depending on the child's anatomy. While most patients suffer a number of physical limitations, many have the potential for functional rehabilitation, which justifies aggressive surgical efforts on their behalf.

Prune-Belly Syndrome (Eagle-Barret Syndrome)

Prune-belly syndrome is characterized by the findings of lax lower abdominal musculature, dilated urinary tract, including the bladder, and undescended testes bilaterally. It occurs almost exclusively in males (Fig. 37-32). Despite the impressive dilatation of the urinary tract, most of the children with prune-belly syndrome have adequate renal parenchyma for growth and development, at least early in life. In the past, aggressive “reconstruction” of the urinary tract was attempted, but currently a conservative approach is favored unless there is a specific area that requires drainage due to obstruction, not just dilation.

The testes are invariably intraabdominal and bilateral orchidopexy can be performed in conjunction with abdominal wall reconstruction at 6 to 12 months of age. Despite orchiopexy, fertility in a boy with prune-belly syndrome is unlikely. Abdominal wall repair is best accomplished with a large curving transverse incision in the lower abdomen extending into the flanks, the so-called “smile” incision. The combination of conservative treatment of the urinary tract and aggressive treatment of the undescended testes and abdominal wall has given many of these children a reasonably normal lifestyle.

Inguinal Hernia

Inguinal hernia results from a failure of closure of the processus vaginalis, a finger-like projection of the peritoneum that accompanies the testicle as it descends into the scrotum. Closure of the processus vaginalis should occur a few months prior to birth, which explains the high incidence of inguinal hernias in premature infants. In children, the male:female incidence of inguinal hernia is 10:1. Infants are at high risk for incarceration of an inguinal hernia because of the narrow inguinal ring. Differentiating between an incarcerated inguinal hernia and a hydrocele can be difficult, even with scrotal transillumination, and the symptomatic infant with a scrotal inguinal bulge must be evaluated by an experienced surgeon. Most often the incarcerated hernia can be reduced. The infant is sedated, and moderate bimanual pressure is applied by compressing the sac from below while a gentle counterforce downward is provided from the examiner's hand above the inguinal ring. Occasionally, these hernias will reduce spontaneously after sedation is given and the continuous struggling and crying are terminated. Following reduction of the incarcerated hernia, the child is admitted for observation, and herniorrhaphy is performed within the next 24 h to prevent recurrent incarceration. If the hernia cannot be reduced, or if intestinal obstruction is obvious, emergency operation, with reduction and repair, is necessary.

When the diagnosis of inguinal hernia is made in an otherwise normal child, operative repair should be planned, since spontaneous resolution does not occur. The frequency of a patent processus vaginalis on the side opposite the obvious hernia remains controversial. Data from laparoscopic evaluation of the normal appearing side revealed an incidence of greater than 30 percent patent processus vaginalis, regardless of the age or gender of the patient.

An inguinal hernia in a female frequently contains an ovary rather than intestine. Although the gonad usually can be reduced into the abdomen by gentle pressure, it often prolapses in and out until surgical repair is carried out. In some patients, the ovary and fallopian tube constitute one wall of the hernial sac (sliding hernia), and in these patients the ovary can be reduced effectively only at the time of operation. If the ovary is irreducible, prompt hernia repair is indicated to prevent ovarian torsion or strangulation.

When fluid fills the processus vaginalis, the patient has a hydrocele. The hydrocele may be associated with a hernia, may have a patent processus vaginalis too small for intestine to pass into (a communicating hydrocele), or may have no communication with the peritoneal cavity. When a hydrocele is diagnosed in infancy and there is no evidence of a hernia, observation is proper therapy until the child is older than 12 months. If the hydrocele has not disappeared by 12 months, invariably there is a patent processus vaginalis, and operative hydrocelectomy with excision of the processus vaginalis is indicated. When the first signs of a hydrocele are seen after 12 months of age, the patient should undergo elective hydrocelectomy, which in a child is always performed through a groin incision. Aspiration of hydroceles is discouraged, since almost all without a patent processus vaginalis will resorb spontaneously, and those with a communication to the peritoneum will recur and require operative repair eventually.

Operative repair of an inguinal hernia or hydrocele is accomplished through a small incision in a skin crease in the groin directly over the internal inguinal ring. The external oblique muscle is opened, and the cord structures and hernia sac are elevated into the wound, taking care not to grasp the vas deferens. The hernia sac is dissected to the internal ring and doubly suture ligated. The distal part of the hernia sac is dissected down to the testicle and widely opened to prevent a postoperative hydrocele. When the hernia is very large and the patient very small, tightening of the internal inguinal ring or even formal repair of the inguinal floor may be necessary, although the vast majority of children do not require any treatment beyond high ligation of the hernia sac. Inguinal hernias in children recur in less than 1 percent of patients, and recurrences usually result from missed hernia sacs at the first procedure, a direct hernia, or a missed femoral hernia. All children should have local anesthetic administered either by caudal injection or by direct injection into the wound. Spinal anesthesia in preterm infants decreases the risk of postoperative apnea when compared with general anesthesia.

GENITALIA

Cryptorchidism

The term *undescended testicle* describes that testicle that has been interrupted in its normal route of descent into the scrotum. Such a testicle may reside in the posterior abdomen, in the internal inguinal ring, in the inguinal canal, or even at the external ring. The testicle begins as a thickening on the urogenital ridge in the fifth to sixth week of embryologic life. In the seventh and eighth months the testicle descends along the inguinal canal into the upper scrotum, and with its progress the processus vaginalis is formed and pulled along with the migrating testicle. At birth, approximately 95 percent of infants have the testicle normally positioned in the scrotum.

A distinction should be made between the undescended testicle and the ectopic testicle. An ectopic testis, by definition, is one that has passed through the external ring in the normal pathway and then has come to rest in an abnormal location overlying either the rectus abdominis or external oblique muscle, or the soft tissue of the medial thigh, or behind the scrotum in the perineum. A congenitally absent testicle results from failure of normal development or an intrauterine accident leading to loss of blood supply to the developing testicle.

Clinical Manifestations

Usually a unilateral undescended testicle can be felt in the inguinal canal or in the upper scrotum. Occasionally, the testicle will be difficult or impossible to palpate, indicating either an abdominal testicle or congenital absence of the gonad. Patients with a nonpalpable testicle are routinely laparoscoped before inguinal exploration to identify an abdominal testicle. In patients with bilateral undescended testicles, it is appropriate to study the serum gonadotropin level, since the serum luteinizing hormone level is elevated in patients without gonadal tissue.

Reasons for surgical placement of the testicle in the scrotum are (1) diminished spermatogenesis, (2) malignant degeneration, (3) increased trauma (to a testicle located at the pubic tubercle), (4) increased incidence of torsion, and (5) psychological. The reason for malignant degeneration is not established, but the evidence points to an inherent abnormality of the testicle that predisposes it to incomplete descent and malignancy rather than malignancy as a result of an abnormal environment.

Males with bilateral undescended testicles are infertile. When the testicle is not within the scrotum, it is subjected to a higher temperature, resulting in diminishing spermatogenesis. Mengel and coworkers studied 515 undescended testicles by histology and demonstrated a decreasing presence of spermatogonia after 2 years of age. Consequently it is now recommended that the undescended testicle be surgically repositioned by 2 years of age. Despite orchidopexy, the incidence of infertility is approximately two times higher in men with unilateral orchidopexy compared to men with normal testicular descent.

Treatment

The use of chorionic gonadotropin occasionally has been effective in patients with bilateral undescended testes, suggesting that these patients are more apt to have a hormone insufficiency than children with unilateral undescended testicle. If there is no testicular descent after a month of endocrine therapy, operative correction should be undertaken. A child with unilateral cryptorchidism should have surgical correction of the problem.

Some patients who have an absent testis are greatly bothered by this anatomic deficiency. Prostheses of all sizes are now available and can be simply inserted into the scrotum, achieving normal appearance and a normal structure for palpation. Any patient who has an undescended testicle corrected surgically should be examined yearly by his surgeon until his midteen years. At that time, the individual should undergo thorough explanation about the possibility of malignant degeneration and be instructed in self-examination, which should be carried out at least twice a year for life.

Vaginal Anomalies

Anomalies of development of the vagina constitute a spectrum from simple defects (imperforate hymen) to more complex forms of vaginal atresia, including distal, proximal, and, most severe, complete. These defects are produced by abnormal development of müllerian ducts and/or urogenital sinus. The diagnosis is made most often by physical examination. Secretions into the obstructed vagina produce hydrocolpos, which may present as a large abdominal mass. US may help delineate the anatomy. The type of surgical repair depends on the extent of the defect. Simple imperforate hymen or a low transverse septum may be excised readily by a perineal approach. Other forms of vaginal atresia require complex reconstructive surgery, which may include mobilizing vaginal remnants down to the perineum, creating skin flaps, or using a colon segment as a neovagina.

Ovarian Cysts and Tumors

Ovarian cysts and tumors occur infrequently in childhood. They may be classified broadly as nonneoplastic or neoplastic. Nonneoplastic lesions include cysts (simple, follicular, inclusion, paraovarian, or corpus luteum), endometriosis, and inflammatory lesions. Neoplastic lesions are classified based on the three primordia that contribute to the ovary: mesenchymal components of the urogenital ridge, germinal epithelium overlying the urogenital ridge, and germ cells migrating from the yolk sac.

Most commonly, children with an ovarian mass complain of abdominal pain. Other signs and symptoms include a palpable abdominal mass, evidence of urinary obstruction, symptoms of bowel obstruction, and endocrine imbalance. The surgical approach depends on the appearance of the mass at operation, i.e., whether it is benign-appearing or is suspicious for malignancy. In the case of a simple, benign ovarian cyst, conservative excision is recommended, sparing ovarian tissue. If a lesion appears malignant, (1) ascites and peritoneal washings should be collected for cytologic study, (2) the liver and diaphragm are inspected carefully for metastatic disease, (3) an omentectomy is performed, (4) pelvic and paraaortic lymph nodes are biopsied, (5) the primary tumor is resected completely, and (6) the contralateral ovary is carefully inspected and, if appropriate, bisected and biopsied.

An increasing number of ovarian cysts are being detected by prenatal US. In the past, surgical excision was recommended for all cysts greater than 5 cm in diameter because of the perceived risk of ovarian torsion. More recently, it has become apparent from serial US examinations that many of these lesions will resolve spontaneously. Complex cysts of any size require surgical intervention. Surgeons have four options for managing simple cysts: (1) follow the lesion sonographically and operate if it fails to resolve over several months, (2) drain the cyst percutaneously by ultrasound- guided needle aspiration, (3) drain the cyst by laparoscopically guided needle aspiration, and (4) excise the cyst by conventional operation.

Ambiguous Genitalia (Intersex Syndromes)

Normal sexual differentiation occurs in the sixth fetal week. In every fetus, wolffian (male) and müllerian (female) ducts are present until the onset of sexual differentiation. Normal sexual differentiation is directed by the Y chromosome in a genetic switch initiated by the sex determining region of the Y chromosome (SRY). SRY is located on the distal 1A1 end of the short arm of the Y chromosome, adjacent to the pseudoautosomal region where small deletions in this area were known to be associated with sex reversal. Thus, SRY provides a genetic switch that initiates gonadal differentiation in the mammalian urogenital ridge. Müllerian inhibiting substance (MIS), a member of the transforming growth factor (TGF- β) family, results in regression of the müllerian duct, the anlagen of the uterus, Fallopian tubes, and the upper vagina. MIS gene transcription is initiated by tissue specific (gonadal ridge) and developmentally appropriate expression of the triple alpha-helical SRY protein that binds to the promoter region of the MIS gene in a sequence specific manner.

Simplistically, in the absence of SRY in the Y chromosome and in the presence of a duplicated X chromosome, the urogenital ridge differentiates as an ovary. In the absence of MIS that is produced by the Sertoli cells of the differentiated testis, the müllerian duct derivatives, i.e., the vagina, uterus, and Fallopian tubes, are preserved or fail to undergo regression. Thus the female phenotype prevails. In order for the male phenotype to develop, the embryo must have a Y chromosome, the SRY must be normal without point mutations or deletions, testosterone and MIS must be produced by the differentiated gonad, and the tissues must respond to these ligands via an appropriate receptor. Testosterone stimulates maturation of wolffian duct structures into epididymis, vas deferens, and seminal vesicles; simultaneously, the MIS produces regression of the female structures. Any disruption of the orderly steps in sexual differentiation may be reflected clinically as variants of the intersex syndromes (Fig. 37-33). These may be classified as (1) true hermaphroditism (with ovarian and testicular gonadal tissue), (2) male pseudohermaphroditism (testicles only), (3) female pseudohermaphroditism (ovarian tissue only), and (4) mixed gonadal dysgenesis (usually underdeveloped or imperfectly formed gonads).

Male pseudohermaphroditism is found in genotypic males with bilateral testes; however, the duct structures of many of these patients differentiate partly as phenotypic females, resulting from defects in androsynthesis or incomplete müllerian regression. Female pseudohermaphroditism is found most commonly in patients with congenital adrenal hyperplasia who are unable to synthesize cortisol due to abnormally low levels of any one of five enzymes. In 90 percent of cases, deficiency of 21-hydroxylase causes adrenocorticotrophic hormone (ACTH) to stimulate the secretion of excessive quantities of adrenal androgen, which masculinizes the developing female. These infants are prone to

salt loss and can experience collapse at about 1 week of age. The rarest intersex form, that of the true hermaphrodite, is usually found with XX karyotype. These children have ambiguous genitalia and their gonad pattern may show an ovary and a testicle, or an ovotestis.

In the differential diagnosis of patients with intersex anomalies, the following diagnostic steps are necessary: (1) evaluation of the genetic background and family history; (2) assessment of the anatomic structures by physical examination and x-ray studies; (3) chromosome studies; (4) determination of biochemical factors in serum and urine; and (5) when necessary, laparotomy and gonadal biopsy. Genetic females should be assigned the female gender, regardless of the degree of virilization. For genetic males, the gender assignment will depend on anatomic considerations, specifically the size of the phallus, since at this time satisfactory surgical techniques do not exist to reconstruct an inadequate phallus. The physicians presented with a newborn infant with ambiguous genitalia must make the gender assignment rapidly, to minimize the emotional impact on the family, and accurately, since a change later causes irreparable anguish. Assignment to one of four major groups of defects described above can be made readily using two screening criteria: (1) the presence of gonadal symmetry or asymmetry, and (2) the presence of a chromatin mass or Barr body or the absence of fluorescence of the distal long arm of the Y chromosome. After the correct designation to one of the major categories, the specific pathoetiology can be determined at a more deliberate pace.

Gonadal symmetry is determined by the position of one gonad relative to the other, either above or below the external inguinal ring. If the gonadal position is symmetric, both gonads lie either above or below the inguinal ring, suggesting a diffuse etiologic basis for the abnormality. For example, in either male or female pseudohermaphroditism, a biochemical defect affects both gonads equally, such that the gonads are symmetric. In contrast, in mixed gonadal dysgenesis or true hermaphroditism, asymmetry is observed if there is a predominance of testicular tissue on one side and ovarian tissue on the other side, in which case the testicular gonad is descended and the ovarian gonad is undescended.

With more than one X chromosome present in the karyotype, the inactivated X chromosome (Barr body) can be found tangentially in the nucleus, which is the case with female pseudohermaphroditism (adrenogenital syndrome) or in most cases of true hermaphroditism. Conversely, Barr bodies are absent in patients with male pseudohermaphroditism and mixed gonadal dysgenesis. At centers that perform analyses to mark the distal long arm of the Y chromosome, instead of chromatin analysis with Barr bodies, fluorescent Y positive nuclei are seen with male pseudohermaphrodites and mixed gonadal dysgenesis, and fluorescent Y negative nuclei are seen with female pseudohermaphrodites and true hermaphrodites. With this algorithm, a preliminary diagnosis that assigns the patient to one of four major groups can be established with 80 to 90 percent accuracy within the first 24 hours of life.

After complete evaluation, certain plastic surgical procedures are required to harmonize the external genitalia with the sex of rearing. Operations to reduce the size of the enlarged clitoris have been developed that spare the sensation and function of the clitoris. Plastic procedures to exteriorize the vagina or separate it from the urethra are necessary in patients born with a urogenital sinus. When male assignment is appropriate for an infant with ambiguous genitalia, hypospadias repair will be necessary. When contradictory gonads or ovotestes are present, removal of these structures is required to prevent the possibility of hormone secretion or malignant degeneration. For psychological adjustment of some teenage male patients with inadequate or absent gonads, the insertion of testicular prostheses may prove beneficial. Children with endocrine deficiency may require lifetime exogenous supplementation. Prompt recognition of infants with intersex anomalies, followed by appropriate sex assignment and proper treatment, prevents the social and psychological derangements that have occurred in the past because of delayed diagnosis or inappropriate gender assignment.

NEOPLASTIC DISEASE

Cancer is the second leading cause of death in children. Approximately 11 percent of the deaths of children in the United States are due to malignant diseases (Fig. 37-34). In the past 30 years there has been a marked increase in the survival of patients with childhood cancer. This improvement can be attributed to better diagnostic imaging techniques; new chemotherapeutic agents; collaborative approaches to surgery, chemotherapy, and radiation therapy; and multi-institutional studies evaluating new treatments and protocols. This unified approach to diagnosis, staging, and therapy has proved useful in the management of the more common pediatric solid malignancies, such as Wilms' tumor, hepatic tumors, rhabdomyosarcoma, neuroblastoma, and teratoma. Advances in the understanding of molecular biology, immunology, pharmacology, and tumor biology surely will lead to improvements in care in the next decade. Best results are achieved in institutions with special skill in the care of children and with access to the latest multimodality and multidisciplinary protocols.

Wilms' Tumor

Wilms' tumor is an embryonal neoplasm of the kidney that usually presents as an asymptomatic mass in the flank or upper abdomen. The peak age of incidence is between 1 and 5 years. This tumor has been associated with congenital anomalies such as aniridia, the Beckwith-Wiedemann syndrome, urinary tract defects, hemihypertrophy, and chromosomal deletion, suggesting a possible hereditary influence. Specific Wilms' tumor suppressor genes have been discovered and cloned (WT1 and WT2), mutations or deletions of which may be responsible for the development of Wilms' tumors.

Before operation, all patients suspected of Wilms' tumor should be evaluated radiographically. Abdominal and chest CT scans and US are the mainstays of the work-up. US confirms the presence of a mass and localizes it to the kidney. It can demonstrate the presence of renal vein or vena caval extension, crucial information before embarking on surgical intervention. The CT scan usually will show a large intrarenal mass (Fig. 37-35). Furthermore, the scan provides important preoperative data such as the tumor status of the contralateral kidney and the presence of local or distant spread. MRI is finding an increasing role in the imaging of these tumors.

The key to staging and treatment is surgical intervention. A generous transabdominal transverse incision is used to ensure adequate exposure for removing the primary tumor, to evaluate the opposite kidney, and to inspect the rest of the abdomen. The tumor and kidney are completely resected with care to avoid tumor spillage. Whenever possible the renal vein should be clamped before the tumor is mobilized, to prevent tumor embolization. If there has been contiguous spread of the tumor into adjacent organs, the operation may be expanded to include part of the liver, spleen, diaphragm, pancreas, or stomach. Lymph nodes are sampled to assist with staging.

Following operative removal of a Wilms' tumor, the need for chemotherapy and/or radiation therapy is determined by the histology of the tumor and the clinical stage of the patient (Table 37-1). Patients with disease confined to one kidney, totally removed surgically, receive a short course of chemotherapy and can expect a 97 percent 4-year survival, with tumor relapse rare after that time. Patients with Stage II and III disease with favorable histology receive a longer course of combination chemotherapy. Patients with Stage IV disease or unfavorable histology require longer and more intensive chemotherapy. Radiation therapy is used for the advanced stages. Even in Stage IV, cure rates of 80 percent are being achieved. The survival rates are worse in the small percentage of patients considered to have unfavorable histology. Actinomycin D and vincristine are the major chemotherapeutic agents used. Adriamycin and cyclophosphamide are used in advanced stages. Despite the enormous contribution of chemotherapy and radiation to improved survival, surgical excision remains the primary treatment. In selected patients, preoperative chemotherapy is used to shrink massive tumors or to reduce tumor bulk when bilateral disease is present. In the case of massive tumors, this approach may reduce tumor spill and make operations safer. With bilateral disease it allows for conservation of renal tissue at the time of resection.

Current protocols designed by the National Wilms' Tumor Study (NWTs) focus on decreasing therapy for low-risk patients and on intensifying therapy for high-risk patients. Additionally, new strategies are required for children with recurrent disease, since the present outlook with conventional therapy is dismal. Long-term follow-up is required for all children treated for Wilms' tumor because of the increasing recognition of second malignancies.

Neuroblastoma

Neuroblastoma is the third most common pediatric malignancy. Over 60 percent arise in children under the age of 5 years. Neuroblastomas arise from the neural crest cells and show different levels of differentiation. The tumor originates most frequently in the adrenal glands, posterior mediastinum, neck, or pelvis but can arise in any sympathetic ganglion. The clinical presentation depends on the site of the primary and the presence of metastasis.

Two-thirds of these tumors are first noted as an asymptomatic abdominal mass. The tumor may cross the midline, and a majority of patients will already show signs of metastatic disease. The patient should be evaluated by CT scan and US, which usually show displacement and occasionally obstruction of the ureter of an intact kidney. Since these tumors derive from the sympathetic nervous system, catecholamines and their metabolites will be produced at increased levels and may produce symptoms. Measurement of VMA and HVMA in serum and urine aids in the diagnosis and in monitoring adequacy of future treatment and recurrence. Elevations of serum ferritin and neuron-specific enolase levels correlate with poor prognosis. Increased copy number of the N-myc gene is an important indicator of poor prognosis.

Although CT scan, bone marrow biopsy, and radionuclide scans are important adjuncts to preoperative staging, surgical evaluation and excision, when possible, are essential. The main hope of cure is to resect the tumor completely, because chemotherapy and radiation therapy have not altered survival significantly in the past 20 years. Abdominal tumors are approached through a transverse incision, and every attempt is made to resect the tumor completely. Thoracic tumors may be approached through a posterolateral thoracotomy and often have an intraspinal component necessitating laminectomy and intraspinal removal. Many staging classifications have been developed to predict survival and to stratify patients for treatment protocols (Table 37-2). The most commonly employed are those developed by the Children's Cancer Study Group and the International Neuroblastoma Staging System.

If the tumor is not resectable, a generous biopsy is obtained for histopathologic classification and determination of N-myc gene copy number, both important determinants of prognosis. Under current experimental protocols, the children receive several courses of chemotherapy to shrink the tumor and then undergo a "second-look" attempt at resection. Protocols incorporating bone marrow transplantation have been used for patients with advanced disease. Even with aggressive chemotherapy protocols, the 2-year survival rate for older children with metastatic disease is 20 percent.

Owing to variability in sites and in biologic behavior, the treatment of neuroblastoma must be individualized. In infants favorable survival may be realized even when there is metastatic spread to bone marrow, liver, or skin (Stage IVs). Over 85 percent of these patients, if less than one year old, will be completely and spontaneously cured. The spontaneous regression of tumor seen in many of these patients has led to much investigation of immunotherapy as a possible therapeutic tool. Although cure rates of greater than 80 percent have been achieved in Stages I and II, the poor survival in the higher stages is discouraging and calls for new therapies and further research.

Rhabdomyosarcoma

Rhabdomyosarcoma is a primitive soft tissue tumor that arises from mesenchymal tissues. The head and neck (36 percent), extremities (19 percent), genitourinary tract (21 percent), and trunk (9 percent) are the most common sites of origin, although the tumor can arise virtually anywhere. The clinical presentation of the tumor depends on the site of origin. The diagnosis is confirmed with incisional or excisional biopsy after evaluation by MRI, CT scans of the affected area and the chest, and bone marrow biopsy. The tumor grows locally into surrounding structures and metastasizes widely to lung, regional lymph nodes, liver, brain, and bone marrow. The staging system for rhabdomyosarcoma is shown in Table 37-3.

The overall survival is 60 percent at 5 years. Prognosis relates to the site of origin, clinical grouping, and histopathology. The prognosis in patients with embryonal pathology is much more favorable than that in patients with alveolar histology. Five-year survival rates are 81 percent in clinical group I, 80 percent in clinical group II, 65 percent in clinical group III, and 27 percent in clinical group IV. The Intergroup Rhabdomyosarcoma Study (IRS) has recently developed a staging system based on TNM principles that may serve as the basis for future studies designed to improve survival.

Wide local excision with adequate “clean” margins is the optimal surgical treatment for localized forms of rhabdomyosarcoma. The potential morbidity of radical amputations or exenteration can be avoided because of progress in chemotherapy and radiation therapy in recent years. Often the approach involves initial biopsy, several courses of chemotherapy, and resection at a “second-look” operation. The standard chemotherapy is a combination of actinomycin D, vincristine, and cyclophosphamide. Treatment is individualized based on the site, group, and histology, but all will receive multimodality therapy.

Teratoma

Teratomas are tumors composed of tissue from all three embryonic germ layers. They may be benign or malignant, may arise in any part of the body, and are usually found in midline structures. Thoracic teratomas usually present as an anterior mediastinal mass. Ovarian teratomas present as an abdominal mass often with symptoms of torsion, bleeding, or rupture. Retroperitoneal teratomas may present as a flank or abdominal mass. The location incidence of teratomas in children is seen in Table 37-4. The goal of therapy is complete surgical excision because the success of chemotherapy and/or radiation therapy seen in other pediatric tumors has not been realized in this group of tumors.

Sacroccocygeal Teratoma

Sacroccocygeal teratoma usually presents as a large mass extending from the sacrum in the newborn period. Many tumors have been diagnosed prenatally by US. In fetuses with evidence of hydrops and a large sacroccocygeal teratoma, prognosis is poor, thus prenatal intervention has been advocated in such patients. The mass may be as small as a few centimeters in diameter or as massive as the size of the infant (Fig. 37-36). There is a form of tumor that does not present externally but grows in the presacral space and often presents later in childhood. The differential diagnosis consists of neural tumors, lipoma, and myelomeningoceles.

Most tumors are identified at birth and are benign, but with advancing age the potential for malignant degeneration is high. Complete resection of the tumor as early as possible is essential. The rectum and genital structures are often distorted by the tumor but usually can be preserved in the course of resection. The cure rate is excellent if the tumor is excised completely. Perioperative complications of hypothermia and hemorrhage can occur with massive tumors and may prove lethal in “benign” lesions. With discovery of the tumor in older infants and children, the results are poor because of the high incidence of malignant degeneration. Chemotherapy is indicated in these older children.

Liver Tumors

More than two-thirds of all liver tumors in children are malignant. Hepatoblastoma is the most common malignancy of the liver in children, with most of these tumors diagnosed before 4 years of age. Hepatocellular carcinoma is the next most common, with a peak age incidence between 10 and 15 years. Malignant mesenchymomas and sarcomas are much less common but constitute the remainder of the malignancies. Most children with a liver tumor present with an abdominal mass that is usually painless, which the parents note while changing the child's clothes or while bathing the child.

The patients are rarely jaundiced but may complain of anorexia and weight loss. Most liver function tests are normal. Alpha-fetoprotein levels are elevated in 90 percent of children with hepatoblastomas but are increased much less commonly in other liver malignancies. Radiographic evaluation of these children should include a flat-plate abdominal radiograph, an US study, and a contrast-enhanced abdominal CT scan.

Complete surgical extirpation of the tumor is the primary goal and is essential for cure. For resectable tumors, attempts to reduce the bulk of massive tumors by pretreatment with chemotherapy have met with limited success. The value of radiotherapy is also limited. Previously, hepatic transplantation for unresectable primary lesions had a high incidence of carcinomatosis and death; however, recent results are more encouraging. For hepatoblastoma, about half the patients have completely resectable lesions (Fig. 37-37), and about 80 percent of these can be cured with adjunct chemotherapy. Of all patients with hepatoblastomas, only about 70 percent are long-term survivors. Patients with hepatocellular carcinoma have a worse prognosis because the tumors are frequently multicentric or metastatic. The cure rate for all children with hepatocellular carcinomas is only 25 percent.

TRAUMA

Injuries account for almost half of all pediatric deaths, more than cancer, congenital anomalies, pneumonia, heart disease, homicide, and meningitis combined. Motor vehicle collision involving car occupants or pedestrians is the most common cause of accidental death in children; drowning, burns, and firearms account for a significant segment of the remaining group. Not evident in the mortality statistics is the number of children who sustain injury and recover with a subsequent disability. Improved methods of communication, immediate prehospital care, and rapid transportation to regional centers increase the probability of survival after a major traumatic injury. Because of the complex requirements for resuscitation and treatment of an injured child, it is essential that each regional referral center establish a systematic approach to care.

Management of trauma in children requires immediate recognition and treatment of life-threatening injuries to the head, thorax, and abdomen. Recognition that the child with severe head injury manifests a different physiologic response from that of the adult forms the foundation of successful therapy. Aggressive therapy of elevated intracranial pressure results in improved survival; 30 percent of the childhood victims of serious head trauma die, whereas the remainder make a good to excellent recovery. Thoracic trauma occurs in 5 percent of all injured children. Blunt abdominal trauma is common in the pediatric population, the spleen and liver are the most frequent intraabdominal organs sustaining injury. Nonoperative management is a reasonable approach in the child with blunt hepatic or splenic trauma with stable vital signs. Exploratory laparotomy is indicated for a physiologically unstable child despite appropriate fluid resuscitation. These children should undergo laparotomy for hemostasis, and an attempt at splenic salvage by splenorrhaphy is warranted if the child's condition permits. Splenic salvage is important in children to minimize the potential for future overwhelming sepsis. Lap belt injuries are increasingly common in children due to mandatory seat belt laws. Lap-belt injuries may cause serious intestinal trauma and lumbar spine fractures and are poorly diagnosed by the abdominal CT. Peritoneal lavage or laparotomy is indicated in this group of injured children.

Consideration of the unique requirements of the injured child during evaluation and therapy improves outcome and minimizes mortality. Aerophagia proceeding to gastric dilatation can compromise respiration or mimic an abdominal injury; therefore, nasogastric decompression is essential in all children with significant injury. The temperature of injured children might drop precipitously because of the relatively large surface area to body mass ratio. Hypothermia potentiates the deleterious effects of shock and metabolic acidosis. Hypovolemic shock with attendant cellular hypoperfusion requires treatment by infusion of crystalloid solutions at a rate of 20 mL/kg, I.V. push. Urine output (1mL/kg/h) in children is a useful indicator of adequate fluid resuscitation. Once initial assessment and resuscitation are complete, diagnostic measures with conventional x-rays and CT scan are important. Prompt movement to a pediatric intensive care unit is essential unless a decision for operation is necessary. Complete rehabilitation of the child and the family should be the ultimate goal of every pediatric trauma center. The solution to pediatric trauma is not improved schemes for treatment but rather prevention of injury.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 38 - Urology

Margaret S. Pearle
John D. McConnell
Paul C. Peters

ANATOMY

Kidney

The kidneys are paired retroperitoneal organs that weigh approximately 160 g each in the healthy adult. They are situated within the fascia of Gerota, and a variable amount of perinephric fat is present between the capsule of the kidney and this fascial envelope. Further protection of the kidneys is provided dorsally (posteriorly) by their relationship with the lower ribs, the quadratus lumborum, and the psoas muscles. The ventral (anterior) relationships of the kidney are illustrated in Fig. [38-1](#). The kidneys lie beneath the diaphragm, with the right kidney usually 1 to 2 cm lower than the left kidney (Fig. [38-2](#)).

The renal arteries originate from the aorta at the upper level of the second lumbar vertebra; the right renal artery is longer than the left because it crosses the midline behind the vena cava. In about two-thirds of the population, each kidney has a single renal artery. Multiple renal arteries occur in normal kidneys but are more prevalent in congenitally malformed or malpositioned kidneys. The main renal artery divides into five major branches, each of which represents an end artery supplying a renal segment. Partial occlusion of any branch produces ischemia, and complete occlusion produces infarction of the involved segment. The ureteral and capsular vessels can provide significant blood flow in the presence of marked renal artery occlusion. The left renal vein crosses the midline ventral to the aorta after it has received blood from the left adrenal and gonadal veins. The renal lymphatics empty through hilar trunks in the region of the renal artery and vein. Capsular lymphatics travel through the perinephric fat to infradiaphragmatic periaortic nodes. The renal nerves, which receive contributions from T4 to T12, from the vagus nerve via the celiac axis, and from the splanchnic, contain vasomotor and pain fibers. After total renal denervation, such as occurs with transplantation, no persistent abnormalities in renal function occur. The renal pelvis, which usually contains approximately 5 mL of urine, lies dorsal to the renal vessels and has a transitional cell epithelium. The kidney's position, with respect to the vertebral bodies and the axis of a line drawn through the upper and lower calyces, offers a clue to focal diseases of the kidney and adjacent organs. The axis of the kidney often parallels that of the psoas muscle edge, as noted on x-rays of the abdomen (see Fig. [38-2](#)).

Ureter

The ureters are muscular tubes that connect the renal pelvis to the bladder, traversing the retroperitoneal space in a linear course just lateral to the transverse processes of the lumbar vertebrae and crossing the common iliac arteries at their bifurcation. The lower ureter follows the contour of the pelvis and deviates laterally and then medially in a gentle curve to reach the trigone of the bladder above a line drawn between the spinous processes. The normal adult ureter is 28 to 30 cm long and about 5 mm in diameter and has an oblique course through the wall of the bladder. The ureteral orifice is an oblique slit in the trigone.

The function of the ureter is to transmit urine from the renal pelvis to the bladder. Ordinarily, urine is not modified in the ureter. Normally, there are about four peristaltic waves per minute, developing pressures of about 30 mmHg.

The ureteral blood supply originates from the renal, aortic, iliac, mesenteric, gonadal, vasal, and vesical arteries. Free intercommunication between these vessels permits extensive ureteral mobilization and transposition. Pain fibers refer stimuli to the T12 through L2 segments, whereas the autonomic innervation is associated with intrinsic parasympathetic motor and sympathetic vasomotor ganglia. The lymphatic drainage is to segmental periaortic and caval nodes. The ureter may be drawn medially in retroperitoneal fibrosis and laterally as a result of enlargement of periaortic lymph node involvement with tumor or an aortic aneurysm. It is essential to be aware of the course of the ureter during aortic and pelvic surgery and in difficult dissections of adjacent organs.

Bladder

The urinary bladder is a muscular pelvic organ lined with transitional cell epithelium. It is related to the peritoneum superiorly and posteriorly, the sigmoid colon and rectum in the male, and the uterus, cervix, and vagina in the female. In the male, the bladder is spherical; in the female, the uterus indents the dome. The smooth muscle (*detrusor*) is capable of stretching to a marked degree. The major blood supply originates from the superior, middle, and inferior branches of the hypogastric arteries. The lymphatics drain to the perivesical, hypogastric, and periaortic nodes. The autonomic nerve supply to the bladder is derived from the sacral cord and from the presacral and epigastric plexuses of nerves.

Prostate and Seminal Vesicles

The chestnut-shaped adult prostate surrounds the proximal male urethra. It is firmly attached to the bladder neck and symphysis pubis. Posteriorly, the fascia of Denonvilliers intervenes between the rectum and the prostate and seminal vesicles (Fig. [38-3](#)). Caudad, the prostate rests on the pelvic diaphragm that contains the voluntary external sphincter. The blood supply derives from the inferior vesical, middle hemorrhoidal, and internal pudendal arteries; the venous drainage communicates with an extensive pelvic plexus that empties into the hypogastric veins. This plexus also communicates with Batson's veins, partially explaining the frequently encountered metastatic spread of prostatic carcinoma to the bony pelvis and lumbar vertebral bodies. The prostate receives secretory and motor (parasympathetic) innervation from S3 and S4 and vasomotor (sympathetic) fibers from the hypogastric plexus. The lymphatics drain into the obturator nodes, the external, internal, and common iliac nodes, and then to the periaortic nodes.

The seminal vesicles lie behind the bladder lateral to the ampullae of the vasa deferentia and are closely related to the ureters. Their secretions are rich in fructose, which may be of importance to sperm motility.

Penis and Urethra

The penis is composed of two lateral spongy erectile bodies (*corpora cavernosa*) and a single ventral body (*corpus spongiosum urethrae*) through which the urethra passes. The latter terminates in the glans penis, which is also composed of erectile tissue. As the urethra emerges from the pelvic diaphragm, it enlarges to become the bulbous urethra. The urethra continues as the pendulous urethra, which terminates in the tip of the glans penis (Fig. [38-4 A](#)). The female urethra corresponds to the prostatic and membranous urethra in the male. It averages 4 cm in length and 32 mm in circumference. The principal blood supply to the penis and urethra originates from the internal pudendal arteries. Somatic sensory innervation of the penis is from S3 and S4 via the ilioinguinal and genitofemoral nerves. Sympathetic vasomotor innervation derives from the hypogastric plexus, whereas the parasympathetic innervation originates from S2, S3, and S4 via the nervi erigentes. The nervi erigentes give rise to the cavernosal nerves that course on the posterior surface of the prostate in route to the corpora cavernosa of the penis. Injury to the cavernosal nerves during pelvic surgery will lead to impotence. The lymphatic drainage is to the superficial and deep inguinal nodes and then to the external iliac and hypogastric nodes.

Testis and Epididymis

The testis is an ovoid, firm scrotal organ that measures 4 × 2.5 × 2.5 cm. The left testis commonly resides lower in the scrotum than the right. The testis weighs approximately 20 g and is covered by a tough membrane, the tunica albuginea, except at its dorsal aspect, where the epididymis and vascular pedicle are attached. The epididymis is a crescent-shaped body that curves around the dorsal portion of the testis. The vas deferens is a 4-mm-thick walled, firm tubular structure that originates at the inferior pole of the epididymis and follows a cranial course with the spermatic vessels (see Fig. [38-4](#)). The arterial blood supply to the testis and epididymis originates from the aorta just below the renal arteries. The left spermatic vein empties into the left renal vein; the right spermatic vein empties directly into the inferior vena cava. The primary lymphatic drainage from the testis is to the periaortic nodes in the vicinity of the kidney.

DIAGNOSIS

Accuracy of diagnosis is a characteristic feature of the practice of urology. In order to maintain the achievable degree of accuracy, it is essential that a clinical problem be approached in a logical sequence. History, physical examination, laboratory tests, and imaging tests lead to a reasonable estimate of diagnosis. Rarely is a urologic diagnosis made on the basis of surgical exploration.

History

A number of symptoms and signs are characteristic and often diagnostic of pathologic conditions involving the genitourinary system. The patient may spontaneously offer these clues, but in most instances it is necessary to ask proper questions to obtain this information.

Gross Hematuria

The presence of gross blood in the urine is a significant sign that is quite alarming to the patient and warrants further evaluation. The degree of bleeding is not proportionate to the severity of the underlying disease. Even small amounts of blood in the urine occurring on one occasion may be the only indication of a malignant process in the urinary tract. Intermittent bleeding is common; large tumors may manifest with only a small amount of bleeding, whereas small tumors may produce considerable loss of blood. The bleeding originates from lesions that cause erosion or disruption of blood vessels or from inflammatory changes. The common causes are inflammation (including infections), tumors, calculi, and trauma.

In young individuals, gross hematuria is more likely to be the result of infection, and in older patients, it is more likely to be related to tumors and prostatic disease. Any patient with unexplained *asymptomatic* gross hematuria should be evaluated for possible tumor.

It is important to determine whether hematuria appears at the beginning or end of voiding. Initial hematuria suggests lesions distal to the bladder neck, namely, the prostatic and membranous urethra. Predominantly terminal hematuria usually indicates involvement of the proximal urethra, bladder neck, or trigone. Uniform hematuria occurs with lesions of the bladder, ureter, or kidney. The passage of blood clots suggests that the bleeding is quantitatively great because urine inhibits coagulation of blood by the presence of citrate. Long, wormlike clots suggest formation in the ureter. Reddish clots occur with recent bleeding, whereas brownish or grayish white clots indicate a time lapse between the bleeding episode and their passage. These “localizing” signs are not reliable enough to pinpoint the diagnosis. Complete imaging and endoscopic evaluation of the entire urinary tract is usually warranted.

Acute Urinary Retention

Acute urinary retention (AUR) is a term that applies to a complete inability to urinate, and this is one of the most distressing symptoms. It should not be referred to as “anuria” because the latter indicates that urine is not being formed. The patient usually complains of lower abdominal distress, and the bladder is percussible or palpable as a tender suprapubic mass.

Benign prostatic hyperplasia (BPH) is the most common cause of AUR in men. Most have a long-standing history of increasing difficulty in voiding, but occasionally an acute event (e.g., surgery, decongestant medications, etc.) may precipitate AUR in a patient with minimal prior symptoms. By contrast, *carcinoma of the prostate* is accompanied by a much shorter duration of symptoms. The history may be of months rather than years, and the urinary difficulties increase more rapidly. The patient with carcinoma of the prostate seldom presents with urinary retention when the lesion is early and amenable to surgical treatment. The presence of retention in a cancer patient therefore suggests advanced disease, and symptoms of advanced malignancy such as malaise, weight loss, anorexia, and back pain are likely to be encountered. Carcinoma of the prostate can coexist with BPH.

In young males, *prostatic infection* may cause AUR. This results from acute prostate swelling. There is usually an antecedent history of frequent urination and dysuria progressing to acute retention. Fever, chills, and symptoms of systemic infection are common, and abscess formation occurs occasionally. AUR retention may be a manifestation of *urethral stricture*. These patients often have an antecedent history of gonorrhea, transurethral instrumentation, trauma, or radiation therapy. Many will have noted a gradual decrease in the force of the stream.

Neurogenic bladder dysfunction may lead to a progressive increase in residual urine and eventually to complete retention. This may be the first indication of spinal cord disease. In other patients, the retention may be preceded by difficulty in initiating voiding and overflow incontinence related to changes in abdominal pressure exerting force on a markedly distended bladder. Acute retention on a neurogenic basis may follow trauma, pelvic surgery, the administration of general anesthetics, spinal anesthesia, or the administration of certain drugs that influence the innervation of the bladder, bladder neck, proximal urethra, or external sphincter. AUR in women is unusual and may be caused by neurogenic and psychogenic factors or urethral obstructions secondary to carcinoma, stricture, vaginal lesions, or cervical fibroids. A cystocele is seldom the cause of AUR.

Incontinence

Incontinence is the involuntary loss of urine that the patient views as socially unacceptable. *Total incontinence* is continuous and not associated with urgency or stress. A vesicovaginal, ureterovaginal, or vesicoperineal fistula may produce total incontinence. Occasionally, an ectopic ureter will open into the vagina, and this also will produce a constant leakage of urine. *Stress incontinence* is the leakage of urine associated with increase in abdominal pressure from coughing, lifting, etc. It is the result of lax pelvic support in women (following childbirth or menopause) and resulting urethral hypermobility or sphincteric dysfunction in men (following prostatectomy). *Urgency incontinence* refers to urinary leakage preceded by the sensation of an urgent need to urinate. This is often the result of a uninhibited contraction of the bladder due to infection, bladder cancer, neurogenic disease, or idiopathic causes. *Overflow, or paradoxical, incontinence* represents a fourth type, in which the bladder retains a large amount of residual urine. With movement, the increased abdominal pressure may cause overflow of a small amount of urine from a distended bladder. This condition may accompany neurogenic diseases of the spinal cord, or it may be the result of long-standing obstructive uropathy with detrusor decompensation. These patients may never really void, intermittently passing small quantities of urine without control. Incontinence of urine is a common manifestation of various neurologic diseases and spinal cord trauma.

Ureteral Colic

Acute obstruction of the kidney or ureter produces very painful distention. Typically, there is sudden unilateral severe crescendo pain in the posterior subcostal region. The localization is frequently so precise that it can be defined by fingertip palpation. Obstruction at the ureteropelvic junction is associated with pain at the costovertebral angle, whereas obstruction lower in the ureter may have the added component of pain in the ipsilateral lower abdomen. With obstruction of the intramural ureter, pain also may be referred to the corresponding side of the scrotum or labia majora. Ureteral colic related to intramural obstruction is often associated with a sudden onset of urgency and frequency.

Lower Urinary Tract Symptoms (LUTS)

Frequency refers to the patient's voiding an excessive number of times. Frequency of urination should not be confused with *polyuria*, where the patient excretes large volumes of urine at more frequent intervals than normal. Most normal, well-hydrated individuals void four to six times daily and do not arise from sleep to empty the bladder. Frequency may be related to reduction in bladder capacity associated with inflammation, bladder outlet obstruction, effective reduction in functional bladder capacity by increased residual urine, or uninhibited bladder contractions. Frequency is also a symptom of psychological stress.

When patients complain of frequency, it is essential to question them carefully regarding fluid and caffeine intake. It is sometimes of value to have the patient record accurately the time and volume of the amounts of fluid ingested as well as the amount excreted in a 24-h period (frequency- volume chart). It is also essential to note the medications the patient is taking because this may influence urinary frequency considerably. Additional information about the relationship of urinary frequency to work, stress, weekend activity, and vacations is extremely helpful. Many patients seek help from other physicians for this disturbing symptom, which may limit their activities because of the need to be close to a bathroom at all times. A careful history may lead to information that will allow simple modification of habits, intake, and drug therapy to bring relief from this distressing and sometimes incapacitating problem.

Nocturia, awakening at night to void, may be caused by consumption of an excessive amount of fluids before retiring, or it may be an expression of generalized restlessness. With cardiac decompensation, fluids that have accumulated in dependent portions of the body during the day are restored to the circulating blood when the patient maintains a horizontal position, thus causing nocturia. Patients with chronic renal disease who excrete large volumes because of an inability to concentrate the urine and patients on diuretics frequently experience nocturia. Prostatic obstruction and acute or chronic infection of the urinary tract cause nocturia by means of the same mechanisms described for frequency.

Urgency is due to an increase in bladder pressure. Urgency and frequency usually occur together in obstruction and infection. However, urgency can be a normal consequence of voiding postponement for prolonged periods of time. The symptom may be so severe that the patient cannot restrain voiding, resulting in *urgency incontinence*. Difficult or painful urination is known as *dysuria*. The sensation is commonly described as burning and may be referred to the glans penis or perineum. The symptom may be caused by urinary infection, the passage of clots or calculi, or primary bladder disease. Bladder cancer and other diseases must be excluded when the urine culture is sterile. Severe pain at the termination of or subsequent to voiding is referred to as a *strangury* and may be associated with the presence of a bladder calculus or infection. Difficulty of urination is usually the result of an obstructive uropathy. *Hesitancy*, indicating delayed voiding in response to mental command, is another symptom of chronic obstructive uropathy or sphincter dyssynergia. *Intermittency* refers to the involuntary stopping and starting of the stream during voiding and may be a symptom of obstruction.

Urinary Stream

The stream may lack force and have reduced projection in patients with lower urinary tract obstruction. Prostatic enlargement and urethral stricture are the most common causes. Stenosis of the meatus, which determines the caliber of the urinary stream, may be evidenced by a thin, deviated, or duplicated stream.

Erectile and Ejaculatory Dysfunction

Disorders of sexual function are usually the result of systemic disease processes, such as atherosclerosis. A careful history about medications is essential because certain drugs produce erectile and ejaculatory disturbances. The cause of loss of erection may be endocrinologic, vasculogenic, or neurogenic. In some situations, however, the underlying factor is situational or psychogenic, and this may become apparent during careful history taking. In instances where physical or anatomic abnormality is found to account for impotence, it may be correctable. It can be restored in most patients by medical or surgical therapy.

Physical Examination

Renal Areas

The kidney region can first be examined with the patient in the upright position. Observation may reveal obvious bulging or asymmetry of the costovertebral region. Scoliosis may be present from guarding in the presence of unilateral pain. Herpetic lesions are occasionally encountered on the skin surface of this area and offer a clue to the cause of pain. Gentle palpation of the costovertebral angle areas should be followed by sharp percussion when attempting to disclose an underlying obstructed or infected kidney. If the patient complains of pain in the flank region before examination, it is wise to start the evaluation on the contralateral side.

Thorough examination of the kidney areas should be carried out with the patient supine, knees flexed, and arms at the side. The examiner should stand on the side being examined. The posterior hand is placed parallel to the twelfth rib and below it. The anterior hand is placed 4 cm below the anterior rib cage and parallel to the posterior hand. Renal and retroperitoneal masses can be balloted between the two hands. In slender individuals, the lower pole of the right kidney is often palpable. In others, the kidney may be palpable with deep inspiration. With unusually mobile kidneys, the organ may be more readily palpable in the upright position. Tenderness is usually related to obstruction or infraction and is uncommon with uncomplicated tumors or cysts. The examiner again looks for evidence of asymmetry, rigidity of the costovertebral angle, tenderness, or bulging, which may suggest the presence of an underlying abscess, obstructed kidney, inflammation, or retroperitoneal extravasation of urine or blood.

Ureters

The deep retroperitoneal location of the ureters does not lend itself to palpation. It is unusual to be able to feel them even when they are grossly dilated or to be able to localize a pathologic area within them.

Bladder

Examination takes place with the patient in the supine position. The empty bladder is neither percussible nor palpable. In the markedly distended condition, occurring with chronic obstruction, the bladder may be visible as a large abdominal mass rising out of the pelvis. Under ordinary circumstances, the bladder is not percussible until it contains approximately 150 mL of urine. Persistence of a low abdominal mass following emptying of the bladder by catheterization documents the extra vesical nature of the lesion.

Penis

The penis is examined with the patient in the upright position facing the examiner, who is sitting in a chair, or with the patient in a supine position on the examining table. If the patient is not circumcised, the foreskin should be retracted so that underlying glans, urethral meatus, and inner aspect of the foreskin can be visualized. Additional evaluation of the meatal caliber may be carried out with appropriate instrumentation, if indicated. A recommended means of evaluating the presence of a meatal obstruction is observation of the patient voiding.

Scrotum

Examination of the scrotum is carried out in conjunction with examination of the penis and with the patient in the positions noted above. Some patients are extremely sensitive to examination of this area, and the examiner may elicit a vasovagal response from the patient who is in the standing position. Careful observation and palpation of this region and the use of a small flashlight to transilluminate lesions offer the best means of making an accurate diagnosis.

Scrotal masses have a variety of causes (Fig. 38-5). An indirect *inguinal hernia* may present as a scrotal mass. The enlargement extends up into the inguinal region and often may be reduced with the patient in the supine position.

Epididymitis

Acute *epididymitis* is commonly a result of retrograde extension via the vas deferens from a focus of infection in the prostate, urethra, or bladder. It may follow prostatic massage or instrumentation. The patient is frequently febrile and has a typical straddling gait to minimize contact with the inflamed scrotal contents. The scrotum is exquisitely tender and may contain a mass. The overlying skin is red and edematous. *Chronic epididymitis*, which is a result of incompletely resolved acute epididymitis, presents with an indurated mass that may or may not be tender. *Tuberculous epididymitis* is characteristically nontender, stony hard, and may be associated with irregular indurated beadings of the vas deferens. A *sterile* or *chemical epididymitis* may occur with the retrograde extravasation of fluid into the epididymis secondary to marked increase in intraabdominal pressure associated with heavy lifting while the bladder is distended. The urine is usually sterile in this situation, in contrast to the pyuria often present with acute epididymitis from infection.

Varicocele

There is a predilection for the left side, presumably because the left spermatic vein drains into the left renal vein, resulting in higher venous pressures than on the right. Acute onset of a left varicocele after the age of 40 suggests left renal vein occlusion, commonly related to renal tumor. A varicocele on the right side may be secondary to vena caval obstruction or occlusion. This occasionally occurs with tumor thrombus from a renal cancer (hypernephroma) invading and involving the lumen of the vena cava.

The usual type of left varicocele is observed best with the patient in the standing position facing the examiner. The characteristic “bag of worms” appearance and feeling of the scrotum is noted. With the patient in the supine position, the varicocele collapses and usually cannot be palpated. The presence of a varicocele in a patient being examined for infertility can be a significant finding. This is especially true if the patient has a low sperm count with reduced motility of the sperm and changes in sperm morphology.

Hydrocele

Primary idiopathic hydrocele can be unilateral or bilateral and represents an increased collection of fluid between the tunica vaginalis and its contents. It is manifested as a nontender mass that is translucent and often obscures palpation of the testis and epididymis. It is ventral, superior, and inferior to the testis and may become very large and symptomatic. *Secondary hydroceles* are the consequence of serous effusion from epididymitis, tuberculosis, trauma, mumps, and occasionally tumors. Aspiration is contraindicated if tumor is suspected. A *communicating hydrocele* is present in a patient with a patent processus vaginalis. In this situation, the scrotal mass fills and empties with varying positions and, in small children, with crying.

Spermatocele

This is a cyst of an efferent ductule of the rete testis. It originates at the head of the epididymis and presents as a transilluminable cystic mass discrete from the testis. Spermatoceles are often bilateral and may be multiple. They are of no consequence except when they reach large proportions or become symptomatic with pain or heaviness. Simple cysts of the epididymis also may occur.

Testicular Tumor

A palpable nodule within the testis represents a malignant tumor until proven otherwise. The lesion is usually firm and nontender and is not transilluminable. It is often first noted by the patient or on a routine examination of this area. Ultrasonography of the scrotum may help in defining the size and location of the lesion. Prompt surgical exploration through an inguinal incision is indicated. One cannot stress enough the importance of being able to detect small lesions of the testis by careful palpation and the significance and urgency of proper management.

Mumps Orchitis

The lesion usually occurs in postpubertal males with acute parotitis. It commonly becomes apparent as the parotitis subsides, 7 to 10 days following the onset of the illness. Patients experience high fever, malaise, and marked testicular swelling. Orchitis can be caused by other viruses as well.

Torsion of the Testis Appendages

Torsion of the testis generally refers to torsion of the spermatic cord with a characteristic rotation of the testis (Fig. 38-6) and often an anterior presentation of the epididymis. The patient presents with sudden onset of pain accompanied by scrotal enlargement and edema. The testis, which appears to be elevated because the cord is shortened by twisting, is exquisitely tender. The differential diagnosis between acute epididymitis and torsion of the testis may be difficult. Isotopic scanning of the testes or Doppler ultrasonography may disclose the characteristic ischemia of the involved testis, but false-negative studies are problematic. Because prompt treatment is important to save the testis from infarction, immediate exploration is appropriate, knowing that the intraoperative diagnosis of acute epididymitis may be necessary on occasion to prevent loss of the testis in patients with torsion. Approximately 4 h of torsion appears to be the limit after which the testis is irreversibly damaged. Because of the high incidence of bilateral anatomic predisposition to torsion, bilateral orchidopexy should be performed when surgical treatment for unilateral torsion is undertaken.

Torsion of the testis can occur when there is a long mesorchium connecting it to the epididymis. In such a case, the testis, which is likened to a clapper in a bell, can twist inside the tunica vaginalis and undergo infarction. In this circumstance, the epididymis may lie in its normal position, and the testis, which is enlarged, tender, and hard, is not drawn up high in the scrotum. Another condition of torsion involves the appendix testis. This normally present embryologic remnant lies at the upper pole of the testis. When it undergoes torsion, the patient presents with a painful lump on the surface of the upper pole of the testis. Transillumination before major swelling occurs may reveal a characteristic “black dot” sign, which is visualization of the infarcted appendage. Since testicular tumors are rarely tender, they are not likely to be confused with torsion of the appendix testis.

Digital Rectal Examination (DRE)

The *prostate gland* can be evaluated best after the patient has emptied his bladder, and it is preferable for the patient to be bending forward over the examining table or bed during the examination. If the patient is unable to stand, he should be examined while lying on his side with knees flexed and facing the standing examiner. If the examiner is right-handed, the patient should lie on his right side. Rectal examination should be carried out carefully, with very little pressure exerted while passing the finger through the anus and while the examining finger is on the prostate gland. Exertion of pressure in this area may produce considerable discomfort and limit the examination. The normal prostate is about 2.5 to 3 cm wide with a 0.5-cm- deep sulcus between the two lobes. It is about 3 to 4 cm from apex to base. Rectal examination is inadequate to define the precise size of the gland; a more accurate estimate by ultrasonography is not routinely warranted.

The consistency of both the normal prostate and the prostate with benign hyperplasia is firm but not hard. Carcinoma of the prostate feels hard, indurated, or nodular. Crepitations are related to multiple prostatic calculi. With benign hyperplasia, the gland can be delineated from surrounding tissues, and the contours are smooth. With advanced carcinoma of the prostate, the surrounding tissues are frequently fixed to the gland, thus eliminating the usual discreteness of the prostatic contour. Extensions of the tumor may proceed into the seminal vesicle area superior and lateral to the prostate. Because 90 percent of carcinomas arise in the posterior portion of the gland, many are readily detectable on rectal examination as induration or a nodular mass. Many men with small prostate cancers have a normal DRE. Acute suppuration of the prostate is accompanied by tenderness or fluctuation and requires an atraumatic examination to avoid dissemination of the infection.

Female Urethra

Pelvic examination of the female patient in the lithotomy position is essential for evaluation of the lower urinary tract. Careful visualization of the urethral meatus can be carried out in this position. The presence of a cystocele or urethrocele can be determined with the patient bearing down and, during examination, with the patient standing. A urethral diverticulum may be detected by gentle pressure against the anterior vaginal wall, milking the urethra from the bladder to the meatus. Pressure in this region against the symphysis may result in release of purulent material or urine from a diverticulum of the urethra.

Urinalysis

A carefully obtained, clean-voided midstream specimen from men and women is usually adequate. A midstream specimen from the circumcised male, without prior cleansing of the penis, is usually satisfactory for culture and microscopic examination. Occasionally, catheterization of the female patient is required to obtain an uncontaminated specimen. The specimen should be examined while fresh, since refrigeration leads to sedimentation of phosphates and storage in a warm environment results in deterioration of formed elements and growth of bacteria.

A clear urine is usually normal. A cloudy urine may be normal if a few drops of dilute acetic acid render it clear, as in phosphaturia. The color may indicate disease, ingestion of certain foods, or administration of medications. Pyridium results in an orange-red color. Methylene blue incorporated in urinary medications leads to varying shades of green and blue. Pink urine may be due to beets or food dyes or to medications such as selenium. Various degrees of bluish gray and brown may be associated with acute porphyria, and bile is readily discernible. The degree and origin of bleeding often can be best estimated by examination of the gross urine specimen.

The odor is often characteristic. A mousy odor that accompanies infection is related to formation of ammonia. A pungent odor is characteristic of necrotic bladder tumors or calculous pyonephrosis. Certain medications, such as ampicillin, have a characteristic and readily identifiable odor in the urine.

Screening examination routinely includes testing for the presence of blood, albumin, glucose, and acetone and determining pH. Because there are no proteins in the secretions of the male genital organs, albuminuria cannot be ascribed to them. The various “dipstick” tests are usually quite reliable and sensitive.

Casts, crystals, and clumps of epithelial cells are identified under low-power microscopic examination, whereas the nature of cells, crystals, and bacteria is determined under higher powers. Following centrifugation, staining with methylene blue facilitates the diagnosis of bacteriuria. Cytologic examination of urine for exfoliated cells may be helpful and may demonstrate clusters of cells with nucleocytoplasmic disparity characteristic of malignancy. Flow cytometry may add additional information regarding possible neoplastic changes. The identification of bacteria in an unspun urine specimen is strongly suggestive of the presence of a urinary tract infection rather than contamination and usually represents more than 100,000 bacteria per milliliter.

Genital Secretions

Urethral Discharge

Collection is accomplished on a glass slide before the patient urinates. Noninfected urethral and prostatic secretions are usually whitish and opalescent; infected secretion is usually yellow and purulent. A heat-fixed and Gram-stained specimen identifies organisms. Gonococcal urethritis is diagnosed by the presence of gram-negative intracellular diplococci. Examination of the wet specimen is helpful in identifying trichomonads. Confirmation of infection is established by adequate culture of the urethral discharge.

Expressed Prostatic Secretions (EPS)

The specimen is obtained by gentle massage of the gland and is indicated if symptoms suggest prostatic inflammation. The acutely inflamed prostate should not be massaged. Examination of the normally opalescent fluid reveals a few white blood cells per high-powered field and tiny refractile cephalin bodies. Epithelial cells are present in small numbers. Seminal vesicle secretion presents as a gelatinous and fibrinous tree-shaped cast of the seminal vesicle containing strands and granules. In the presence of prostatic inflammation, the secretions become granular, and many white blood cells (710 per high-powered field) and large clumps are readily seen. A urine culture after prostate examination is required to determine if the inflammation is due to bacterial prostatitis.

Semen Analysis

For purposes of standardization, the specimen should be obtained by masturbation and collected in a dry container after 48 h of sexual abstinence. Examination, performed less than 1 h after ejaculation, normally reveals a volume of 3 to 5 mL with liquefaction complete within 1 h. There should be over 20 million spermatozoa per milliliter with at least 60 percent demonstrating motility and 60 percent appearing morphologically normal. More sophisticated studies of semen may be performed when indicated for the study of infertility.

Endoscopy

Cystourethroscopy

This investigative procedure should be performed in most patients with gross or microscopic hematuria and in selected patients with symptoms of lower urinary tract obstruction or infection. It often can be performed as an office procedure. Flexible and rigid instruments are available. Instillation of a local anesthetic agent into the urethra makes the procedure tolerable. The smallest-caliber instrument that will provide adequate visualization of the bladder and urethra is used. The presence of very small tumors of the bladder can be detected readily, as well as small calculi, the configuration and location of the ureteral orifices, prostate size, urethral strictures or valves, and other pathologic conditions that may be present in the bladder or the urethra.

Ureteropyeloscopy

The ureteral orifice can be dilated, and under direct vision, the full length of the ureter and much of the renal pelvis and calyceal system can be visualized with special instruments. Stones can be manipulated and fragmented with ultrasonic or electrohydraulic probes or with laser, and lesions can be biopsied and treated. This is performed under anesthesia.

Therapeutic Catheterization

An indwelling catheter affords temporary relief of obstruction. Because maintenance for more than 3 days is usually attended by bacteriuria, external collection, such as condom drainage, is more appropriate to improve nursing care of the incontinent patient who has no obstruction. Relative contraindications to urethral instrumentation include acute cystitis, urethritis, prostatitis, and a coagulopathy. In these instances, it is best to pretreat the patient for the infection or bleeding disorder. An exception to the rule is that if obstruction is the predisposing cause of infection, e.g., BPH with retention, immediate drainage by catheterization. Temporary drainage during the acute infection stage may be carried out by suprapubic tap and the insertion of a small polyethylene or plastic tube connected to constant drainage. Therapeutic instrumentation may be applied in the endoscopic removal of calculi or foreign bodies; biopsy or excision of tumors, cysts, or other obstructive lesions; drainage of prostatic abscesses; dilation or incision of urethral strictures or valves; and transurethral removal of prostatic obstruction. Institution of catheter drainage may represent the most important therapeutic measure in obstructions secondary to prostatic hypertrophy or impacted ureteral calculus. Endoscopic insertion of a ureteral stent to bypass an obstructing ureteral stone with concomitant infection can be a lifesaving maneuver.

Catheterization should be performed with aseptic technique and minimal trauma. For routine catheterization in adults, a #16 French or #18 French urethral catheter is the easiest to pass and the least traumatic. Traumatic catheterization may lead to bleeding and urethral strictures. After traumatic catheterization, closed-system drainage should be instituted and efforts made to improve urine output.

Percutaneous Nephrostomy

This procedure is carried out under local anesthesia as a diagnostic or therapeutic measure. It is used most often to relieve obstruction of the kidney secondary to stone or tumor. In the case of extensive pelvic malignancy (rectal, prostate, etc.), percutaneous nephrostomy may reverse renal failure caused by ureteral obstruction. Stones, pelvocalyceal lesions, and congenital ureteropelvic obstruction can be treated under general anesthesia in an antegrade fashion. An indwelling catheter can be left in place if the obstruction is not relieved.

Imaging Studies

Excretory Urography

Certain intravenously administered organic substances are filtered and excreted by the kidney. When rendered opaque by iodination, they opacify the renal parenchyma and collecting system. These agents are potent nephrotoxins and should be used with caution in patients with compromised renal function. It is important to hydrate patients by encouraging fluid intake following examination. Occasionally, the study is followed by anuria and renal failure in elderly, dehydrated, debilitated patients who have evidence of renal dysfunction. The use of other imaging studies is advised for this patient population. Nonionic contrast agents decrease the risk of an allergic reaction, but they are significantly more expensive.

Excretory urography provides excellent visualization of the kidneys (see Fig. [38-2](#)), ureters, and bladder. It is an excellent initial imaging study in patients with hematuria or suspected stone.

The adult male kidney averages 13 × 6.2 cm on pyelography. The left kidney is usually a few millimeters longer and broader. The kidneys of females are approximately 5 mm smaller in both dimensions. The upper pole of the left kidney lies at mid T12, whereas the right kidney is a half vertebral body lower. In the upright position, the kidney descends one vertebral body. If lines are drawn through the uppermost and lowermost calyces of the two kidneys, they should parallel the lateral margins of the psoas muscles (Fig. [38-7](#)). Any deviation from this axis suggests the presence of a pathologic condition or a variant of normal. The calyces should be adequate in number and delicately cupped. The infundibula are straight and fine. The pelvis points medially and describes a smooth curve without redundancy or tortuosity. There may be several areas of incomplete ureteral visualization due to peristalsis. The diameter of the ureter usually does not exceed 4 to 5 mm. Late upright films should demonstrate emptying of the collecting system and visualization of the urinary bladder. On supine films the contrast material may pool in the fundus of the bladder and lead to the erroneous interpretation of an outlet filling defect. A postvoiding film is helpful in revealing the lower ureter and indicating the amount of residual urine.

Nephrotomography

More detailed and accurate visualization of the kidney and pelvocalyceal system is available by this excretory urographic technique. Several “slices” of the kidneys are obtained, beginning posteriorly and advancing anteriorly. This eliminates the overlying gas and fecal material in the bowel. Lucent areas such as fat and cysts are more readily identifiable. The poorly prepared patient may be evaluated more satisfactorily than with the standard excretory urography.

Retrograde Pyeloureterography

This study may be performed during cystoscopy by catheterization of the ureters and is indicated to further evaluate lesions of the pelvocalyceal system and ureter. “Filling defects” of these structures may not be evaluated adequately by excretory urography. Urine may be collected from the kidney for cytologic study and also to determine differential renal function. Instillation of a contrast medium provides a more detailed visualization of these structures and is best done under fluoroscopic control. Improved techniques of excretory urography and nephrotomography have significantly decreased the need for retrograde pyeloureterography. But patients who manifest allergic response to intravenous contrast materials and patients with nonvisualizing or nonfunctioning kidneys can still be evaluated by this method.

Cystourethrography

The bladder and urethra may be evaluated by antegrade or retrograde studies. Fluoroscopic examination of the patient voiding may reveal the dynamics of micturition and evidence of obstruction or reflux of urine. Voiding cystourethrography is a useful diagnostic tool in the evaluation of children with voiding problems or recurrent urinary tract infections to rule out vesicoureteral reflux or urethral obstruction. Isotopic cystograms may be employed for the evaluation of reflux in children and offer decreased radiation exposure.

Antegrade Pyelography

Percutaneous insertion of a small catheter into the pelvocalyceal system may be a valuable diagnostic and therapeutic tool. The procedure is performed under local anesthesia and with fluoroscopic or ultrasonic control. Once the small tube is placed properly within the drainage structures of the kidney, adequate drainage of an obstructed, infected pyohydronephrosis may be instituted, and adequate x-ray visualization of the upper urinary tract may be obtained. Occasionally, a ureter cannot be catheterized from below and may have to be visualized by this technique. It is an extremely valuable technique to use in the seriously ill, toxic patient. Adequate drainage may convert the situation from a surgical emergency to a relatively simple elective procedure in a nontoxic patient. Stone manipulations and ureteral dilations also may be carried out. The percutaneous tract can be dilated, and an instrument can be passed into the kidney, allowing for direct-vision stone fragmentation (nephrostolithotripsy).

Renal Arteriography

Percutaneous transfemoral renal arteriography is useful in the evaluation of possible renal vascular hypertension, and therapeutic dilatation of narrow arteries can be performed (angioplasty). The technique also can demonstrate vascular lesions and thus define the surgical approach. With congenital anomalies of the kidney, arteriography is also helpful, since vascular anomalies frequently coexist and have an important bearing on surgical treatment. Arteriography is also applied to the differential diagnosis of renal masses; a characteristic neovasculature and pooling of opaque material are noted in tumor vessels within the parenchyma, diagnostic for a hypernephroma (Fig. [38-8](#)). Therapeutic infarction of kidneys also may be carried out by this approach.

Digital Subtraction Angiography

This technique provides a means of visualizing the arterial supply of the kidneys on an outpatient basis, with decreased morbidity as compared with standard renal arteriography. A bolus of contrast material is injected intravenously instead of intraarterially, and a computerized subtraction system provides clear visualization of the renal arteries and their branches, as well as of the aorta and other abdominal visceral arteries. Renovascular pathology, such as stenosis or anomalies, can be seen, as well as the presence of multiple renal arteries, which might preclude use of a kidney for donor transplantation.

Vena Cavography

Percutaneous catheterization of the femoral vein with instillation of contrast material provides adequate visualization of the inferior vena cava. This study is helpful in evaluating patients with carcinoma of the kidney, although it has largely been replaced by magnetic resonance imaging (MRI). Intrinsic involvement and obstruction of the vena cava and renal veins may be present with carcinoma of the kidney. Preoperative evaluation with this study may help determine the most suitable type of surgical procedure.

Renography and Renal Perfusion Scan

A number of isotopes, most of which are gamma emitters, provide an isotopic evaluation of the function and drainage of each kidney. Dehydration, sodium depletion, and tubular blocking agents may alter the curves, and technical factors such as placement of detectors, dose of isotope, and paravenous injection may complicate interpretation. When properly performed, the renogram can be helpful in recognizing nonfunction or reduced function and delayed excretion seen with obstruction. A furosemide (Lasix) challenge may demonstrate delayed "washout" indicative of obstruction. After renal arterial surgical procedures or homotransplantation, renography provides a simple and dependable method of evaluating the vascular status of the organ. Illustrative curves are shown in Fig. [38-9](#). The use of properly selected isotopes may provide more information about renal perfusion, drainage, morphology, and differential renal function.

Ultrasonography

Improved technology and experience in interpretation have increased the accuracy and applicability of this noninvasive diagnostic modality. The differential diagnosis of renal and scrotal masses has been made easier with ultrasonography. Cystic renal lesions may be readily distinguished from solid tumors. The technique also has provided a means of accurately guiding needles into lesions for biopsy or fluid aspiration. Renal cysts may be punctured and aspirated percutaneously and contrast material instilled into the cystic cavity under ultrasonic control for diagnosis and therapy. The noninvasive character of the procedure and the lack of radiation exposure and intravenous injection of contrast material make it a useful outpatient screening test. This is especially important in children, in pregnant women, and in those with compromised renal function. Obstructive uropathy can be ruled out when contrast excretory urography is contraindicated or is not feasible. Transrectal ultrasonography provides a means of early detection and staging of prostatic carcinoma (Fig. [38-10](#)). Biopsies of the prostate can be obtained with ultrasound guidance as an outpatient procedure without an anesthetic.

Computed Tomography (CT)

This is one of the most useful and accurate means of evaluating intraabdominal pathology, and it has significantly enhanced evaluation of patients with renal, ureteral, bladder, prostate, and testicular lesions. In some instances it has replaced other modalities such as renal cyst aspirations, angiography, and lymphangiography as a means of diagnosing and staging disease as well as of following patients after a diagnosis has been established and therapy instituted. It is performed with and without contrast material, and lesions of the kidneys can be readily identified as to size, location, extent, and density. Therefore, cystic lesions can be differentiated from solid lesions. Nonopaque filling defects of the urinary tract can be separated readily into stones or soft tissue lesions. Tumor extension, lymph node involvement, and metastases to other organs in the abdomen, chest, and brain can be demonstrated by CT scans.

Magnetic Resonance Imaging (MRI)

This imaging study, which requires no ionizing radiation, is applied with increasing frequency for the diagnosis of urologic pathology. The use of contrast agents and MR angiography have increased the diagnostic yield.

Urodynamic Tests: Pressure-Flow, Uroflow, Cystometrics, Urethral Pressure Profiles, and Sphincter Electromyography

These studies are useful diagnostic tools for the evaluation of some patients with micturition dysfunction, including problems from obstruction, neurologic disease, functional disturbances, and dyssynergia. These tests are most commonly indicated to evaluate lower urinary tract symptoms (LUTS) when initial evaluation fails to establish a clear diagnosis.

Biochemical Tests and Radioimmunoassays

Evaluation of renal function, hypertension, electrolyte disturbances, calculous disease, impotence, and neoplasms of the genitourinary tract requires studies involving the use of the biochemistry and radioimmunoassay (RIA) laboratories. Examples include blood urea nitrogen, creatinine, creatinine clearance, renin, electrolyte, calcium, phosphorus, serum testosterone, acid and alkaline phosphatase, *prostate-specific antigen (PSA)*, beta subunit of human chorionic gonadotropin, α -fetoprotein, and parathormone determinations.

BLADDER FUNCTION AND DISORDERS

Anatomy and Physiology

Normal micturition requires two simple functions: storage and evacuation of urine. The lower urinary tract includes the bladder, bladder neck, proximal urethra, and external urethral striated sphincter. Smooth muscle in the bladder wall relaxes with filling and accommodates large volumes of urine at low intravesical pressures. The bladder neck and proximal urethra remain closed during urine storage by a combination of mechanisms. First, alpha receptors on bladder neck smooth muscle cells are activated by the sympathetic nervous system, and an increase in muscular tone helps maintain a closed outlet. Second, the proximal urethra shares excursions in intraabdominal pressure exerted on the bladder to avoid an intraluminal pressure difference between the bladder and the urethra. Hypermobility or descent of the bladder neck and base out of a retropubic position can lead to stress urinary incontinence. Third, an increase in external urethral sphincter muscle contraction occurs with transmitted abdominal pressure to the bladder (the guarding reflex) and closes the distal urethra. Normal micturition requires coordination between the bladder and outlet structures whereby a bladder contraction and drop in outlet resistance occur in rapid sequence. The storage and evacuation of urine are regulated by the nervous system.

Smooth muscle cells in the bladder are innervated by short postganglionic axons from ganglia located in the adventitia of the bladder. This short axon system receives afferent input from the pelvic plexus. The pelvic plexus integrates sympathetic and parasympathetic outflow via the hypogastric and pelvic nerves, respectively; these are mixed sensory and motor nerves. The pudendal nerve innervates the external urethral striated sphincter and provides sensory pathways from the penis. Motor pathways to the bladder (pelvic nerve) and external urethral sphincter (pudendal nerve) originate in the S2 to S4 segments of the sacral spinal cord. Sympathetic fibers in the hypogastric nerve arise from T9 to L2 spinal cord segments. The sensations of bladder filling, distention, and pain are mediated by the hypogastric and pelvic nerves. Ascending and descending pathways converge on the dorsal and ventral horns of the spinal cord to regulate the sensory and motor function of the lower urinary tract. Neurons in the pontine reticular formation receive sensory input with bladder filling and coordinate bladder contraction with concomitant lowering of outlet resistance at the bladder neck and external striated urethral sphincter.

Diagnostic Studies

Cystometry and urodynamic testing are used to evaluate abnormalities of urine storage and evacuation. Cystometry is the recording of bladder pressure with filling (gas or water) at a constant infusion rate. Bladder sensation and smooth muscle activity during filling, capacity, and compliance are derived from the cystometrogram (CMG). The CMG is a storage study only and does not provide data on the ability of the bladder to completely evacuate urine. The normal CMG demonstrates an adult bladder capacity of 400 to 600 mL, filling pressure below 15 cmH₂O, and the first sensation of bladder filling at one-third bladder capacity. A strong continuous bladder contraction with reflex neurally mediated relaxation of the bladder neck, proximal urethra, and external urethral sphincter characterizes normal micturition. Urodynamic testing with multiple pressure recordings, urinary flow rate, and electromyographic recording from the external urethral sphincter may be needed to diagnose complex voiding dysfunction.

Neurogenic Bladder Dysfunction

The neurologic level of a lesion disrupting the nerve supply to the lower urinary tract can assist in the diagnosis of neurogenic bladder dysfunction.

Sacral and Intrasacral Disorders

Lesions of the conus medullaris, cauda equina, or pelvic nerves related to trauma, congenital lesions, or neoplasia can disrupt sensory and motor innervation of the bladder and external urethral sphincter (Fig. [38-11](#)). Occult or overt myelodysplasia represents the most common congenital lesion. Sacral agenesis, partial or total, and other anomalies in the caudal regression syndrome often are associated with neurogenic bladder dysfunction. Operation for imperforate anus can damage the pelvic plexus. Most patients suffer from recurrent urinary tract infection and overflow incontinence.

Children with spina bifida usually present with an areflexic bladder and open (fixed) bladder neck. If the meningocele or myelomeningocele has damaged the pathways from the pontine reticular formation, one might see detrusor sphincter dyssynergia, with the potential danger of renal compromise from high lower-urinary-tract pressure (Fig. [38-12](#)). No correlation exists between the neurogenic bladder dysfunction and the level of the vertebral defect. Therefore, urodynamic testing in children with myelodysplasia plays a pivotal role in diagnosis and treatment selection.

Voiding dysfunction following radical pelvic surgery (e.g., abdominoperineal resection, proctocolectomy, radical hysterectomy, low anterior resection) is common and usually results from direct injury to the innervation of the lower urinary tract. Most patients recover function within 4 to 6 months, and conservative management is warranted. Urodynamic testing reveals a smooth, large-capacity bladder with no reflex activity. Intermittent self-catheterization at periodic intervals should be instituted early to prevent bladder overdistention, recurrent infection, and overflow incontinence. Long-term antimicrobial therapy is not recommended because of the potential development of resistant organisms. The goals of treatment are preservation of renal function and establishment of urinary continence.

Sensory disorders of bladder innervation may cause the loss of sensation to bladder filling. Tabes dorsalis, posterior spinal cord degeneration, herpetic infection, or diabetes mellitus often leads to chronic overdistention of the bladder and high postvoiding residual urine volumes. Intermittent self-catheterization at periodic intervals is the treatment of choice in these patients.

Suprasacral Spinal Disorders

Spinal cord trauma, vascular compromise, and multiple sclerosis are the most common causes of suprasacral disorders. Spinal cord injury from penetrating trauma or a motor vehicle accident is characterized by an early period of spinal shock and bladder areflexia. A short period of indwelling bladder catheterization is appropriate until the patient is stable. As soon as possible, intermittent catheterization should be started and anticholinergic medication added when bladder function returns. The long-term use of a Foley catheter or suprapubic tube should be avoided because they often lead to chronic resistant infection, urethral erosion, fistula formation, stones, and loss of the bladder as a urinary reservoir.

After the period of spinal shock, most patients develop detrusor-sphincter dyssynergia and are at risk for loss of renal function. Intermittent catheterization, anticholinergic medication, and transurethral sphincterotomy (resection of the external sphincter) with condom catheter drainage have been the mainstays of bladder management in spinal cord injury. With time, the bladder may lose compliance and capacity, thereby placing the kidneys at risk despite frequent catheterization. Augmentation cystoplasty may be used to increase bladder capacity, restore urinary continence, and reduce high intravesical pressures.

Autonomic dysreflexia is a serious medical condition that must be recognized and treated promptly in the spinal cord injury patient. Patients with injuries at T6 or higher are at risk for this autonomic reflex, usually triggered by visceral distention. Elevated blood pressure, sweating, pounding headache, and bradycardia occur with visceral autonomic nervous system activation (bladder distention, fecal impaction, bowel obstruction). Immediate treatment includes relief of the obstruction and the use of alpha-adrenergic blocking agents such as prazosin, terazosin, or nifedipine.

Neurogenic bladder dysfunction is common in patients with multiple sclerosis. This diffuse demyelinating disorder can present with a variety of voiding problems due to the haphazard location of the plaques in the white matter of the spinal cord. Urgency and urgency incontinence are common symptoms. Many patients with multiple sclerosis have detrusor-sphincter dyssynergia with residual urine and must be treated with intermittent catheterization and drugs. Because the disease is characterized by periods of exacerbation and remission, conservative management should be followed along with periodic urodynamic testing.

Supraspinal or Suprapontine Disorders

Coordinated bladder contraction and urethral relaxation are directed by the pontine “micturition center” in the brainstem. Neurologic lesions above the pons usually leave the “micturition reflex” intact. When the lesion affects voiding, it generally results in a loss of voluntary control over urination. Cerebrovascular accident, normal-pressure hydrocephalus, tumor, dementia, and Parkinson's disease represent suprapontine lesions. Urinary incontinence from bladder hyperreflexia presents with complaints of urgency and frequency. Since the lesions are above the pons, detrusor- sphincter dyssynergia does not develop, and renal impairment occurs only when chronic indwelling catheterization is used to manage the urinary incontinence. These patients have lost the ability to inhibit the micturition reflex and are managed best with drugs to inhibit bladder hyperreflexia and timed voiding to keep the bladder empty. Cystometry may show uninhibited bladder contractions at low volumes, which characterize the hyperreflexia seen with suprapontine lesions.

Bladder Rehabilitation

The level and degree (partial versus complete) of a neurologic lesion will determine the pattern of voiding dysfunction and provide prognostic information. Central nervous system structures do not regain function, and complete lesions, after a period of spinal shock, become fixed neurogenic bladder disorders. Maintenance of renal function is the primary goal of management. Intermittent catheterization should be used whenever possible so that the reservoir function of the bladder can remain intact and be augmented if bladder compliance deteriorates. Supravesical urinary diversion is indicated in rare instances, and often a continent pouch or catheter-free diversion can be constructed utilizing segments of bowel. Reconstructive lower urinary tract surgery for neurogenic bladder dysfunction should be done only when a fixed and stable deficit is encountered with the potential loss of renal function or debilitating urinary incontinence.

INFECTIONS

Acute Infections

Pathogenesis

The most common port of entry for *urinary tract infection (UTI)* is the urethra. Ninety-six percent of pediatric UTIs occur in girls because of the shortness of the urethra. Similarly, in the adult, most UTIs develop in women as a consequence of the short urethra and the superimposed factors of pathogenic organisms, vaginal cell-bacterial adherence, and sexual intercourse. The normal bladder possesses defense mechanisms capable of coping with introduced bacteria. Complete emptying so that no residual urine remains is an additional defense factor. In the obstructed, inflamed, or ulcerated organ, the defense mechanisms are compromised.

Ascending infection is potentiated by any obstructive process, including obstructive uropathy secondary to prostatic hyperplasia, foreign bodies, or bladder tumors. Reflux of infected urine from the bladder can lead to contamination of the ducts of Bellini and to the development of pyelonephritis, predominantly involving the medullary portion of the kidney.

UTIs occasionally are secondary to distant infections such as cutaneous abscesses that undergo hematogenous spread. Hydronephrosis and renal calculi potentiate the risk of hematogenous infection. UTIs rarely are caused by spread from contiguous intestinal infections such as appendicitis, diverticulitis, and regional enteritis.

Bacteriology

The most commonly implicated organism is *Escherichia coli*. Other organisms include enterococci, *Staphylococcus aureus* and *Streptococcus epidermidis*, *Proteus mirabilis*, *Klebsiella*, and *Pseudomonas*. Because most forms of *Proteus* split urea, resulting in an alkaline urine and precipitation of calcium, infections caused by these organisms are often associated with calculi.

Clinical Manifestations

Acute *cystitis* presents with frequency, urgency, and dysuria, with minimal systemic symptoms. The fever, if present, is usually low grade. By contrast, *pyelonephritis* is usually accompanied by chills, temperatures of 39°C, nausea, and flank pain. The diagnosis of a UTI is established by microscopic analysis of the urine demonstrating bacteria and white blood cells and by culture of the urine to establish the specific pathogens and the sensitivity to antibiotics.

Treatment

The ability to treat UTIs with drugs is enhanced by the unique ability of normally functioning kidneys to concentrate a variety of antibacterial agents in the urine. Drug selection is facilitated by culture and determination of sensitivities, but because these are usually not available in the early stages of treatment, initially the drug is prescribed empirically based on the probable pathogen. Further considerations include the age and general health of the patient, kidney function, history of allergy, previous drug therapy, relative cost of therapy, and possible complications of the drug to be administered.

Drugs that are rapidly excreted by the kidney in high concentrations are ideally suited for the treatment of *uncomplicated lower UTIs*, such as acute cystitis. Most acute UTIs infections are caused by *E. coli* and will respond rapidly to drugs that are usually administered for a period of 7 days or less. Short courses, consisting of single-dose treatment or 1 to 3 days of therapy, also have proved to be efficacious. Some drugs commonly used for the treatment of uncomplicated UTIs are sulfonamides, ampicillin, cephalexin, nitrofurantoin, nalidixic acid, trimethoprim-sulfamethoxazole, and quinolones (e.g., floxacin).

In the patient with manifestations of *acute pyelonephritis* or *sepsis*, drugs that produce high blood and tissue levels are indicated, and drugs that yield only high urinary concentrations are not adequate. Parenteral therapy may be required, and a more prolonged course of other therapy is indicated. When a *Pseudomonas* infection is suspected, the aminoglycosides such as gentamicin, tobramycin, or amikacin may be used. Combination with ampicillin or a cephalosporin provides a broad initial coverage until culture and sensitivity reports are available, at which time specific treatment can be instituted. For less severe “cases,” oral quinolones may be sufficient.

The acute infection is sometimes secondary to an anatomic factor (e.g., obstruction or stone) that may require correction to prevent recurrences. Although it is justifiable and essential to treat acutely ill patients before a causative diagnosis is made, a more complete investigation should be carried out as soon as possible. This is especially indicated in patients with complicated UTIs, such as acute pyelonephritis and sepsis. Uncomplicated cystitis in the adult female may require no further evaluation, since the excretory urogram and cystoscopic findings are usually unremarkable in such patients. The male with a single UTI requires evaluation to rule out obstruction, stone disease, or tumor because ascending infection from the urethra is not common except as related to prostatitis in the adult.

The patient who has had a UTI is at increased risk for recurrences, and repeat urinalysis and possibly cultures are indicated in follow-up. Examination of a midstream urine specimen from the male or a clean-catch urine specimen from the female should be performed before or shortly after discontinuing antibiotic therapy. If recurrences occur in the absence of any correctable anatomic abnormality, such as in cystitis in the adult female, long-term, low-dose, suppressive therapy may be required. Drugs suitable for long-term therapy include nitrofurantoin and trimethoprim- sulfamethoxazole. The length of therapy usually is a minimum of 3 months.

Gram-Negative Bacteremia

The urinary tract is a common source of sepsis. Frequently, the bacteremia follows instrumentation in the presence of preexisting UTI that has been ineffectively treated. Parenteral antibiotics should be administered as a prophylactic measure before urethral instrumentation in patients who have had an indwelling catheter for a prolonged period of time.

The patients experience a shaking chill followed by high fever, peripheral vasodilation, and hypotension. The organisms responsible are frequently resistant to the usual drugs and require large doses of parenterally administered antibacterial agents. Bactericidal drugs are preferable to prevent liberation of additional quantities of the endotoxin by the multiplying bacteria. An aminoglycoside, such as gentamicin, tobramycin, or amikacin, and ampicillin, or a similar drug used in combination, has proved to be quite effective for the management of gram-negative bacteremia.

Perinephric Abscess

This usually follows perforation of a renal infection or abscess into the perinephric tissues. In some instances, perinephric abscess may result from direct hematogenous spread or from microabscesses in the kidney that are not apparent. The perinephric infection may be secondary to and is often associated with a struvite renal stone or a staghorn calculus.

The patient presents with an extremely high fever and flank pain. Radiographs reveal nonvisualization of the psoas shadow and concavity of the spine to the side of the lesion. The kidney may be nonfunctioning, and a stone may be seen on the film. Gas can be seen within or outside the kidney shadow if the organism is gas-producing. *Emphysematous pyelonephritis* occurs more frequently in diabetic patients. CT scan provides the most accurate definition of the location and extent of the process.

Papillary Necrosis

Necrosis of renal papillae occurs in patients with diabetes, sickle cell disease, tuberculosis, and excessive ingestion of phenacetin or nonsteroidal anti-inflammatory drugs. In addition to the symptoms related to the associated acute infection of the kidney, renal colic may result from ureteral obstruction by the sloughed papillae. Diagnosis is made radiographically by demonstrating sloughed papillae or by the characteristic appearance of the intravenous pyelogram showing loss of one or more of the papillae (Fig. 38-13). When amputation of the papilla at the cortical- medullary junction is complete and contrast material surrounds the papilla floating free in the calyx, a “ring sign” is present. Loss of renal function may attend papillary necrosis, but recovery usually occurs, and initial measures are directed toward supporting the patient during the acute episode. Specific antibacterial treatment for the involved organism is administered. If the patient remains febrile or septic and there is ureteral obstruction, endoscopic, percutaneous, or open surgical intervention may be required to remove the obstructing soft tissue and to provide drainage.

Urethritis

Acute urethritis is usually a sexually transmitted disease (STD). Included in the acute types of urethritis are those caused by gonorrhea, *Ureaplasma*

urealyticum, *Chlamydia*, and *Trichomonas vaginalis*. Diagnosis is established by microscopy of the gram-stained discharge and confirmed by adequate culture techniques. Reiter's syndrome, consisting of the triad of arthritis, nongonococcal urethritis, and conjunctivitis, is considered by many to be of venereal origin and should be included in the differential diagnosis.

Gonorrhea is a common STD. The incubation period is 7 to 9 days, after which the symptoms of acute urethritis are manifested. Urethral stricture is a sequela if the original infection is not adequately treated. Metastatic gonococcal infections include arthritis, tenosynovitis, bursitis, ophthalmitis, perihepatitis, and endocarditis. Gonococcal arthritis, which represents the common manifestation of metastatic involvement, is polyarticular in 80 percent of patients and accompanied by tenosynovitis in 50 percent of the cases. The diagnosis of gonococcal infection is established by demonstration of *Neisseria gonorrhea* (gram-negative diplococci) intracellularly in the urethral discharge and by culture of the organism. Because syphilis and gonorrhea are contracted by the same exposure in some cases, the latter should be excluded by serologic tests. Gonococcal urethral infections are best treated by adequate dosage of penicillin-type drugs. Resistant strains have appeared with increasing frequency; spectinomycin and other antibiotics have been effective in these cases.

Nonspecific urethritis (NSU), or nongonococcal urethritis (NGU), is the most common venereal disease in the male. This is manifest by a translucent, sticky drop on the underwear or liquid that can be milked from the urethra on a daily basis. The organisms most commonly implicated in this infection are *Chlamydia trachomatis* and *U. urealyticum*, which usually can be treated satisfactorily with the tetracyclines. Recurring infections indicate the need to treat the patient and his sexual contacts. *Trichomonas* infections of the urethra and prostate are best treated with metronidazole (Flagyl), including treatment of sexual partner(s).

Bacterial Prostatitis

Acute bacterial prostatitis is usually caused by the same organisms that produce UTIs, and frequently these organisms can be found in the urine. The bacteria may enter the prostate through the urethra and may lie dormant in the many prostatic ducts. Active infection may be precipitated by urethral trauma or instrumentation. Acute prostatitis is particularly common in young, healthy males and is also seen in men in the benign prostatic hyperplasia age range. The organisms associated with urethritis also may be the pathogens of acute prostatitis.

Initial symptoms may be mild and consist of perineal discomfort with increasing frequency, initial hematuria, dysuria, and occasional retention. With progression, the perineal pain becomes continuous and severe and is attended by chills, high fever, malaise, and prostration. The prostate is usually very tender, and examination should be gentle. Leukocytosis is characteristic, and bacteremia is common. In extreme cases, the gram- negative bacteremia may be associated with endotoxic shock. Liquefaction necrosis of the local infection results in the formation of abscesses that may rupture into the urethra spontaneously or as a result of catheterization or transurethral resection.

Culture of the urine should precede administration of antibacterial agents, but these should not be withheld until the culture reports are available. In mild cases, an oral broad-spectrum antibiotic such as a quinolone should be administered for a minimum of 10 to 14 days. Parenteral antibiotics are indicated for patients with chills, high fever, malaise, and/or hypotension. If necessary, the drug may be changed at a later date. The acutely ill patient should be at bed rest, hydrated, and treated with antibiotics until the acute phase of the process subsides. Patients with retention should be catheterized gently with a small indwelling catheter or a percutaneous suprapubic tube. Rectal examination reveals a tender and sometimes warm prostate and should be performed infrequently and gently with great care to avoid dissemination of the infection. Several days of appropriate antibacterial treatment should result in symptomatic improvement and return of temperature to normal. Persistence of symptoms and fever, coupled with a diminished size and force of the urinary stream, suggests the presence of an abscess that may require drainage. CT scan aids in localization.

A significant number of men with the symptoms of chronic prostatitis (pain, frequency, urgency) have sterile prostate fluid and no evidence of inflammation in the expressed prostatic secretions. These men are believed to have chronic pelvic pain syndrome termed *prostatodynia*.

Epididymitis

This condition is characterized by rapid progression of pain and swelling of the epididymis and is usually associated with leukocytosis, fever, malaise, and chills. A UTI often present, with symptoms of frequency, urgency, and dysuria. When this is the case, the offending pathogen can be cultured from the urine. When this is not the case, it is difficult to identify the organism. There can be an underlying prostatic infection. The differential diagnosis includes acute torsion of the spermatic cord or hydatid of Morgagni, spontaneous hemorrhage of a testicular tumor, trauma, or thrombosis of the pampiniform plexus. A radioisotopic scan of the scrotum or ultrasound examination may aid in the diagnosis.

Acute epididymitis can occur after open prostatectomy or might be related to infection secondary to long-term urethral catheterization. Common organisms producing this infection are gonococcus, *E. coli*, and *C.*

trachomatis. The latter organism is difficult to culture and is usually surmised to be present.

Treatment consists of symptomatic measures such as elevation of the scrotum, the use of a scrotal suspensory or an ice bag, and broad- spectrum antibiotic therapy with ampicillin, tetracycline, or aminoglycosides. Epididymoorchitis and abscess formation may occur and require surgical intervention for drainage or excision.

Acute onset of epididymal swelling and tenderness following scrotal trauma or strain in lifting a heavy object provides a problem from a compensation standpoint regarding the entity described as *traumatic epididymitis*. Treatment is usually as noted for infectious epididymitis, but the antibiotic therapy may be withheld. Infection often cannot definitely be ruled out despite the negative urinalysis and cultures.

Chronic Infections

Bacterial Prostatitis

Chronic bacterial prostatitis is characterized by recurring UTIs, low back and perineal discomfort, urinary frequency, urgency, and dysuria. The symptoms are variable, and there may be long periods of remission between symptomatic episodes or signs of UTIs. The prostate usually is not tender, but the expressed prostatic secretion obtained by massage usually shows an increased number of white blood cells with some in clumps. The properly collected specimen may reveal the offending organism on culture.

The commonly prescribed antibacterial agents that are used for the treatment of a UTI or acute bacterial prostatitis may be ineffective in the management of chronic bacterial prostatitis. Failure to achieve adequate levels of these drugs in the nonacutely inflamed prostate may account for this phenomenon and results in a low cure rate for this troublesome condition. The quinolones have largely resolved this problem, although several weeks of therapy may be required. Surgical intervention, such as transurethral resection, does not seem to benefit unless associated obstruction is present because infected prostatic tissue remains after resection.

Nonbacterial prostatitis produces symptoms identical to chronic bacterial prostatitis and may be more common. Inflammatory changes of the prostate are present, with increased white blood cells in the expressed prostatic secretions and negative cultures for bacterial pathogens. *Prostatodynia* is present if both the expressed prostatic secretions and cultures are negative. Symptoms may be similar to those seen with chronic bacterial prostatitis, and therapy is empirical and usually unsuccessful. Management consists of reassurance, as well as muscle relaxants, warm baths, prostatic massage, and other "symptomatic relief" measures.

Cystitis

Chronic cystitis is seldom the end result of recurrent uncomplicated acute bacterial cystitis. The inflammatory process usually extends beneath the mucosa into the submucosal and muscular layers of the bladder and may be associated with white blood cell infiltration and varying degrees of fibrosis, which compromises detrusor function and decreases bladder capacity. Predisposing factors usually contribute to the progressive changes and associated symptoms. These may include urethral stricture, bladder calculi, tumor, diverticula, neurogenic bladder dysfunction, long- standing indwelling catheter drainage, tuberculosis, parasitic infestation, such as schistosomiasis, and vesicoenteric fistula.

Chronic pyuria and often bacilluria are noted, and frequency, urgency, and dysuria are usually present unless catheterization or diversion has been instituted. The diagnostic work-up is directed toward accurate identification of the predisposing factors and the offending pathogen(s). Endoscopic evaluation of the urethra and bladder is usually carried out, as well as biopsy of suspicious lesions, cystourethrography, and cystometrics as indicated. Careful urine collections and appropriate cultures are performed to determine the predominant organisms. Unlike the *E. coli* of acute uncomplicated cystitis, the pathogen is usually one of the more resistant types such as *Pseudomonas*, *Serratia*, and cocci, and changing organisms are often encountered. When possible, the predisposing factor should be eliminated if symptoms are to be relieved and preservation of the remaining detrusor function is to be accomplished. Chronic antibacterial therapy is often required, and some permanent dysfunction is usually present after treatment.

A form of abacterial chronic cystitis has been identified and termed *interstitial cystitis*. It occurs more commonly in females than in males. The cause remains unknown, and the predominant symptoms are marked urinary frequency, urgency, and suprapubic pain. Cystoscopy may reveal a decreased bladder capacity and increased irritability. Bleeding from the mucosa may occur with bladder distention and production of the so-called Hunner's ulcer.

Biopsies of the bladder mucosa and exfoliative cytology may be indicated to rule out carcinoma in situ or dysplasia of the bladder. Cultures are usually negative, and symptomatic treatment is usually prescribed. As usually is the case in such situations, a number of medical and surgical regimens have been tried. Periodic instillations of dimethyl sulfoxide (DMSO) have proved of little value.

Epididymitis

Persistent induration of the epididymis characterizes this condition. Systemic symptoms and signs are usually absent, and tenderness in the indurated area may be minimal and intermittent. A history of previous acute epididymitis, which may be of the recurrent type, usually can be elicited. A definite cause may be difficult to establish, since culture of a pathogen may be impossible. The differential diagnosis of spermatocele, adenomatoid tumor, tuberculosis, trauma, or underlying testicular tumor has to be considered, especially in the patient in whom there is no previous history of acute infection. Ultrasound evaluation may help rule out a testicular neoplasm or spermatocele. Treatment consists of empirical antibacterial therapy, excisional biopsy, or just periodic observation depending on the size of the lesion, symptoms, and the age of the patient.

Pyelonephritis

The histologic picture of chronic pyelonephritis is characterized by a lack of specificity. Inflammatory changes associated with fibrosis and scarring are noted. Radiologically, there is evidence of loss of parenchyma, calyceal blunting, contracture, cortical scars with an irregular contour, and decreased perfusion and excretion of contrast material. Predisposing factors include vesicoureteral reflux, calculous disease, ureteral stricture, neurogenic bladder dysfunction, urinary diversion, indwelling nephrostomy, or bladder catheter. Treatment is directed toward correction of the predisposing factor and antibacterial therapy. The goal of therapy is preservation of renal function. End-stage renal disease may require dialysis and possibly renal transplantation.

Xanthogranulomatous pyelonephritis is a variant that characteristically presents in middle-aged patients with a poorly functioning kidney. There is often accompanying calculous disease and *Proteus* infection. Tumefaction can occur with rupture of the process through the renal capsule. It can be confused histologically with renal cell carcinoma because of

areas of clear cells that are lipid-laden macrophages. The lesion is occasionally bilateral, does not respond to antibiotics, and progresses to end-stage renal disease (Fig. [38-14](#)).

Urinary Tuberculosis

The incidence of urinary tuberculosis is increasing. Renal tuberculosis is the result of hematogenous spread from primary pulmonary or intestinal lesions. Although the involvement is probably bilateral and cortical in most individuals, some lesions progress to healing, whereas others assume clinical significance. Renal tuberculosis is generally undetected and asymptomatic until it ulcerates into the collecting system of the kidney. At that time, the symptoms include gross hematuria, dysuria, frequency, and pyuria with an absence of bacteria on Gram stain. Many of the symptoms are a consequence of cystitis arising from organisms localizing in the bladder secondary to the renal involvement.

Special culture techniques reveal the presence of *Mycobacterium tuberculosis*. Intravenous pyelography demonstrates calcification of caseous abscesses, ulceration of calyces, and failure of visualization of calyces filled with caseous debris (Fig. [38-15](#)). One drug therapy regimen consists of treatment with isoniazid (INH), ethambutol, rifampin, and pyridoxine. The patient is followed by repeated urine cultures and repeated pyelography, which should be performed at 3-month intervals during the first year. Nephrectomy is seldom indicated for renal tuberculosis.

Genital tuberculosis may accompany renal tuberculosis or may exist independently as a result of hematogenous spread. The epididymis is most frequently involved, and spread may take place from this source up the vas deferens to the prostate or through the scrotum in the form of a fistula. The vas deferens may become beaded with involvement. Epididymal tuberculosis may be managed successfully by chemotherapy, and epididymectomy is held in reserve.

NEPHROLITHIASIS

Nephrolithiasis is one of the more common urologic disorders, affecting approximately 5 to 10 percent of the population at some time during their lifetime. Without preventive treatment, the incidence of recurrence is as high as 50 percent within 5 years of the first stone event. Even patients with asymptomatic calculi stand a nearly 50 percent chance of becoming symptomatic within 5 years of diagnosis. Thus the economic and clinical impact of nephrolithiasis is considerable; the total cost of evaluation, hospitalization, inpatient and outpatient treatment, and indirect costs including lost wages has been estimated at \$1.83 billion annually.

The surgical management of stone disease has changed dramatically in the past 15 years. Advances in noninvasive and minimally invasive techniques have virtually eliminated the need for open stone surgery; in 1997, less than 1 percent of patients required open stone procedures. However, the need for a medical prophylactic program has not diminished. Stone recurrences after surgery can be reduced significantly by the initiation of selective medical therapy. With current diagnostic protocols, metabolic abnormalities are identified in more than 95 percent of patients, and in the vast majority of these patients, directed medical and dietary therapy can have a significant impact on the course of the stone disease.

When a patient presents to the emergency room or physician's office with symptoms attributable to a renal or ureteral calculus, the goal is to establish a diagnosis, relieve pain and/or obstruction, identify stone characteristics and anatomic features that affect management, select conservative management versus surgical treatment, and proceed with treatment if indicated. To this end, a careful history, thorough physical examination, selective laboratory tests, and radiographic studies must be obtained (Fig. [38-16](#)).

History

Patients with obstructing renal or ureteral calculi or calculi in transit from the kidney to the bladder typically present with a constellation of symptoms known as *renal colic*. Renal colic is characterized by the abrupt onset, often at night or early morning, of sharp pain originating in the flank and radiating around to the groin or testicle (labia). The sensation of pain in the groin area is attributed to the close proximity of the origins of the blood supply and autonomic nerve fibers supplying the kidney and testicle (ovary). The pain is brief, lasting a few minutes at a time, but recurs in paroxysms. Because the celiac axis innervates the kidneys as well as the stomach and small intestine, renal colic is frequently associated with nausea and vomiting, often confusing the diagnosis with gastrointestinal diseases (acute appendicitis, perforated viscous, colitis, diverticulitis).

The specific location of the pain may reflect the location of the obstructing stone: Stones in the proximal ureter or ureteropelvic junction typically cause flank pain radiating to the groin, stones in the middle ureter (overlying the bony pelvis) are often associated with anterior lower quadrant pain, and stones in the distal ureter may cause bladder irritative symptoms such as urinary frequency and urgency.

Physical Findings

Unlike the patient with an acute abdomen, patients with renal colic characteristically writhe in pain or pace the floor in an attempt to gain relief. Physical examination is remarkable only for abdominal or flank tenderness overlying the site of stone impaction. Palpation of the testes or ovaries fails to reproduce the pain and may rule out the gonads as the underlying source of pathology.

An elevated pulse and blood pressure are common due to pain and are nonspecific. If the obstructing stone is associated with infection, fever or chills may be present; noninfected stones are occasionally associated with a low-grade fever (100°F).

Laboratory Studies

The minimum laboratory evaluation of the patient with suspected renal colic consists of a complete blood count (CBC), serum electrolytes, blood urea nitrogen and creatinine, and a urinalysis. Mild leukocytosis (10,000 to 15,000/mm³) is common in patients with renal colic due to demargination of white blood cells in response to stress. A white blood cell count greater than 15,000/mm³ should alert the clinician to the possibility of obstructive or nonobstructive pyelonephritis.

An assessment of renal function is indicated to determine the impact of renal obstruction on overall renal function, to determine if intravenous contrast material can be administered safely for radiographic evaluation, and to assess fluid status. Serum electrolytes should be checked to identify any metabolic abnormalities such as acidosis, alkalosis, or hypokalemia.

The urinalysis is an important part of the laboratory assessment because the presence of microscopic hematuria may suggest the diagnosis of renal calculi. The absence of red blood cells in the urine does not preclude the presence of stones; in 10 to 25 percent of patients with an acute stone event, no red blood cells are detected in the urine, particularly in the face of complete obstruction. Mild pyuria may accompany hematuria but often reflects inflammation rather than infection. Significant pyuria, particularly when associated with fever and/or leukocytosis, should raise the suspicion of a UTI, and a urine culture should be obtained.

Additional urine parameters may be helpful in the overall evaluation of the stone patient but contribute less to the acute stone event. Urine pH may suggest the etiology of stone disease. A pH > 7 suggests the presence of urea-splitting organisms and possible struvite stones; in contrast, a pH < 5.5 is associated with uric acid stones. The presence of crystals may reveal the stone composition (hexagonal crystals in cystinuria, pyramidal crystals in calcium oxalate stones).

Radiographic Evaluation

The diagnosis of urinary tract calculi ultimately relies on radiographic confirmation. Radiographic imaging should determine the presence, number, size, location, and occasionally, composition of stones within the urinary tract. For radiopaque calculi, comprising 90 percent of upper tract calculi, a plain abdominal radiograph (kidneys, ureter, and bladder, or KUB) may demonstrate radiodensities highly suspicious for renal or ureteral stones (Fig. [38-17](#)). The addition of oblique views may increase the sensitivity of plain-film radiography by altering which dimension of the stone is penetrated by the x-ray or by projecting the stone away from the spine or bony pelvis, thereby allowing the stone to be visualized. Plain nephrotomograms may prevent overlying bowel gas from obscuring small or faintly opaque renal calculi. Table [38-1](#) lists various stone compositions in decreasing order of radiodensity.

Extraurinary calcifications may mimic urinary tract calculi and must be distinguished from renal, ureteral, or bladder calculi with the aid of oblique or lateral views that project these calcifications away from the urinary tract. Gallstones, which are characteristically multifaceted, lie in the right upper quadrant and project anterior to the kidney on a right posterior oblique view. Vascular calcifications are often curvilinear and usually identifiable. Renal artery aneurysms are typically circular with a lucent center. Pelvic phleboliths, which are calcified thrombi in pelvic veins, are generally round or oval and contain a lucent center. Phleboliths usually require opacification of the urinary tract to distinguish them from ureteral calculi.

The *intravenous urogram (IVU)* remains the mainstay of diagnosis of renal and ureteral calculi (Fig. [38-18](#)). The IVU can precisely identify calculi within the urinary tract, assess the degree of obstruction, demonstrate anatomic abnormalities that may predispose to stone formation (medullary sponge kidney, ureteropelvic junction obstruction) or affect treatment decisions (ureteral stricture, calyceal diverticulum), and suggest the presence of sequelae from the obstructing stone (loss of renal parenchyma, forniceal rupture with extravasation of contrast material, mass effect from perinephric abscess). Although radiolucent calculi are not visible on the scout film, opacification of the collecting system may demonstrate a filling defect associated with ureteral obstruction.

Careful monitoring of the IVU as it progresses is critical to the successful diagnosis of the stone and to maximize useful information about the stone and kidney. The study should be modified based on the results of early images (e.g., delayed images or oblique views or the patient in a prone position).

In the acute setting, patients with renal insufficiency or a contrast material allergy can be imaged with a plain abdominal radiograph (KUB) and renal sonography or by retrograde pyelography. Renal sonography is highly sensitive for renal calculi greater than 5 mm in diameter and for detecting hydronephrosis (Fig. [38-19](#)). The quality of the sonogram depends on patient size and imaging technique (i.e., stones in obese patients may easily be missed). Ureteral calculi are poorly imaged, particular in the absence of ureteral dilation, although distal ureteral calculi occasionally may be visualized through the window of a fluid-filled bladder.

Retrograde pyelography is indicated for the patient in whom administration of intravenous contrast material is contraindicated or when the IVU or renal sonogram is inconclusive. Ureteral calculi typically produce a filling defect in the ureter and cause a discrepancy in the caliber of the ureter above and below the defect (dilated above, narrow below).

Non-contrast-enhanced CT scans are very sensitive for detecting calculi of any composition in any location in the urinary tract (Fig. [38-20](#)). However, thin-cut (5-mm) images are necessary to ensure adequate overlap and prevent overlooking small stones. Helical CT, in which images are acquired in a spiral configuration during a single breath hold, eliminates respiratory motion artifact and allows three-dimensional reconstruction of the entire urinary tract. At some institutions, non-contrast-enhanced helical CT has replaced the IVU for the diagnosis of acute flank pain in the emergency room setting. Early results suggest that helical CT is more sensitive than IVU in detecting ureteral stones; even in the absence of a clear-cut stone on CT, secondary signs of obstruction such as perinephric stranding or circumferential soft tissue attenuation can confirm the diagnosis of a ureteral stone with a high degree of accuracy. The advantages of non-contrast-enhanced helical CT are the rapidity of acquisition, the high degree of diagnostic accuracy, and no need for intravenous contrast material; however, the cost of spiral CT currently exceeds that of IVU at most institutions.

Initial Management

The initial effort is directed toward prompt pain relief. Most narcotic analgesics, such as morphine or meperidine, provide adequate analgesia. Intravenous nonsteroidal analgesics, such as Ketorolac, are effective, but caution must be exercised with the administration of these agents because of reported cases of renal failure, particularly in patients with impaired renal function or dehydration. As with other nonsteroidal anti-inflammatory medications, Ketorolac causes inhibition of platelet aggregation and may predispose to bleeding, thereby interfering with planned surgical intervention or nephrostomy drainage.

Some investigators have advocated the use of a variety of agents aimed at reducing ureteral peristalsis and renal colic and facilitating stone passage. Intranasal desmopressin, aminophylline, indomethacin, nifedipine, and methylprednisolone have all been used to treat renal colic and facilitate stone passage with variable success. The use of these agents remains sporadic and has not gained widespread acceptance.

Immediate management of patients with renal colic depends on their response to analgesics, their ability to take oral fluids, and the presence of infection. A partially or completely obstructing stone associated with infection, as indicated by a fever, a markedly elevated white blood cell count, or a urinalysis suggestive of infection, mandates immediate decompression of the collecting system. The optimal method of draining the collecting system, whether retrograde placement of a ureteral catheter or placement of a percutaneous nephrostomy tube, is controversial. The advantage of nephrostomy tube placement includes accessibility of the external catheter for irrigation or to monitor urine appearance and output and the option of placing directly into the kidney a larger-caliber tube than is generally accommodated by the nondilated ureter. The advantage of ureteral catheterization is presumed greater patient comfort and less inconvenience with an internalized drainage tube. Regardless of the method of drainage employed, the kidney should be drained promptly and broad-spectrum antibiotics initiated. Urine cultures should be obtained from a voided specimen in the emergency room and a specimen obtained from the kidney at the time of drainage; antimicrobial coverage can be customized based on the identification and sensitivity of the cultured organisms.

The indications for hospitalizing a patient suffering from an acute stone event include failure to gain adequate pain relief with oral analgesics, inability to retain fluids or food because of nausea or vomiting, suspicion of infection, and anuria in patients with a solitary kidney (immediate decompression of the kidney is warranted). The patient may then receive intravenous narcotic analgesics, fluids, antiemetics, and antimicrobial agents if necessary.

Surgical intervention to remove the stone is undertaken if the stone is unlikely to pass spontaneously (≥ 6 mm in diameter), if the patient experiences persistent or recurrent colic or has persistent high grade obstruction (>1 week), or if occupational requirements mandate prompt treatment (airline pilots).

Despite dramatic advances in the surgical management of stones that has rendered virtually all stones in the urinary tract amenable to a noninvasive or minimally invasive treatment, many ureteral calculi will pass spontaneously without the need for treatment; thus conservative management of selected stones should not be overlooked. The likelihood of spontaneous passage depends on the size and location of the stone at the time of diagnosis. Small stones are likely to pass spontaneously regardless of their location in the ureter; in contrast, stones larger than 6 mm rarely pass spontaneously (<10 percent). Likewise, stones located in the distal ureter at the time of diagnosis stand a better chance of passing than more proximally located stones. A recent review of the literature estimated rates of spontaneous stone passage at 12, 22, and 45 percent for stones in the proximal, middle, and distal ureter, respectively.

The relative indications for treatment of stones in the kidney include large size, pain, location in the renal pelvis with the potential for obstruction, and associated infection or loss of renal function. Asymptomatic calyceal stones have an approximately 50 percent chance of becoming symptomatic within 5 years, and consequently, some authors have advocated prophylactic treatment of these stones as well. Asymptomatic stones that demonstrate continued growth despite preventative medical measures also warrant treatment. Staghorn calculi (large branched stones occupying a substantial portion of the collecting system) always require treatment because of the high incidence of renal failure, sepsis, and loss of renal function (Fig. 38-21).

Surgical Intervention

Appropriate surgical management of renal or ureteral calculi depends on the size, location, and composition of the stone, as well as the anatomy of the collecting system and ureter. Large obstructing renal or ureteral calculi or staghorn calculi may warrant an assessment of renal function before treatment. Substantial loss of renal function from long-standing obstruction or infection may favor nephrectomy over renal-preserving surgery. Radionuclide renal studies assess renal function, perfusion, and drainage with minimal radiation exposure. Unlike radiographic imaging studies, the renogram is unaffected by patient size or body habitus. The finding of 20 percent or less differential function in the stone-bearing kidney suggests an unsalvageable renal unit, and nephrectomy should be considered.

Management of renal and ureteral calculi includes noninvasive, minimally invasive, and open surgical approaches. Open surgery is reserved for complex stones with extensive associated anatomic abnormalities (such as a staghorn calculus in a kidney with numerous stenotic infundibula requiring extensive reconstruction of the collecting system) or stones that have failed minimally invasive approaches (e.g., an impacted ureteral calculus in a fixed, scarred ureter that is inaccessible by endoscopic means). The majority of stones, however, are amenable to minimally invasive approaches such as shock-wave lithotripsy, ureterorenoscopy, percutaneous nephrostolithotomy, or laparoscopic uretero- or pyelolithotomy.

Shock-Wave Lithotripsy

The study of shock waves was born from an observation by German scientists that aircraft returning from supersonic flight had pitting on their surfaces, reportedly due to the generation of shock waves that were focused by the contours of the craft onto the surface of the plane. Applying this principle to human study, they observed that shock waves generated under water could be transmitted safely to the body without loss of energy because of the similar acoustic density of water and body tissue. Once the shock wave encountered an object of differing acoustic density, such as a calculus, the energy was released, thereby resulting in stone fragmentation and spontaneous discharge of fragments down the ureter and into the bladder. In 1981, the introduction into clinical practice of this completely noninvasive means of treating stones with shock waves revolutionized the contemporary management of renal and ureteral calculi.

The first lithotripter, designed by Dornier Medical Systems, used an underwater spark plug to generate shock waves that were then focused by means of an ellipsoid reflector (Fig. 38-22). Using biplanar fluoroscopy, the stone was targeted and positioned at the shock-wave focal point. Patients were partially submerged in a large water bath in order to couple the shock waves to the patient. Newer-generation lithotripters differ in the source of shock-wave generation, the means of coupling, the focusing mechanism, the imaging modality, and the anesthesia requirements of the treatment.

With current technology, in excess of 85 percent of stones are amenable to *shock-wave lithotripsy (SWL)*. Most patients are treated on an outpatient basis with the expectation of resuming work within 1 to 2 days. Although SWL has a remarkably low morbidity, it is associated with some complications, either from direct injury to the kidney or from problems associated with passage of stone fragments. Most patients will experience a minimal degree of renal injury, manifested by gross hematuria, but few patients will demonstrate clinically significant hemorrhage (0.1 to 0.7 percent incidence of clinically significant perinephric hematoma).

Incomplete stone fragmentation or fragmentation of a large stone may result in the accumulation of obstructing stone fragments in the ureter (*steinstrasse*) (Fig. 38-23). While some steinstrasses resolve spontaneously, others require intervention such percutaneous nephrostomy drainage or ureteroscopy with intracorporeal lithotripsy to clear the fragments.

The overall success rate of SWL for renal calculi averages 70 percent using the first-generation Dornier HM3 lithotripter. Success rates vary according to the size, location, and composition of the stone. Stone-free rates vary inversely with stone size; for stones greater than 2 cm in size, stone-free rates drop off substantially. Shock-wave monotherapy for treatment of staghorn calculi is generally discouraged because of the high retreatment rate and the frequent complications attributable to passage of a large volume of stone fragments. SWL monotherapy is reserved exclusively for small-volume staghorn calculi in nondilated collecting systems, a subgroup comprising less than 20 percent of staghorn calculi. The use of a ureteral stent has been shown to decrease the rate of complications and the need for secondary procedures when treating stones greater than 2.5 cm in diameter.

Stone location also influences the outcome of SWL treatment. Higher stone-free rates are achieved with SWL therapy of renal pelvic stones compared with treatment of calyceal stones. In particular, lower-pole stones are associated with the lowest stone-free rates of any renal calculi because of difficulty clearing fragments from the dependent lower-pole calyces. A recent metaanalysis comparing SWL and percutaneous nephrostolithotomy for the treatment of lower-pole stones favored a percutaneous approach in all size categories.

Stone composition bears directly on the success of SWL, although stone analysis is often unavailable before treatment. Cystine stones are relatively resistant to SWL, and as such, alternative endoscopic treatment is recommended for cystine stones greater than 1 cm in diameter. Calcium oxalate monohydrate stones are hard stones with variable success when treated with SWL. In general, calcium oxalate dihydrate and uric acid stones are easily fragmented with SWL.

Ureteral stones in all locations have been managed successfully with SWL. Modifications of standard patient positioning, such as treatment in the supine or prone position, have enabled treatment of stones in the middle (overlying the bony pelvis) or distal (below the bony pelvis) ureter, which are unable to be targeted with patients on the standard gantry. Early investigators advocated retrograde displacement of ureteral stones into the renal pelvis or bypass of the stone with a ureteral stent prior to SWL therapy because of perceived lower success rates with in situ treatment of ureteral stones. However, several prospective, randomized series have shown that endoscopic manipulation prior to SWL is unnecessary for proximal ureteral stones. No randomized series have evaluated the need for ureteral stents for middle or distal ureteral stones, although several retrospective series suggest that stent placement is unnecessary for stones in these locations as well.

With the introduction of SWL came the sometimes indiscriminate treatment of nearly all renal and ureteral calculi with this noninvasive modality. Now, after nearly 15 years of experience with SWL and extensive analyses of SWL failures, the indications for SWL have been defined more precisely. Staghorn calculi, renal stones greater than 2.5 cm in size, and ureteral stones greater than 1.5 cm in size are best managed endoscopically. Likewise, lower-pole calculi greater than 1 to 2 cm in size may be inappropriate for SWL and will be treated more successfully percutaneously or with ureterorenoscopy. Renal or ureteral calculi associated with distal obstruction, such as ureteropelvic junction obstruction or ureteral stricture, should be managed with an antegrade percutaneous or retrograde ureteroscopic approach in which the stricture can be addressed endoscopically at the time of stone removal. Obese patients pose unique problems for SWL; weight limitations on the gantry or table, difficulty visualizing the stone fluoroscopically or with sonography, and inability to position the stone at the shock-wave focal point because of a skin-stone distance that exceeds the distance to the focal point may all preclude successful treatment.

Percutaneous Nephrostolithotomy

Percutaneous nephrostolithotomy (PCNL) is the endoscopic removal or fragmentation of stones through a tract from the skin into the kidney through the flank. PCNL is generally reserved for large or complex renal or ureteral calculi or for stones associated with anatomic obstruction precluding spontaneous passage of stone fragments after SWL. PCNL fell

into disuse after the introduction of SWL, and many urologists lost their percutaneous skills; however, PCNL has experienced a reemergence as the indications for SWL have become more refined.

PCNL requires a general anesthetic and may be performed in conjunction with an interventional radiologist. The procedure is facilitated by placement of a ureteral occlusion balloon catheter into the renal pelvis to opacify the collecting system and to prevent distal migration of stone fragments during intracorporeal lithotripsy. With the patient in the prone position, a small- gauge hollow-bore needle is positioned under fluoroscopic guidance directly into the stone-bearing calyx or into the calyx providing optimal access to the stone. A guidewire is passed through the needle and maneuvered into the collecting system and preferably down the ureter into the bladder. A second guidewire is placed to serve as a safety wire. The nephrostomy tract is then dilated using a high-pressure balloon or sequential graduated dilators, culminating in the placement of a working sheath (usually #30 French in size, or 10-mm inner diameter) through which the endoscope is passed directly into the collecting system.

Stones 1 cm or less in size can be retrieved intact; stones greater than 1 cm in size require some form of power lithotripsy, usually either ultrasonic or pneumatic lithotripsy, to fragment the stone into small pieces that can be extracted or aspirated using suction attached to the end of the hollow lithotripter probe (Fig. 38-24). The use of a flexible nephroscope with fiberoptic imaging allows access to stones in calyces remote from the nephrostomy tract and to stones in the ureter; these stones can be retrieved with flexible graspers or baskets passed through the working channel of the endoscope, or the stone can be fragmented in situ with a flexible laser fiber or electrohydraulic lithotripsy probe.

At the conclusion of the procedure, a ureteral catheter is passed over the working guidewire into the bladder and a large-caliber nephrostomy tube (#22 French) with a central hole in the tip is placed over the ureteral catheter into the renal pelvis to provide urinary drainage and to tamponade the nephrostomy tract. After 24 to 48 h, a nephrostogram is obtained to assess drainage from the kidney and to identify extravasation of contrast from an inadvertent perforation in the collecting system. If residual stone fragments are detected on postoperative imaging studies, “second look” flexible nephroscopy can be performed under local anesthesia with intravenous sedation to retrieve the remaining stones.

Unlike SWL, stone-free rates for PCNL are independent of stone size and stone location; stone-free rates of 70 to 100 percent have been reported with PCNL. The most common complication associated with PCNL is bleeding; the average decrease in hemoglobin is 1.2 to 2.8 g/dL. In fewer than 1 percent of patients, persistent bleeding from pseudoaneurysm formation necessitates arteriography and transcatheter arterial embolization. The need for renal exploration or nephrectomy is exceedingly rare. Access- related injury to adjacent structures most commonly involves the pleura, which can be traversed in the case of an upper-pole supracostal access. Injury to the spleen, liver, colon, and duodenum have been reported but are rare and generally can be managed conservatively.

PCNL can be performed in association with renal reconstructive surgery; a stricture or narrowing of the infundibulum leading to a stone-bearing calyx (infundibular stenosis) or a short stricture at the ureteropelvic junction (congenital or acquired ureteropelvic obstruction) can be incised endoscopically with a “cold” or “hot” knife, and the stones may be removed simultaneously at the same sitting.

According to the American Urological Association Nephrolithiasis Clinical Guidelines Panel, first-line treatment for staghorn calculi is PCNL, with the addition of SWL and repeat PCNL as needed. Complex stones may require multiple accesses into the kidney in order to render the patient free of stone. In some cases, combination therapy with SWL may be indicated; “sandwich therapy” offers the high success rates of PCNL with the lower morbidity of SWL. Sandwich therapy is a multistage procedure that consists of PCNL followed by SWL and always concludes with flexible nephroscopy for removal of residual fragments remaining after SWL.

Ureterorenoscopy

The introduction of the Hopkins rod lens system and fiberoptic technology paved the way for the development of semirigid and flexible endoscopes that could be passed through the urethra and bladder into the ureter and kidney for inspection and treatment of a variety of pathologic conditions, including stones. Under direct endoscopic vision, the stone can be visualized and extracted intact with a grasper or basket, or it can be fragmented in situ with a variety of intracorporeal lithotripsy modalities: laser (pulsed-dye, alexandrite, neodymium:YAG, or holmium:YAG), electrohydraulic (in situ shock-wave generation), or pneumatic (ballistic energy).

Ureteroscopy begins with the successful passage of a guidewire through the ureteral orifice up the ureter and into the kidney via a cystoscope passed into the bladder; for flexible ureteroscopy, a second, safety guidewire is placed. The semirigid ureteroscope is advanced directly into the ureter under endoscopic vision *alongside* the guidewire until the stone is reached. In contrast, the flexible ureteroscope is passed *over* the guidewire and advanced under fluoroscopic guidance to the site of pathology; the guidewire through the ureteroscope is then removed, leaving only the safety guidewire in place (Fig. 38-25). Working instruments are passed through the working channel onto the stone to either retrieve the stone intact or to fragment it into spontaneously passable fragments. A ureteral stent is routinely left in place after ureteroscopy to prevent obstruction and renal colic due to edema at the site of stone manipulation. The stent is usually removed in the office with a rigid or flexible cystoscope 3 to 7 days after the procedure.

Recent technology has resulted in miniaturization of ureteroscopes as small as #6.9 French for semirigid and #7.5 French for flexible ureteroscopes; indeed, most small-caliber ureteroscopes can be passed routinely into the undilated ureter. Small-caliber ureteroscopes allow access to virtually any stone in the kidney or ureter. Improvements in ureteroscope design have led to better imaging and larger-caliber working channels, thereby expanding the armamentarium of instruments that can be used in conjunction with the ureteroscope. The introduction of the holmium:YAG laser lithotripsy has raised the upper limit of stone size treatable endoscopically and has made ureteroscopy a reasonable alternative to SWL or PCNL for some moderate to large renal calculi. Ureteroscopy can be performed on an outpatient basis with minimal morbidity. Ureteral perforation and late development of ureteral stricture remain significant, albeit rare, complications of ureteroscopy that render ureteroscopy second-line or salvage therapy for most renal or ureteral calculi after failed SWL.

Laparoscopic Stone Removal

The safety and feasibility of laparoscopic pyelolithotomy and ureterolithotomy have been demonstrated in anecdotal reports in the literature. Laparoscopic stone removal remains a salvage therapy and serves as an alternative to open surgery in the rare case of SWL or endoscopic failure.

Medical Management

Despite dramatic advances in the minimally invasive surgical management of stone disease, the need for a medical prophylactic program remains unchanged. Surgical treatment of renal calculi does not eliminate the risk of stone recurrence; indeed, in one retrospective study, selective medical therapy reduced the risk of stone recurrence after successful SWL from 0.65 to 0.09 per year.

The selection of patients in need of a metabolic evaluation is controversial. The evaluation of recurrent stone formers or high-risk first-time stone formers (patients with a strong family history of stone disease, associated medical conditions, or nephrocalcinosis) is clearly indicated. The evaluation of first-time stone formers without obvious risk factors is debatable. Nevertheless, first-time stone formers have been shown to have the same incidence and severity of metabolic abnormalities as recurrent stone formers. The decision to pursue a detailed metabolic evaluation must be shared by the patient and the physician.

The first step in evaluation of the stone-forming patient is a careful history. Identification of risk factors such as a family history of stone disease, gastrointestinal disorders, bone disease, nephrocalcinosis, chronic UTIs, or a history of gout should prompt a detailed metabolic evaluation. For first- time stone formers without significant risk factors, a simplified evaluation may suffice.

It is important to identify any medical conditions that may be complicated by renal calculi. Chronic diarrheal syndrome as a result of intestinal disease (Crohn's disease, ulcerative colitis), bowel resection, or pancreatic insufficiency may be associated with uric acid or calcium stones because of dehydration, hypocitraturia, and enteric hyperoxaluria. Gout is frequently associated with uric acid or calcium oxalate stones because of low urine pH and hyperuricosuria. Bone disease may be a result of hyperparathyroidism but in association with brain tumors or peptic ulcer disease may represent multiple endocrine neoplasia syndrome. Recurrent UTIs, particularly with *Proteus*, *Klebsiella*, or *Pseudomonas*, predisposes to struvite stone formation.

A careful dietary and medication history can elicit environmental factors that predispose to stone formation, such as inadequate fluid intake, salt abuse, a high calcium or oxalate diet, or a purine-rich diet. Stone- provoking medications also may contribute to the risk of stone formation. Acetazolamide, a carbonic anhydrase inhibitor, is associated with metabolic acidosis and may lead to hypocitraturia, hypercalciuria, and an alkaline urine. Calcium supplements or vitamin D, commonly taken for osteoporosis, may cause hypercalciuria. Large doses of vitamin C, which is metabolized to oxalate, may promote calcium oxalate stone formation.

A simplified evaluation includes a careful history, a serum chemistry profile, urinalysis, plain radiograph, and stone composition. The serum chemistry profile may identify systemic risk factors, such as hypercalcemia associated with primary hyperparathyroidism, hyperuricemia associated with gout, hypokalemia, hyperchloremia, and low serum bicarbonate associated with distal renal tubular acidosis or hypophosphatemia associated with renal phosphate leak. Microscopic inspection of the urine may reveal pathognomonic crystals, such as cystine crystals, and dipstick urinalysis may reveal extremes of pH; pH > 7.5 is associated with infection by urea- splitting organisms with subsequent struvite stone formation, whereas pH < 5.5 leads to precipitation of uric acid and subsequent formation of uric acid or calcium stones. In some cases, the stone composition may suggest the underlying metabolic abnormality responsible for stone formation, such as cystine stones and cystinuria, struvite stones and recurrent UTIs, and uric acid stones and low urine pH.

A more extensive evaluation includes 24-h urine collections on a random diet and after 1 week of a low-calcium, low-sodium diet. The urine is assessed for total volume, pH, calcium, oxalate, sodium, uric acid, citrate, magnesium, and phosphorus. A urinary calcium-creatinine ratio measured first while fasting and then after an oral calcium load can distinguish the various causes of hypercalciuria. Bone densities are obtained to assess bone loss prior to consideration of calcium restriction.

With an extensive ambulatory protocol, a diagnosis can be established in more than 95 percent of stone-forming patients. Treatment is then directed at correction of the underlying metabolic derangements and elimination of environmental risk factors. The pathophysiologic categories of stone formation are characterized below (Table 38-2).

Hypercalciuria

Hypercalciuria is a manifestation of a variety of disorders. Absorptive hypercalciuria, the most common form of hypercalciuria, is due to intestinal overabsorption of calcium. The mechanism responsible for calcium overabsorption is unknown; consequently, treatment is aimed at reducing urinary calcium excretion rather than normalizing intestinal calcium absorption. Renal hypercalciuria is less common than absorptive hypercalciuria. The primary defect in renal hypercalciuria is impaired renal tubular reabsorption of calcium. Thiazide diuretics are the mainstay of treatment for both types of hypercalciuria; thiazides directly stimulate calcium reabsorption in the distal tubule and indirectly stimulate reabsorption in the proximal tubule. The hypocalciuric action of thiazides becomes attenuated after 18 to 24 months in absorptive hypercalciuria but is maintained indefinitely in renal hypercalciuria. Consequently, the search for more effective drugs for the treatment of absorptive hypercalciuria has led to the development of a slow-release form of neutral potassium phosphate that has been shown to reduce intestinal calcium absorption and enhance excretion of urinary inhibitors. Preliminary results are encouraging, and clinical trials are currently in progress.

Resorptive hypercalciuria is usually due to primary hyperparathyroidism. Excessive secretion of parathyroid hormone from a parathyroid adenoma leads to bone resorption and increased renal synthesis of 1,25-(OH)₂D, which in turn enhances intestinal absorption of calcium; the net effect is elevated serum and urinary calcium levels. Primary hyperparathyroidism is a surgical disease, and parathyroidectomy is the optimal treatment.

Hyperuricosuria

Hyperuricosuria constitutes a risk factor for both uric acid and calcium oxalate stone formation. The most common cause of hyperuricosuria is a purine-rich diet. At low urine pH (<5.5), the undissociated form of uric acid, which has a low solubility, precipitates to form uric acid stones. At pH > 5.5, sodium urate formation predominates and promotes calcium oxalate stone formation by serving as a nidus for calcium oxalate crystallization.

Hyperoxaluria

Enteric hyperoxaluria, associated with inflammatory bowel disease or bowel resection, is the most common form of hyperoxaluria. Intestinal malabsorption of fat enhances saponification with divalent cations (magnesium and calcium) in the intestinal lumen, thereby reducing complexation with oxalate and increasing the reservoir of free oxalate available for absorption. Furthermore, poorly absorbed bile salts and fatty acids are thought to increase colonic permeability to oxalate, further increasing serum and urine oxalate levels. Intestinal disease is associated with other metabolic abnormalities as well, such as dehydration, metabolic acidosis, and hypocitraturia. Treatment is aimed at correcting dehydration and reducing the metabolic acidosis. Potassium citrate therapy corrects the acidosis and normalizes urinary citrate. Limitation of dietary oxalate, and in some cases supplementation with calcium citrate, may correct the hyperoxaluria.

Hypocitaturia

Hypocitaturia is associated with a variety of conditions, such as chronic diarrheal syndrome, renal tubular acidosis, or thiazide therapy, or it may be idiopathic and associated with calcium oxalate stones. Potassium citrate supplementation restores urinary citrate, raises urinary pH, and lowers the saturation of calcium-forming salts.

Gouty Diathesis

Gouty diathesis describes a stone-forming propensity characterized by a low urinary pH level, which may or may not be associated with gouty arthritis or hyperuricemia. Gouty diathesis is typically associated with uric acid stone formation but also predisposes to calcium stone formation. In patients with uric acid stones, treatment is directed at raising urinary pH above 5.5 and maintaining a normal urinary uric acid level. Correction of dietary overindulgence in purine-rich foods (meat, fish, poultry) or the administration of allopurinol should normalize the urinary uric acid level. Potassium citrate raises the urinary pH level and results in dissolution of existing uric acid stones and inhibition of new stone formation.

Cystinuria

Cystinuria is an inherited transport disorder that results in excessive urinary excretion of dibasic amino acids (cystine, ornithine, lysine and arginine). In patients with the disorder, the urinary cystine concentration frequently exceeds the solubility of cystine, resulting in stone formation. High fluid intake is critical for maintaining the cystine concentration below the solubility limit. The solubility of cystine varies according to pH and is highest at a pH of more than 7.5. Consequently, alkali therapy may result in dissolution of preexisting stones and prevention of new stone occurrences. When fluid and alkali therapy alone are inadequate to prevent stone formation, the addition of a chelating agent such as d-penicillamine or a-mercaptopyronylglycine reduces cystine concentration by forming a more soluble cysteine-sulphydryl complex.

Infection Lithiasis

The formation of ammoniomagnesium phosphate, or struvite, stones is due to chronic infection with urea-splitting organisms. The hydrolysis of urea catalyzed by these bacteria leads to ammonium production and an alkaline urine. In this environment, the components of struvite become supersaturated, and crystals form. The mainstay of therapy for struvite stones is surgical removal of the stone and eradication of the offending organism. In some recalcitrant cases, the administration of acetohydroxamic acid, a urease inhibitor, can facilitate stone dissolution and prevention.

BENIGN PROSTATIC HYPERPLASIA

Benign prostatic hyperplasia (BPH) is the most common disorder of the prostate gland. Although seldom life-threatening, BPH can have a significant impact on the aging male's quality of life. In the majority of men affected, the disease produces urinary symptoms that are slightly inconvenient. In others, however, the presence of more bothersome symptoms or the development of more serious sequelae leads to therapeutic intervention. Although surgical removal of the obstructing prostatic tissue offers the highest probability of symptomatic improvement, less effective medical therapies are attractive to some patients because they have less risk.

Natural History

Hyperplastic growth of the prostate begins in many men by the fifth decade of life. By age 80, almost 90 percent of men have histologic evidence of BPH. The normal adult prostate reaches a plateau at a weight of about 20 g at age 30. Then the weight remains stable until approximately age 50, when a process of growth is initiated at an average of 0.5 to 0.8 g/year. In populations, the probability of developing symptoms increases with increasing prostate size. In individual patients, however, the correlation between prostate weight and symptom severity is poor.

The clinical progression of BPH is variable. Approximately one-third of patients being followed by “watchful waiting” will have improvement in symptoms over time, whereas an equal number will have no worsening. The remainder will have symptomatic progression. A small number of men in the latter group will develop more serious sequelae of the disease: urinary retention, recurrent UTIs, bladder stones, bladder failure, or renal dysfunction. The risk of acute urinary retention in men with BPH is approximately 1 percent per year.

Pathogenesis

BPH first develops in the periurethral transition zone of the prostate. The transition zone consists of two separate lobules of tissue immediately external to the preprostatic sphincter (Fig. [38-26](#)). The main ducts of the transition zone arise on the lateral aspects of the urethral wall at the point of urethral angulation near the verumontanum. Proximal to the origin of the transition zone ducts are the glands of the periurethral zone, which also undergo hyperplastic growth. Stromal (smooth muscle and extracellular matrix) and epithelial nodules develop in these zonal compartments, leading to an overall increase in the size of the gland. As the prostate enlarges, the peripheral zone, the most common site for the development of prostatic adenocarcinoma, is compressed between the transition zone and the capsule of the gland.

One of the unique features of the human prostate is the presence of the prostatic capsule, which plays an important role in the development of prostatism. In the dog, the only other species known to develop naturally occurring BPH, symptoms of prostatism rarely develop because the canine prostate lacks a capsule. Presumably the capsule transmits the pressure of tissue expansion to the urethra, leading to an increase in urethral resistance.

Growth of the prostate in aging men is not a uniform process. Some patients develop “global” enlargement of the entire gland, whereas others develop more prominent growth of specific regions of the gland (Fig. [38-27](#)). Extensive growth of the periurethral zone may lead to a “middle lobe” and a resulting ball-valve type of obstruction.

Smooth muscle cells within the prostatic capsule, stroma, and bladder neck have a high density of a ₁-adrenergic receptors on their surfaces. Thus tone in the prostatic urethra is influenced by the degree of adrenergic stimulation to the gland, which contributes in a dynamic way to outflow obstruction.

The bladder's initial response to prostatic enlargement is the development of compensatory muscular hyperplasia. Intravesical pressure increases to maintain flow in the face of outflow resistance. Unfortunately, the adaptation is not perfect: Bladder hyperplasia leads to urinary frequency and urgency. In advanced cases, the bladder wall becomes fibrotic, loses compliance, and fails to empty completely. The development of postvoiding residual urine, however, is common and does not invariably lead to urinary retention and hydronephrosis.

The molecular pathogenesis of BPH is uncertain. It is clear that the process requires aging and testicular androgens. BPH does not develop in men castrated before puberty. Testosterone, produced by testicular Leydig cells, is the major androgen in the circulation. After testosterone diffuses into the prostate stromal cell, most of it is converted into a much more potent androgen, dihydrotestosterone (DHT), by the enzyme 5α-reductase (Fig. [38-28](#)). The importance of this amplification step in growth of the prostate is clear from studies of patients with deficiency of the 5α- reductase enzyme who have virtually absent prostates. Both testosterone and DHT bind to the androgen receptor in the epithelial cell to stimulate androgen-dependent cell growth and inhibit the programmed cell death seen on androgen withdrawal. Although androgens do not cause BPH, DHT plays a central role in the pathogenesis of the disease.

Clinical Manifestations

Symptoms

Men with clinically significant BPH have a mixture of symptoms that are often termed *LUTS*: lower urinary tract symptoms. Symptoms of urinary frequency and urgency, decreased force of the urinary stream, hesitation in the initiation of flow, intermittency, a sensation of incomplete emptying, and nocturia occur to variable degrees. These symptoms can be quantified by use of a standardized questionnaire developed by the American Urological Association (Table [38-3](#)). Men with AUA symptom scores of 7 or less are seldom bothered by their symptoms and usually can be managed successfully by watchful waiting.

Unfortunately, LUTS is not specific for BPH. UTI, prostatitis, prostate and bladder cancer, urethral stricture disease, and neurogenic bladder disease may all produce identical symptoms. Therefore, further evaluation is necessary (Fig. [38-29](#)).

Baseline Evaluation

A careful history should be taken to exclude causes of urinary tract dysfunction. A history of urethritis, straddle injury, or lower urinary tract instrumentation puts the patient at risk for urethral stricture disease. The presence of dysuria or perineal/suprapubic pain may indicate UTI or prostatitis. Many neurologic diseases, such as stroke, Parkinson's disease, diabetic neuropathy, and disc disease, may alter bladder function. Congestive heart failure, diabetes, and renal insufficiency may lead to polyuria, resulting in urinary frequency and nocturia.

Physical examination must include a careful digital rectal examination (DRE), a limited neurologic assessment to exclude obvious abnormalities, palpation/percussion of the suprapubic area to detect a distended bladder, and an examination of the external genitalia. The DRE is performed to estimate prostate size and to detect asymmetry, induration, or nodularity that may be suspicious for malignancy. Although the size of the prostate does not determine disease severity, the finding of a normal-sized gland on DRE increases the

probability that the symptoms are not related to BPH.

Urinalysis must be done to exclude UTI and hematuria. Infection may mimic the symptoms of BPH or be a complication of outflow obstruction. The presence of hematuria should prompt endoscopy and imaging of the kidneys to exclude urothelial and renal malignancy, as well as bladder stones. Bacterial prostatitis may not be associated with pyuria. Therefore, examination of the expressed prostatic secretions and localization cultures should be performed in men with prostatism who also have dysuria or perineal/suprapubic pain.

Measurement of the serum creatinine level is important to evaluate renal function. Even mild degrees of renal insufficiency should be evaluated by renal sonography. Patients with minimal prostatism may develop obstructive uropathy, although rarely.

Although BPH does not appear to be a risk factor for prostatic cancer, *both diseases commonly occur in the same patient*. Digital rectal examination and measurement of the serum PSA level are the best tests to detect prostate cancer. Serum PSA determination is a valuable test in men between the ages of 50 and 75 who have at least a 10-year life expectancy, but PSA is not a specific tumor marker. Approximately 25 to 30 percent of men with BPH will have an elevated PSA level. Since PSA elevation usually prompts the performance of transrectal ultrasonography and biopsy of the gland, this high rate of false-positive results is bothersome. Recently, assays that measure the percentage of PSA in the serum that is unbound to plasma proteins (free PSA) have improved the specificity of PSA testing.

In men with only mild symptoms, and in those who elect medical therapy or observation, more diagnostic testing is not always necessary. If, however, invasive therapy is being considered or the results of the initial evaluation are discordant, further diagnostic testing is indicated (see Fig. [38-29](#)).

Diagnostic Tests

The best noninvasive test to evaluate men with prostatism is the uroflow, measured by an electronic device that calculates the velocity of urine flow. The maximal urinary flow rate ($Q_{\max}$) is normally greater than 15 mL/s. A reduction in $Q_{\max}$ is suggestive of obstruction. Men with normal flow rates and lower urinary tract symptoms are not likely to have BPH as the cause. Unfortunately, uroflow does not distinguish obstruction from poor bladder contractility, nor does it localize the site of obstruction, e.g., BPH versus urethral stricture. Despite these limitations, the test is especially valuable in patient follow-up and in quantifying the response to therapy.

Measurement of the postvoid residual (PVR) by catheterization or ultrasonography is helpful in assessing the bladder's response to obstruction. Large residual urine volumes (>300 mL) may be associated with a higher risk of disease progression. PVR, however, has considerable intraindividual variation, and consensus is lacking about the amount of residual urine that should prompt active therapy.

The most sensitive and specific test to confirm the presence of bladder outlet obstruction is pressure-flow urodynamics. In this evaluation, bladder pressure is measured during urination. High bladder pressures and low flow indicate obstruction. Unfortunately, a significant number of tests are equivocal and are invasive and costly. Urodynamic studies are indicated in patients at risk for neurogenic bladder disease and in patients in whom the diagnosis of BPH is still uncertain after noninvasive tests.

Endoscopic evaluation of the lower urinary tract by cystourethroscopy can easily distinguish urethral stricture disease from BPH and provide information on prostatic configuration (i.e., the presence of a middle lobe). Endoscopic demonstration of an "obstructing prostate" and bladder trabeculation, however, is of unproved value. Endoscopy should be considered in men before invasive or surgical therapy but is probably not justified in men being treated conservatively.

Renal imaging by excretory urography or by ultrasonography is indicated in the evaluation of BPH in men who have concomitant UTI, hematuria, renal insufficiency, a history of urolithiasis, or previous urinary tract surgery. Renal imaging is not cost-effective in routine cases. Prostatic imaging by ultrasonography can provide much of the same information provided by endoscopy.

Treatment

Men with only mild symptoms or asymptomatic prostatic enlargement should be managed by watchful waiting. By contrast, men who have developed urinary retention, recurrent infections, bladder stones, or renal insufficiency secondary to BPH should be treated surgically. For men with bothersome symptoms, watchful waiting, medical therapy, and surgery should be presented to the patient as treatment options.

Watchful Waiting

Observation is the most appropriate therapy for men with only mild symptomatology (AUA symptoms score <7) (see Fig. [38-29](#)). A significant number of men with moderate to severe symptoms will not progress or will find their symptoms not bothersome enough to warrant therapy. Patients in this group should be seen annually, and the baseline evaluation repeated.

Surgical Therapy

Transurethral resection of the prostate (TURP) is the most common surgical procedure to treat BPH. TURP is clearly the most effective treatment for BPH. Between 80 and 90 percent of patients undergoing TURP experience significant improvement in urinary symptoms and uroflow. The procedure is usually done under regional anesthesia, with tissue removal achieved by an electrocautery-type loop inside a modified cystoscope (resectoscope). The operative mortality for TURP is 1 percent or less, but morbidity rates can be high in elderly patients with underlying cardiovascular disease. Short-term complications include the rare absorption of hypotonic irrigation fluid leading to hyponatremia (TURP syndrome), blood loss requiring transfusion (5 percent), urinary retention, and infection. Long-term complications include urinary incontinence (1 percent), impotence (5 to 10 percent), retrograde ejaculation (78 to 80 percent), and the need for repeat endoscopic surgery for either recurrence, bladder neck contracture, or urethral stricture disease (10 to 15 percent at 8 years). Transurethral laser prostatectomy and electrovaporization techniques have been evaluated, but these have not been shown to be more effective than the standard TURP.

Transurethral incision of the prostate (TUIP) is almost as effective as TURP in glands smaller than 30 g in size. In this endoscopic procedure, one or two incisions are made through the prostatic capsule posteriorly. Without removing tissue, this procedure reduces urethral resistance. The procedure can be done on an outpatient basis and has less risk of retrograde ejaculation, impotence, and blood loss than TURP.

Because of prolonged resection times and the resulting increase in complication rates, very large prostate glands (>60 to 80 g) can be treated by simple retropubic or suprapubic prostatectomy. In these procedures, a lower abdominal incision is made, and the obstructing prostatic tissue is "enucleated" through an incision in the prostatic capsule (retropubic) or in the anterior wall of the bladder (suprapubic). Complication rates for open prostatectomy are probably no higher than for TURP, but incisional morbidity makes TURP the preferred procedure if the size of the prostate does not preclude it.

Medical Therapy

Currently, there are two medical approaches to the treatment of BPH. α -Adrenergic receptor blockers (doxazosin, terazosin, tamsulosin) relax prostatic smooth muscle and partially relieve the "dynamic" component of obstruction. Selective α_1 antagonists achieve significant symptom and flow rate improvement in 30 to 50 percent of men. Side effects of asthenia, headache, dizziness, and orthostatic hypotension occur in 10 to 15 percent of men.

The 5 α -reductase inhibitor finasteride reduces intraprostatic dihydrotestosterone levels significantly, without lowering plasma testosterone levels. After 6 months of therapy, the average patient will experience a 20 percent reduction in the size of the prostate. Only one-third of men, however, experience significant improvement in symptoms or uroflow. The major advantage of the drug appears to be a 50 percent reduction in the need for TURP and the risk of AUR. Between 10 and 14 percent of men report sexual dysfunction when treated; otherwise, the drug has no significant side effects. Finasteride lowers the serum PSA level, but this can be corrected simply by multiplying the serum PSA value by 2. Cancer detection does not appear to be compromised.

Minimally invasive treatments for BPH include transurethral needle ablation (TUNA) and transurethral microwave thermal therapy (TUMT). These approaches are safe and effective, but their long-term durability is uncertain.

NEOPLASMS

Renal Parenchymal Tumors

Incidence and Etiology

Renal tumors account for approximately 2 percent of cancer deaths. The frequency in males is twice that in females, and there is a relationship between the incidence and smoking. The histologic origin of most carcinomas in males is from the renal tubular cell, as evidenced by the electron microscopic demonstration of a microvillous brush border on the carcinoma cells.

Pathology

Benign renal tumors (adenomas) frequently are noted on postmortem examinations, but they are rarely clinically significant. There are three major types of malignant tumors of the renal parenchyma. Renal cell carcinoma accounts for 85 percent of these lesions. Grossly, the tumor is yellow, at times pink, and often contains cystic areas of hemorrhage and necrosis. There is a tendency for the tumor to invade its own venous system, extending into the vena cava and even into the right side of the heart. Resection of tumors with this marked intravascular extension is feasible, occasionally using cardiopulmonary bypass, and cures have resulted. Microscopically, the dominant cells appear clear but are often admixed with granular cells.

Wilms' tumor (adenomyosarcoma) characteristically occurs in children and accounts for 14 percent of parenchymal malignancies. Sarcomas arising from interstitial tissue constitute 6 percent of malignancies. Tumors of the collecting system make up 14 percent of renal malignancies but are not considered in the parenchymal category.

Clinical Manifestations

The classic triad of pain, mass, and hematuria occurs late and is seen in its entirety in less than half the patients. Marked variations in the prevalence of these symptoms have been noted in reported series: gross hematuria, 31 to 76 percent; pain, 24 to 75 percent; mass, 6 to 50 percent; fever, 16 percent; metastatic pain, 6 percent. Hematuria is a late manifestation that represents erosion of the tumor into the collecting system. A palpable mass is more frequently present if the tumor involves the lower pole of the right kidney. Pain has been related to stretching of the renal capsule, in which instance it is usually mild but persistent. Sudden severe flank pain may accompany bleeding into the tumor. The passage of blood clots mimics colic from a ureteral stone. Fever is probably caused by necrosis of a tumor. Adult parenchymal tumors are usually not associated with hypertension

unless there is a renin-producing tumor. Hypertension is a more frequent accompaniment of Wilms' tumor. Hypertension also has been related to compression of the renal artery with resulting renal ischemia.

Erythrocytemia is present in 2 percent of patients with hypernephroma, which accounts for 7 percent of patients evaluated for polycythemia. Erythropoietin has been extracted from these renal tumors. Hepatic cell dysfunction (Stouffer's syndrome), reversible with removal of the renal carcinoma, has been noted.

Early detection often is the result of abdominal imaging studies, such as CT scans, that have been performed for evaluation of noneurologic conditions. Metastases, which were seen in one-third of patients at the time of diagnosis in the pre-CT scan era (lungs, bones, lymph nodes, and skin), is now seen in fewer than 15 percent of newly diagnosed cases.

Diagnosis

A mass lesion can be seen on an excretory urogram. The uninvolved, functioning portion of the kidney is visualized, whereas the tumor is not. If the renal vein is occluded, the kidney fails to visualize.

Although ultrasonography is helpful in forming the diagnosis, CT usually provides more meaningful information, differentiating between solid and cystic lesions and determining the extent of the tumor and operability (Fig. 38-30). MRI has been used to determine the extent of vascular invasion. The use of arteriography has been reduced because of the more liberal application of CT, but selective renal arteriography provides a "road map" of the arterial pattern that may be helpful if partial nephrectomy is contemplated (Fig. 38-31) and to detect recurrence in the tumor bed. A vena cavagram will identify extension of tumor into that vessel (Fig. 38-32), although MRI is more commonly used for this purpose.

Treatment

Radical nephrectomy with removal of the ipsilateral adrenal and hilar nodes is the standard operation for unilateral renal neoplasms. In this operation, the kidney is removed with the surrounding Gerota's fascia and perinephric fat (Fig. 38-33). Early ligation and division of the pedicle minimize blood loss from the usually present extensive collaterals. Preoperative percutaneous angioinfarction, however, is seldom helpful. For small tumors in the cortex or tumors in a solitary kidney, partial nephrectomy is an appropriate option.

Removal of the primary tumor is occasionally followed by regression of metastases. Resection of isolated metastases in the lungs or extremities is associated with prolonged survival. No chemotherapy or immunotherapy has been predictably effective.

Results

Because 80 percent of renal carcinomas are diagnosed by CT in their asymptomatic phase, the 5-year survival now approaches 70 percent. For patients with low-grade local lesions, survival equals actuarial life expectancy. Survival is influenced by histologic type, extension into the perinephric tissues, lymph node, or renal vein involvement, and distant metastases.

Other Parenchymal Tumors

Other solid neoplasms of the renal parenchyma are rarely encountered and are often confused with renal cell carcinoma because of the presence of pain, hematuria, and a solid intrarenal mass.

Oncocytoma

Oncocytoma appears often as a large mahogany-colored, well-demarcated renal neoplasm with a central area of necrosis or a stellate scar, depending on the age of the area of necrosis. The cells have abundant granular eosinophilic cytoplasm (rich in mitochondria) arranged about a space. Most of these oncocytomas have a benign course, but some deaths have occurred from metastatic disease. Some renal cell carcinomas have an oncocytic pattern, making it impossible to distinguish them from oncocytomas.

Angiomyolipoma

Angiomyolipoma (hamartoma) usually occurs as a solitary lesion but sometimes is associated with tuberous sclerosis and should be thought of in patients with adenoma sebaceum, seizures, mental retardation, and renal masses with or without hematuria. The three components vary, but the tumor usually can be differentiated by CT because of its fat content (Fig. 38-34). The lesion is benign and does not of itself usually require surgery but is often operated on because of complications, i.e., hemorrhage or sepsis, or because of confusion with renal cell carcinoma. Bleeding can be controlled by embolization. Rarely the lesion contains an intrarenal neoplasm.

Carcinoma of the Collecting System

Renal Pelvis and Ureter

Tumors of the renal pelvocalyceal system constitute about 14 percent of kidney tumors and are usually transitional cell carcinomas (TCCs); about one-third of patients with TCC of the kidney have TCC of the bladder at the same time. Squamous carcinoma has been associated with calculus in 50 percent of patients. Conversely, only 2 to 3 percent of staghorn calculi have concomitant squamous cell carcinoma.

Transitional cell tumors of the renal pelvis and ureter are histologically similar to their vesical counterparts. They occur more frequently in men and are seldom seen in patients under the age of 50. Gross hematuria and renal colic due to the passage of blood clots are common presenting symptoms. Intravenous pyelography frequently establishes the diagnosis, showing a filling defect in the involved portion of the pelvis or ureter (Fig. 38-35). Differential diagnosis of the filling defects includes blood clot and nonopaque (uric acid) calculi. Calculi can be ruled out by ultrasonography or CT scan.

Urine obtained by ureteral catheter lavage or brushing may demonstrate clusters of abnormal transitional cells. With a flexible nephroscope or ureterscope, direct biopsy can be performed. Transitional cell tumors of the pelvis tend to be multiple, and frequently, there are ureteral implants involving the ureter or the bladder. Treatment consists of removing the kidney, the ureter, and a margin of the bladder around the ureter orifice. The prognosis for survival is markedly reduced when the tumor infiltrates the adjacent parenchyma. Postoperative radiation therapy and chemotherapy have not been demonstrated to be valuable in improving the survival.

Patients with squamous cell carcinoma characteristically have no gross hematuria, flank pain, or mass and frequently present with a resistant urinary infection and markedly reduced function noted on intravenous pyelography. The associated infection or calculus frequently masks the tumor. Ureteral urinary cytology may demonstrate epithelial pearls. The treatment for squamous cell carcinoma is surgical excision, but the prognosis is exceedingly poor because of late diagnosis. Adjuvant radiation therapy and chemotherapy are ineffective.

An algorithm for the diagnosis and treatment of kidney tumors, subject to the age and general health of the patient, is shown in Fig. 38-36.

Urinary Bladder Incidence and Etiology

The estimated incidence of new cases is approximately 45,000. The peak age incidence is 50 to 70 years of age, and the usual lesion is a transitional cell carcinoma. Approximately 7200 men and 3400 women die from cancer of the bladder annually, accounting for 2 percent of cancer deaths.

Bladder tumors have been related to the prolonged exposure to chemicals that are excreted in the urine as conjugated orthoaminophenols. These include b-naphthylamine and para-aminobiphenyl. Certain carcinogenic tryptophan metabolites are present in higher concentrations in the urine of patients with bladder tumors. There appears to be a prevalence of heavy cigarette smokers among the group with bladder tumors. Chronic infection with schistosomiasis and mechanical trauma from calculus also cause metaplasia and neoplasia of the bladder epithelium.

Clinical Manifestations

Gross or microhematuria is the initial sign and is present in 95 percent of patients at the time of diagnosis. Frequency and painful urination are late symptoms that result from sloughing of the tumor, infiltration, and secondary infection. In late cases, the neoplasm can occlude the ureters and lead to hydronephrosis and uremia, manifested by nausea, vomiting, and diarrhea.

The physical findings in early cases are minimal, since the tumor remains within the confines of the bladder. Local extension of the tumor may result in a nodular induration at the base of the bladder with fixation to the wall of the pelvis. This may be palpated on bimanual examination. Lymphatic spread may involve the pelvic and periaortic nodes. Osseous metastatic lesions are usually osteolytic and may cause nerve root pain.

Malignant cells are demonstrated cytologically by examining a urine specimen or bladder washings. The tumor is demonstrated on an excretory urogram or a cystogram (Fig. 38-37). Cystoscopic examination and cold punch biopsy establish the diagnosis and assess the depth of penetration. Preoperative studies to determine the stage include chest x-ray, CT scan of the abdomen and pelvis, and MRI (especially T2-weighted images).

Treatment

Bladder cancers are categorized as superficial non-muscle-invading lesions and deep tumors with invasion of muscle (Fig. 38-38). Endoscopic resection is suitable for superficial well-differentiated lesions. Intravesicle therapy, such as thiotepa, mitomycin, Adriamycin, or BCG (bacille Calmette-Guérin) can be given to decrease the risk of recurrence. Patients with superficial lesions treated by local therapy require periodic cystoscopic examinations.

For tumors infiltrating the bladder wall, cystectomy with supravescical diversion of urine by means of a cutaneous ureteroileostomy (ileal loop), cutaneous ureterocolostomy (colon loop), continent (catheterizable) supravescical diversion, or orthotopic neobladder is indicated (Fig. 38-39). Continent forms of urinary diversion give patients a significantly improved quality of life.

For patients with advanced nonresectable disease, palliation can be achieved by radiation therapy or transurethral local resection followed by the topical application of chemotherapeutic agents. Diversion of the urine relieves suffering by diminishing bladder stretching and spasm and avoiding the discomfort of voiding. Diversion also will reduce the bleeding and prevent sepsis.

Prognosis

The survival depends on the stage: stage A (T 1), 90 percent 5-year survival; stage B (T 2), 55 percent; stage C (T 3), 13 to 20 percent; stage D, less than 5 percent. Skinner has reported a 5-year survival of 36 percent for a series of stage D patients who received a combination of cis- platinum, cyclophosphamide, and Adriamycin (CISCA). Others have

reported equivalent improvement for stage D patients receiving combination chemotherapy.

Prostate

Incidence

In the current era of PSA detection, the incidence of newly diagnosed cases is 400,000 per year, with 45,000 men dying each year. The incidence at autopsy in 65-year-old men is 35 percent; in 90-year-old men it is 70 percent. It is estimated that approximately 5 million men are living with histologic cancer of the prostate. The tumor accounts for 10 percent of male cancer deaths.

Pathology

The causation remains unknown, but an association has been noted to occur when there is atrophy of the gland. Carcinoma can occur in any portion of the gland, and two or three areas are simultaneously involved in about one-third of patients. The apical portion is particularly vulnerable to penetration and spread along the ejaculatory ducts to the base of the gland and the seminal vesicles. The obturator and internal iliac nodes are the first areas of nodal involvement.

The tumor is an adenocarcinoma in more than 90 percent of patients. Criteria of malignancy include small glands in juxtaposition without an intervening stroma, loss of a double layer of nuclei in the acini, capsular invasion, perineural space invasion, and nodal and distant metastases.

Diagnosis

In the past, the diagnosis was suggested by palpating a prostatic nodule on rectal examination. More than 50 percent of prostatic nodules palpated on rectal examination are positive for cancer on biopsy. The liberal use of the PSA screening test has increased the detection rate fourfold in the past decade. PSA levels have been correlated with the clinical extent of disease (Table 38-4). Ultrasonic visualization of hypoechoic areas in the prostate has improved the accuracy of biopsy. An elevated serum acid phosphatase level is characteristic of carcinoma of the prostate, while an elevated alkaline phosphatase level is suggestive of bony metastases.

Early-Stage Carcinoma

This represents disease that is localized to the gland (organ-confined) without extension beyond the capsule or to distant sites. After establishing the diagnosis, the patient should be evaluated for possible radical prostatectomy. The patient's age and general medical condition are important determinants.

The nerve-sparing technique has brought back prostatectomy as the most acceptable treatment for cure. Preservation of the cavernosal nerves responsible for erection and maintenance of the external sphincter mechanism have produced a higher incidence of postoperative potency and continence, respectively. Currently, a patient considering radical prostatectomy should anticipate a more than 90 percent incidence of continence. A 50-year-old man will have a 50 to 70 percent chance of remaining potent postoperatively; if the patient is 70, that chance is reduced to 20 percent.

External-beam radiotherapy is an alternative method of treatment. The 5- and 10-year survivals of 75 and 60 percent, respectively, are essentially the same as those achieved with radical prostatectomy. If the patient is not a candidate for either form of therapy, he can be managed with estrogen therapy or other agents to reduce the androgen level or can undergo orchiectomy. In some instances patients have been followed for long periods without specific treatment.

Advanced Carcinoma

The average age of the symptomatic patient who presents with advanced prostatic cancer is 72. Typically, the symptoms are of a few months' duration, with rapid progression of frequency and nocturia leading to acute retention in 28 percent. Weight loss is common; severe back pain occurs in 14 percent and sciatica in 56 percent. Gross hematuria occurs in only 8 percent of patients, and lymphedema secondary to lymphatic obstruction is present in 4 percent.

Rectal examination characteristically demonstrates a fixed, enlarged, nodular, stony hard prostate. The serum PSA level is usually greater than 20 ng/mL; acid and alkaline phosphatase levels are elevated in two-thirds of patients with extension of tumor or metastatic disease. Combined osteoblastic and osteolytic bony metastases are present in 60 percent, whereas isolated osteolytic metastases are rare. Pulmonary metastases occur in some patients and may be miliary in appearance.

Treatment of symptomatic advanced carcinoma consists of androgen blockade therapy and/or orchiectomy. Reduction in serum testosterone to castration levels may be accomplished with surgical castration (bilateral orchiectomy) or with analogues of luteinizing hormone-releasing hormone (LHRH; medical castration). This treatment rapidly improves pain and anorexia and gradually relieves the obstructive urinary symptoms. The addition of oral antiandrogens is of minimal additional benefit in the average patient. If a large residual urine persists, transurethral prostatic resection may be indicated.

X-ray therapy to isolated metastatic bone lesions is also palliative. Corticosteroids in large doses are symptomatically beneficial but have no direct effect on the tumor. Both 5- and 10-year survival rates depend on the stage of disease at the time of institution of therapy. Chemotherapy regimens for advanced cases after endocrine therapy fails have some benefit, especially the combination of alkylating agents with estramustine. Experimental gene therapy and vaccine studies are underway.

Testicular Tumors

Incidence and Etiology

Testicular tumors account for 2 percent of cancers in the male, and the average age at diagnosis is 32 years. Testicular cancer is seen infrequently in blacks. The cause of human testicular tumors is unknown, but they occur 11 times more frequently in undescended testes and 50 times more often in abdominal undescended testes; they may be related to imperfect embryogenesis. Seminomas represent about 40 percent of malignant testicular tumors, embryonal cell carcinomas and teratocarcinomas about 25 percent each, and adult teratomas 8 percent; choriocarcinoma is limited to about 1 to 2 percent. Embryonal carcinomas and teratocarcinomas are the more common lesions in men in their twenties, whereas older men usually have seminomas. Combinations of tumors are common. A nongerminal neoplasm known as the *Leydig cell tumor* makes testosterone and results in precocious puberty in children but is generally asymptomatic in adults. Lymphomas rarely have been reported in men in their sixth and seventh decades. Benign tumors are very rare.

Clinical Manifestations

Patients usually present with a nonpainful "lump" in the testis. On examination, the lesion is firm, nontender, and solid and does not transilluminate. Occasionally tumors are misdiagnosed as simple hydroceles because of the formation of a secondary hydrocele. The diagnosis of epididymitis may be made erroneously in some situations and delay the appropriate diagnostic surgical intervention. In the patient with a testicular tumor, a large scrotal mass also may develop as a result of rupture of the tumor with formation of a hematocele. Late symptoms of metastatic disease include weight loss, fatigue, hemoptysis, and enlargement of the regional lymph nodes with associated ureteral obstruction.

Diagnosis and Treatment

If a testicular tumor is suspected, surgical exploration is indicated. Ultrasonographic evaluation may be helpful in localization of intrascrotal lesions. Measurement of serum *b* -chorionic gonadotropin (*b*-hCG) and *a* -fetoprotein (AFP) markers should be done prior to surgery, but surgical treatment should not be delayed for the results. If the lesion is confined to the testis on physical examination, an inguinal surgical approach is preferred. The spermatic vessels are occluded before the testis is exteriorized for inspection in order to prevent the spread of tumor. Palpation of induration in the testis is indication for radical orchiectomy, and there is little place for biopsy, because more than 90 percent of solid lesions of the testis in this age group are malignant. In patients with apparent lymphatic or pulmonary spread, the primary tumor should be resected for pathologic evaluation.

Staging has been applied to selection of a therapeutic regimen. Stage A lesions are confined to the testis and cord. Stage B lesions have positive retroperitoneal nodes and are subdivided according to the extent of nodal involvement into B1, B2, and B3. Stage C designates distant metastases above the diaphragm or involvement of visceral organs, including lungs, liver, and brain. Preoperatively, the retroperitoneal nodes are best assessed by CT. In addition to CT-defined nodal involvement, other indications for retroperitoneal node dissection include (1) invasion of lymphatics or blood vessels within the parenchyma by nonseminomatous tumors, (2) invasion of the spermatic cord by nonseminomatous tumors, (3) persistence of a large volume of teratocarcinoma within the testis, and (4) failure of markers to return to normal after removal of testis containing nonseminomatous germ cell elements.

Stage A seminomas usually are treated with 27 to 36 cGy to the abdomen. Stage B seminomas are treated with abdominal radiation therapy and/or chemotherapy with VP-16 and *cis*-platinum. Stage C seminomas receive chemotherapy alone. Stage B and C embryonal cell tumors are treated with chemotherapy alone or followed by retroperitoneal lymph node dissection. Teratocarcinomas are managed with retroperitoneal lymph node dissections. After puberty, teratomas usually are invasive and should be treated in the same fashion as teratocarcinomas. Choriocarcinomas are almost always fatal, but node dissection and chemotherapy are advised.

Prognosis

The 5-year survival depends on the cell type and stage of disease at the time of operation. Currently acceptable 5-year survival figures are listed in Table 38-5.

Patients who are free of disease for 2 years are usually cured. Chemotherapy for germ cell neoplasms has doubled the survival for patients with advanced disease. Five-year survivals have been in the 90 to 95 percent range for patients with stage A and B nonseminomatous tumors. Ominous signs include high markers at presentation (i.e., beta-chain b-hCG > 8000mIU/mL, AFP > 1000IU/mL), failure of markers to normalize, development of intracranial disease, or increasing size of the tumor following node dissection and chemotherapy.

Subsequent therapy depends on the histologic diagnosis. Additional surgical treatment is of little use with choriocarcinoma because this tumor is associated with pulmonary metastases in 81 percent of patients. Pulmonary lesions are present in only 19 percent of patients with seminoma, but bilateral retroperitoneal lymph node resection appears unnecessary for this lesion because it is extremely radiosensitive. Lymph node resection usually is indicated for adult teratomas, embryonal carcinoma, and teratocarcinoma without supradiaphragmatic spread. Lymphadenectomy in such patients increases the 5-year survival and aids in accurate staging of the disease, although observation alone may be considered for select stage A tumors after orchiectomy.

The use of such chemotherapeutic agents as vinblastine, actinomycin, bleomycin, *cis*-platinum, and other drugs in combinations has markedly increased the survival of patients with advanced metastatic testicular tumors. These agents may cause severe hemopoietic changes, especially when combined with x-ray therapy. Adjuvant chemotherapy has been responsible for doubling the number of long-term survivors with advanced disease. Tumor markers (b-hCG and AFP) are valuable aids in diagnosis and follow-up. Figure 38-40 offers an algorithm for the management of the patient with a lesion of the testis.

Carcinoma of the Penis

Carcinoma of the penis develops in the squamous epithelium of the glans and foreskin. The risk of squamous cell carcinoma of the penis is almost eliminated by circumcision in infancy, and the incidence is decreased when the procedure is carried out in adulthood. The lesion is uncommon in the United States, and the average age of onset is over 60. Patients characteristically present with a sore that does not heal or a palpable nodule under the nonretractable foreskin (phimosis). Because of poor hygiene and associated infection, the diagnosis may be difficult. Biopsy provides a definitive diagnosis. Squamous cell carcinoma should be treated by local excision or partial or complete penilectomy and is associated with a 90 percent 5-year cure rate when no distant spread is present. With nodes involved by tumor, the 5-year survival is reduced to 32 percent, and excision of both inguinal and femoral nodes is attended by improved survival.

SEXUAL DISORDERS

Priapism

Priapism is defined as prolonged erection without sexual excitement. It results from a variety of vascular disorders. Blood is trapped within dilated arterioles of the erectile tissue by venous occlusion and cannot escape. Priapism differs from true erection in that the glans is not erect and the bulbospongiosus complex is not erect, whereas the corpora cavernosa are.

The most common cause currently is the injection of erection-producing agents by the patient. The antidote is to slowly inject a few milliliters of 1:100,000 epinephrine directly into the corpora to constrict the arterioles and permit spontaneous drainage. If this fails, injection of heparinized saline can be effective, or incision and drainage of the corpora can reverse the process. The Winter procedure, establishing a fistula between the corpora and the glans by incision or with a Tru-Cut biopsy needle, has been successful. Open operations that create shunts between the corpora cavernosa and corpora spongiosa or saphenous vein are seldom used.

Impotence

Impotence is defined as the inability to generate or maintain an erection adequate for penetration and sexual intercourse. Orgasm is a separate phenomenon, as is ejaculation. In evaluating the patient, history taking should pay particular attention to the use of antidepressant medications, antihypertensives, and other cardiovascular drugs. A thorough physical examination is essential, and laboratory tests for diabetes and vascular disease should be assessed.

Erection is enhanced by α -adrenergic blockers, e.g., dibenzylamine, prostaglandin E₁, phentolamine, and papaverine, and is abated by sympathomimetic agents such as epinephrine. Orgasm is mediated through higher cortical centers. A man in whom all major pelvic motor nerves have been severed can experience an orgasm if cutaneous sensitivity persists. This has been noted in men who have undergone pelvic exenteration and insertion of a penile prosthesis. Dry (no ejaculate) orgasm is possible after removal of the entire bladder, prostate, and rectum.

Erection can be effected by use of a vacuum tube device, injection of prostaglandin E₁, or implantation of a solid or inflatable penile prosthesis. Vascular anastomosis for revascularization, i.e., inferior epigastric artery to penile artery or inferior epigastric artery to dorsal vein, is indicated only in younger men with traumatic arterial injury. Nitric Oxide (NO) has been demonstrated to enhance erections during sexual stimulation. The release of NO activates the enzyme guanylate cyclase, which results in increased levels of cyclic guanosine monophosphate (cGMP). This increased level of cGMP produces smooth muscle relaxation in the corpus cavernosum allowing the inflow of blood into the cavernous tissue. Sildenafil (Viagra™) enhances the effect of NO by inhibiting the enzyme which degrades cGMP. This is the first form of oral therapy for erectile dysfunction that has been proved to be effective following sexual stimulation.

Infertility

In evaluating infertility, both partners must be considered. A careful history must be taken from the man and the woman. Does each understand the appropriate timing of sexual activity for conception? Does the woman menstruate regularly? When did her menarche take place? Has she missed periods? Has she noted mucus show midway between her periods? If the patient's menarche occurred between 11 and 15 years of age, if she has not missed any periods, and if she menstruates regularly and has noted midcycle mucus show, there is a 95 percent chance that she is ovulating. A careful pelvic examination and an ultrasonographic examination of the uterus and ovaries determine the presence of any anatomic abnormality.

The man is exclusively at fault in 20 percent of the cases of infertility and a contributing factor in another 30 percent. His history should include questions related to exposures to unusual heat, including the frequent use of hot baths, and hernia repair during infancy. The medications that he has taken should be defined. Specific questions are important to determine if he experiences retrograde or premature ejaculation and if he achieves normal penetration. Physical examination should assess the testis, vas deferens, and epididymis bilaterally and the prostate.

The semen is best taken after 3 days of sexual abstinence. The specimen is allowed to sit at room temperature approximately 30 minutes for liquefaction before microscopic examination. Patients are classified as azoospermic, oligospermic (<20 million sperm/mL), or normospermic. Normally, there are 30 to 100 million sperm/mL, with at least 60 percent showing purposeful motility and less than 40 percent abnormal forms. The ejaculate should be checked for fructose, the absence of which suggests ductal obstruction. Luteinizing hormone (LH), follicle-stimulating hormone (FSH), testosterone, and prolactin levels should be measured. An FSH level over two times normal indicates a germ cell failure that is generally not reversible by treatment.

Failure to ejaculate can be caused by drugs such as ab blockers, an incompetent bladder neck following transurethral resection, advanced pelvic atherosclerosis, or diabetes mellitus. The causes of azoospermia include ductal obstruction, germinal aplasia, drug toxicity, and Klinefelter's syndrome. Oligospermia can be a consequence of varicocele, cryptorchidism, or endocrine disorders or the sequela of juvenile mumps orchitis. Infertility associated with normal sperm, sperm counts, and sperm motility can be caused by abnormal sexual habits, prostatitis, gynecologic abnormalities, antisperm antibodies, or a defective acrosomal reaction.

If the man is oligospermic and the couple has tried to conceive for a couple of years under a physician's direction, in vitro fertilization or the use of pooled donor insemination can be tried, with a 15 to 20 percent anticipated success rate.

TRAUMA

Urologic injuries are rarely life-threatening, so the urologic surgeon may not be involved in the early resuscitative efforts of the trauma victim. However, the appropriate management of urologic injuries in the early phase of care may have a significant impact on the long-term outcome and quality of life of these patients. As such, timely urologic consultation is paramount in ensuring a good overall outcome.

Urologic trauma is classified according to the location of injury: upper urinary tract (kidneys and ureters), lower urinary tract (bladder and urethra), and external genitalia (penis, scrotum, and testes).

Kidney

The kidney is the most commonly injured structure in the urinary tract, accounting for 1 percent of all traumatic injuries. Nevertheless, operative intervention is pursued in only approximately 10 percent of renal injuries. The trend toward nonoperative management of renal injuries is the consequence of intense scrutiny of the outcomes of operative and nonoperative management; the end result is an algorithm of care for the trauma victim with upper tract injury that relies heavily on complete radiographic assessment.

The mechanism of renal injury, whether blunt or penetrating trauma, is critical to the decision-making process. The majority of renal injuries (80 to 90 percent) result from blunt trauma, which encompasses direct blows to the flank, falls, and motor vehicle accidents; deceleration injuries are particularly dangerous because they may lead to renal vascular injuries, such as renal artery thrombosis or intimal tearing. Penetrating trauma consists of stab wounds and gunshot wounds.

Initial Evaluation

The first step in the evaluation of the renal trauma victim is to obtain a careful history. The size of the weapon or caliber of the bullet in penetrating trauma, as well as the location of injury (chest, flank, abdomen) in blunt or penetrating trauma, is important in assessing the likelihood and severity of injury. In motor vehicle collisions, the speed of the vehicle and the position of the victim in the car can raise the index of suspicion of renal injury. A history of congenital renal anomalies should be elicited because of an increased likelihood of injury with even seemingly minor trauma. Physical examination includes careful inspection of the chest, abdomen, thorax, and flank for external signs of trauma such as bruising, entry or exit wounds, rib fractures, or abdominal distention.

The urinalysis is the single most important laboratory test in the evaluation of renal injuries. In most patients, a urethral catheter will be placed soon after arrival to the trauma unit, providing early assess to a urine specimen; the exception is the patient with suspected urethral injury, in whom a catheter is not passed routinely. The urine "dipstick" is adequate for the detection of hematuria and precludes the need for microscopic urinalysis, which is time-consuming and inconvenient.

Hematuria will be present in more than 95 percent of patients with renal injury and is the single most important indicator of injury to the kidney. The degree of hematuria, however, does not precisely correlate with the severity of injury. Indeed, hematuria may be absent in a small percentage of renal injuries. In particular, renovascular injuries have no associated hematuria in 24 to 60 percent of patients. As such, a high index of suspicion must be maintained in order to determine which patients need further radiographic assessment.

Radiographic Evaluation

The goal of radiographic assessment is to provide complete and accurate staging of renal injuries in order to determine the need for and to plan operative management. Most blunt renal injuries are minor, however, and can be managed conservatively. Consequently, selective imaging rather than generalized radiographic screening in adult patients with blunt renal trauma avoids low-yield studies in the majority of patients. Adult patients sustaining blunt trauma who have microscopic hematuria but no record of hypotension or shock (systolic blood pressure < 90 mmHg) in the field are exceedingly unlikely to have a major renal injury and may forego radiographic evaluation. Using these criteria, well under 1 percent of major renal injuries will be missed. These criteria do not apply to pediatric trauma patients; in this population, significant renal injury can occur even in the setting of microscopic hematuria without hypotension.

In contradistinction, patients with gross hematuria or those with microhematuria and shock are at significant risk for major renal injury; in one series, 12.5 percent of patients satisfying these criteria sustained major renal lacerations or renal vascular injury. Consequently, all adult patients with gross hematuria or microscopic hematuria and shock and all

pediatric patients with hematuria require radiographic imaging (Fig. 38-41).

Patients with penetrating trauma to the kidney require renal exploration unless radiographic imaging clearly delineates injury that can be managed safely nonoperatively in a hemodynamically stable patient. CT scanning is the imaging modality of choice in this setting. For stab wounds, if the site of penetration lies posterior to the anterior axillary line, most injuries can be managed conservatively.

Historically, the IVU has been the initial imaging modality in most cases of renal trauma. This is performed with the rapid intravenous administration of 2 mL/kg of 30% iodinated contrast material after obtaining an initial scout film encompassing the kidneys, ureters, and bladder (KUB). Ideally, the IVU should include nephrotomograms to optimally assess the renal contour. Films are obtained at 1, 5, 15, and 30 minutes after the administration of contrast material if time permits. The IVU should establish the presence of two functioning kidneys, assess renal parenchyma, and delineate the collecting system and ureters. Nonvisualization on IVU implies absence of the kidney or significant renal vascular injury; poor visualization may indicate a major parenchymal injury. Extravasation of contrast material reflects collecting system injury and should prompt further radiographic imaging studies, usually with CT.

In the unstable patient, immediate laparotomy generally precludes a preoperative IVU. In this case, if exploration reveals a retroperitoneal hematoma, an intraoperative IVU should be obtained. An on-table KUB film is obtained, after which 150 mL of iodinated contrast material is administered intravenously. A single postcontrast film at 10 or 15 minutes is usually adequate to demonstrate contralateral renal function.

CT scanning is the most informative renal imaging study because it clearly identifies retroperitoneal or intrarenal hematomas, devascularized renal segments, parenchymal lacerations, and urinary extravasation (Fig. 38-42). A completely nonenhancing kidney suggests renal artery injury. In addition, CT demonstrates associated intraabdominal injuries. Consequently, CT has largely replaced the IVU as the imaging study of choice in the stable trauma patient.

Arteriography demonstrates renovascular injury to the main or segmental renal artery, renal lacerations, and urinary extravasation. However, CT has replaced arteriography in most instances; the advantage of CT over arteriography is its noninvasiveness, the added information provided regarding associated injuries, and clear delineation of a retroperitoneal hematoma (Fig. 38-43). If an arteriogram is performed to evaluate the aortic arch, additional views of the kidney may be helpful. Currently, arteriography is most useful in the delayed setting to evaluate recurrent or persistent bleeding; selective embolization of a bleeding vessel may then be performed concurrently.

Classification

The Committee on Organ Injury Scaling of the American Association for the Surgery of Trauma devised a classification system for renal injuries (Fig. 38-44). Grade I represents renal contusion with or without subcapsular hematoma; these injuries maintain an intact renal capsule and collecting system. Grade II injuries include minor cortical lacerations without injury to the collecting system or extension into the deep medulla. Grade III injuries encompass deep parenchymal lacerations not involving the collecting system. Grade IV injuries involve deep parenchymal lacerations with extension into the medulla and collecting system or contained injury to the main renal artery or vein. Grade V injuries include fractured kidneys or renal artery or vein avulsion.

Management of Renal Injury

Absolute preoperative indications for renal exploration include hemodynamic instability due to renal hemorrhage and grade V renal injury. Intraoperatively, the observation of a pulsatile or expanding retroperitoneal hematoma mandates renal exploration. If a preoperative IVU was not obtained, an on-table one-shot IVU should be performed to document contralateral renal function prior to exploration.

A retroperitoneal hematoma encountered at the time of laparotomy should be explored if preoperative radiographic staging was incomplete. If preoperative or intraoperative staging suggests an injury that can be observed safely, the hematoma is left intact.

Relative indications for renal exploration include extensive urinary extravasation, devitalized parenchymal segments, suspicion of renovascular injury, and incomplete staging. By these criteria, grade I and II renal injuries can be managed safely nonoperatively. In contrast, a high incidence of delayed bleeding (23.5 percent) has been noted in grade III and IV injuries that were managed nonoperatively. Grade III or IV renal injuries associated with a devitalized renal segment have been shown to have a higher incidence of delayed complications such as urinoma or abscess formation. Consequently, management of grade III and IV injuries is determined by the clinical situation; if laparotomy is performed for associated intraabdominal injuries, repair of grade III and IV injuries is recommended. Urologic complications are more likely to occur when intraabdominal injuries are present. In the absence of associated injuries necessitating laparotomy, grade III and IV injuries can be observed safely with bed rest and serial hematocrit testing. Persistent bleeding or instability mandates surgical exploration or restaging. Arteriography may demonstrate a vascular lesion amenable to selective arterial embolization.

The presence of urinary extravasation does not necessarily mandate renal exploration; in most cases the collecting system will seal spontaneously. However, renal pelvis lacerations are unlikely to seal without intervention, and ureteropelvic junction avulsions require repair. Consequently, visualization of the opacified ureter on CT or IVU is critical in excluding a ureteropelvic junction avulsion. Differentiation between renal pelvic and calyceal extravasation is difficult and may not be possible in the acute setting. As such, persistent urinary extravasation should alert the clinician to a possible renal pelvis injury. In this case, placement of a retrograde ureteral catheter may allow the laceration to seal successfully, obviating the need for exploration and repair.

Major renal vascular injuries require immediate exploration and repair. Small segmental arterial injuries often can be managed nonoperatively with a reasonable expectation of an uncomplicated outcome. In some cases, selective renal arterial embolization may obviate the need for exploration in cases of significant hemorrhage.

Renal artery thrombosis is usually associated with deceleration injury and occurs when stretching of the renal artery leads to intimal tearing and thrombosis. Renal artery thrombosis is difficult to diagnose and treat. Even with prompt diagnosis, successful revascularization occurs in fewer than 30 percent of patients. Furthermore, delayed hypertension occurs commonly even in patients with presumed successful revascularization. Consequently, revascularization is reserved for patients with bilateral renal injury or injury to a solitary kidney. Historically, penetrating injury to the kidney mandated renal exploration. Current teaching, however, suggests that accurate preoperative radiographic staging, usually with CT, can distinguish injuries that require exploration and repair from those which can be managed safely nonoperatively. In the presence of associated injuries requiring laparotomy, renal exploration should be performed. This situation is more likely to occur with gunshot wounds rather than stab wounds.

Operative Management

The preferred approach for laparotomy is a midline transabdominal incision, even in cases of suspected isolated renal trauma. This approach allows inspection of intraabdominal organs and vessels in an expeditious fashion. In most cases, repair of intraabdominal organs or vessels will precede renal exploration unless hemodynamic instability is attributed to renal vascular injury or massive parenchymal bleeding.

The need for early vascular control of the renal vessels is controversial but, in general, is recommended to allow rapid control of bleeding once the hematoma is opened. In most cases, occlusion of the renal vessels is unnecessary, but in cases of massive bleeding, rapid vascular occlusion may obviate the need for nephrectomy. The renal vessels are identified by incising a “window” in the posterior peritoneum directly over the aorta, just medial to the inferior mesenteric vein. Superior dissection along the anterior surface of the aorta will lead to the left renal vein as it crosses the aorta. The renal arteries are invariably found on either side of the aorta just posterior and superior to the left renal vein. Identification of the right renal vein may require mobilization of the duodenum. Multiple renal arteries are present in 25 percent of patients and occur equally on the right and left sides. Multiple renal veins occur in 15 percent of patients and appear more commonly on the right side.

Once vascular control is attained, the colon is reflected medially, Gerota's fascia is entered laterally, and the hematoma is evacuated. The kidney is then exposed in its entirety. If the kidney is deemed salvageable, devitalized renal parenchyma is debrided sharply, preserving the renal capsule to facilitate later closure. Deep polar lacerations may devascularize a large segment of tissue and are best managed with amputation. Lacerations in the midkidney or small lacerations are debrided, collecting system violations are closed, and bleeding vessels are oversewn. The margins of the laceration are reapproximated with a running absorbable suture through the renal capsule only. If the integrity of the collecting system cannot be ascertained adequately, injection of indigo carmine–stained saline directly into the renal pelvis will facilitate identification of collecting system openings. If a large parenchymal defect is present, the capsule is closed with interrupted mattress sutures over absorbable gelatin bolsters that are wedged into the defect. In the event that no capsule is available for closure, an omental flap is swung over the defect and sutured to the remaining capsular edges. After repair, the kidney is replaced within Gerota's fascia, and a closed suction drain is placed in proximity to the kidney.

Segmental renal vein injuries may be managed with ligation of the vessel. Segmental or main renal artery and main renal vein injuries should be repaired whenever possible. If acute renal artery thrombosis is diagnosed promptly, repair may be attempted with resection of the involved segment and end-to-end anastomosis, bypass grafting with autologous vein or synthetic graft, or autotransplantation.

Complications associated with renal injuries include urinoma or abscess formation, delayed or persistent bleeding, and hypertension. Most of these complications can be managed without the need for reoperation. For urinary extravasation or urinoma formation, retrograde ureteral catheterization or percutaneous antegrade decompression of the collecting system, with or without percutaneous drainage of the urinoma, is adequate in most cases. A perinephric abscess generally also can be drained percutaneously. Delayed bleeding may be managed with selective transcatheter arterial embolization, thereby precluding the need for open exploration. The development of posttraumatic hypertension may be due to renal artery thrombosis, an ischemic renal segment, arteriovenous fistula, or a Page kidney. The hypertension may be managed medically, or partial or total nephrectomy may be necessary. Arteriovenous fistulas are managed with arterial embolization.

Ureter

Because the ureter is mobile and well protected by surrounding retroperitoneal fat and muscle, ureteral injuries are rare. The ureter may be injured, however, as a result of external violence (penetrating or blunt trauma) or by iatrogenic injury during an operation.

Initial Evaluation and Radiographic Imaging

Unfortunately, acute ureteral injuries are associated with few diagnostic hallmarks; consequently, these injuries are frequently missed until the sequelae are apparent. Blunt trauma is rarely associated with ureteral injury except in cases of ureteral avulsion, primarily in children; as such, a high index of suspicion, particularly in cases of deceleration injury, is necessary to identify these injuries. Penetrating trauma requires radiographic imaging of the upper urinary tract and should detect ureteral injury in most cases.

The intravenous urogram detects ureteral injury with a high degree of accuracy. Extravasation of contrast material is pathognomonic, but delayed opacification of the ureter or ureteral dilatation above the site of injury may raise the suspicion of ureteral injury. CT is very sensitive for detecting extravasated contrast material along the ureter and can

diagnose ureteropelvic avulsion by the absence of contrast material in the distal ureter.

Retrograde pyelography is generally reserved for delayed diagnosis of missed ureteral injuries. In the multiply injured patient, a retrograde pyelogram is impractical and usually unnecessary; most ureteral injuries are associated with intraabdominal injuries requiring laparotomy, at which time the ureter can be explored.

Operative Management

Acute ureteral injuries diagnosed at initial evaluation should be managed with open surgical repair via a midline transabdominal incision. In the absence of upper tract imaging, an on-table IVU should be obtained to evaluate the function of the contralateral kidney. If the mechanism of injury or path of the projectile placed the ureter at risk, the ipsilateral ureter should be mobilized and inspected in its entirety. Intravenous administration of indigo carmine or direct injection of indigo carmine–stained saline into the renal pelvis can identify subtle ureteral injuries. Alternatively, if the bladder is open, retrograde passage of a ureteral catheter can facilitate identification of a full-thickness ureteral injury.

The principles of ureteral repair include debridement of nonviable tissue, spatulation of ureteral edges, creation of a tension-free, watertight anastomosis over an internal ureteral stent, and drainage of the field with a Penrose or closed-suction drain.

Less than full-thickness ureteral injury or suspicion of blast effect in an apparently intact ureter may be treated with placement of an internal ureteral stent and a retroperitoneal drain. For a partial laceration with viable remaining ureter in continuity, the laceration should be reapproximated over a ureteral stent with interrupted absorbable sutures.

Ureteropelvic junction avulsion is managed with debridement and spatulation of the wound edges and then reapproximation. Placement of a ureteral stent and a closed-suction or Penrose drain is advisable.

Ureteral injury above the iliac vessels usually can be managed with primary ureteroureterostomy (Fig. 38-45). If a large defect precludes primary reanastomosis despite mobilization of the kidney, a transureteroureterostomy may be considered. However, it is always preferable to avoid involvement of the noninjured side if at all possible. Alternative options include autotransplantation or ileal ureter substitution; these maneuvers usually are reserved for elective repair of the injured ureter, and in the acute setting the ureter may be ligated and a percutaneous nephrostomy performed to establish drainage until the repair can be performed. In rare cases, if a long ureteral defect is present with significant associated vascular or intraabdominal injuries in an unstable patient, nephrectomy may be the best option.

Injury below the level of the iliac vessels is best repaired with a ureteroneocystostomy, in which healthy ureter is anastomosed to the bladder. Mobilization of the kidney and proximal ureter usually provides ample ureter to reach the bladder. If additional distance is needed, a psoas hitch is performed, in which the dome of the bladder is mobilized and then secured to the psoas tendon with several interrupted nonabsorbable sutures. Healthy proximal ureter is then reimplanted into the bladder. For longer defects, a bladder flap is rolled into a tube into which the ureter is reimplanted submucosally (Fig. 38-46).

The same principles of successful repair apply to ureteral injury caused by surgical misadventure. Recognition of these injuries is often delayed until the patient develops a urinoma, abscess, ureterocutaneous fistula, hydronephrosis, or sepsis. In the delayed setting, retrograde pyelography and CT scan can delineate the injury and associated complications. Injury diagnosed within a week of presentation can be repaired surgically at that time if no infection is present. Beyond that period, it is best to manage the injury with percutaneous nephrostomy drainage and an antegrade ureteral stent; a urinoma or abscess should be drained percutaneously. In some cases, 4 to 6 weeks of urinary diversion and stinting may allow resolution of the injury. If a ureteral stricture forms, endoscopic management with endoureterotomy or balloon dilation may be successful.

Bladder

In the adult, the bladder is an extraperitoneal organ relatively well protected from external trauma by the bony pelvis, the symphysis pubis, pelvic floor, and rectum. In contrast, the bladder is an intraperitoneal organ in the child and is susceptible to external injury.

Initial Evaluation

Penetrating trauma to the lower abdomen from gunshot or stab wounds may injure the bladder. Bladder injury may occur from bony spicules generated from a pelvic fracture. Blunt trauma to the bladder may occur from motor vehicle accidents, blows to the abdomen, falls, or crush injuries to the pelvis. A full bladder is particularly susceptible to intraperitoneal rupture from blunt trauma.

Signs and symptoms related to bladder injury are relatively nonspecific and generally consist of vague suprapubic discomfort or inability to urinate. The accumulation of intraperitoneal urine may lead to referred shoulder pain or respiratory distress. In patients at risk for bladder injury, urine should be collected for urinalysis or “dipstick” evaluation. Gross hematuria will be present in over 95 percent of patients with bladder injury.

Radiographic Evaluation

The diagnosis of bladder rupture is established by static cystogram. A Foley catheter is passed into the bladder if there is no suspicion of urethral injury (or after obtaining a retrograde urethrogram if there is), and a scout film is obtained. Approximately 100 mL of water-soluble iodinated contrast material (15 to 30 percent) is instilled into the bladder by gravity drip, and an anteroposterior film is taken. If no gross extravasation of contrast material is seen, further contrast material is instilled until a total of at least 300 mL has been given. Anteroposterior, both obliques, and postdrain films are obtained. In some cases, subtle extravasation of contrast material is identified only after the bladder has been drained of contrast material. The use of CT cystography has been advocated by some practitioners as a time- saving measure in patients already undergoing CT imaging; the bladder is distended during scanning by clamping the Foley catheter before imaging. The accuracy of this method is controversial; false-negative studies have been reported when the bladder is not actively distended before imaging. However, if retrograde cystography is performed, CT may be used to image the bladder with the expectation of a high degree of accuracy.

Blunt bladder injuries are classified according to the extent and location of injury. Bladder contusion represents a non-full-thickness injury of the mucosa or muscular layer, and no extravasation is identified on cystogram.

Intraperitoneal bladder rupture generally occurs as a result of a sudden blow to the lower abdomen or pelvis, particularly with a full bladder. The dome of the bladder represents the weakest and most vulnerable part of the bladder; as such, the sudden force is transmitted to the dome, which ruptures. On cystography, intraperitoneal extravasation of contrast material outlines loops of bowel and fills the paracolic gutters (Fig. 38-47). Intraperitoneal bladder rupture often occurs in multiply, severely injured patients and, accordingly, is associated with a high mortality rate.

Extraperitoneal bladder rupture occurs with equal frequency compared with intraperitoneal bladder rupture. Extraperitoneal rupture is closely associated with pelvic fractures. Indeed, 89 to 100 percent of extraperitoneal bladder ruptures are associated with a pelvic fracture; conversely, only 5 to 10 percent of pelvic fractures are associated with bladder rupture. The typical finding on cystogram is extravasation of contrast material into the pelvis around the base of the bladder. The characteristic “teardrop” deformity represents bladder compression by a large pelvic hematoma (Fig. 38-48). A combination of intraperitoneal and extraperitoneal bladder rupture may occur, usually in conjunction with a pelvic fracture or penetrating injury.

Treatment

Management decisions are based on the location and extent of injury. Bladder contusions require no specific therapy or short-term catheter drainage only.

Intraperitoneal bladder rupture from blunt or penetrating trauma should be explored and repaired via a transabdominal incision so as to allow exploration of the abdomen for associated injuries. If necessary, the bladder opening is extended to facilitate inspection of the bladder neck and trigone. Any extraperitoneal bladder lacerations should be closed from within the bladder using a single running layer of chromic catgut or polyglycolic suture. The ureteral orifices should be identified and ureteral injury excluded by observing efflux of clear urine from each side; if necessary, intravenous administration of indigo carmine will facilitate identification of the ureteral orifices in an injured, edematous bladder. Passage of ureteral catheters into each ureter helps identification of ureteral or ureteral orifice injury. A large-caliber (24 Fr) suprapubic catheter is placed near the dome of the bladder and brought out through a separate stab incision in the bladder and skin.

Most cases of extraperitoneal bladder rupture due to blunt trauma do not require operative intervention and will heal with 7 to 10 days of urethral catheter drainage only. If laparotomy is performed for associated intraperitoneal injuries, however, it is prudent to open the dome of the bladder and repair the bladder injury intravesically. The perivesical hematoma should be left undisturbed to avoid introducing bacteria. No correlation between the degree of extravasation and the size of bladder perforation has been found; accordingly, the decision to formally repair the bladder rather than manage it with catheter drainage should not be based on the degree of extravasation seen on cystography.

Iatrogenic endoscopic intraperitoneal bladder injuries deserve special mention. Provided the patient is hemodynamically stable and shows no signs of respiratory difficulty due to urinary ascites, drainage of the bladder with a large-caliber Foley catheter may be adequate for small perforations typical of endoscopic injury. Bladder injury occurring during laparoscopy may be repaired laparoscopically.

Complications associated with bladder rupture are usually due to failure to diagnose the injury or to adequately drain extravasated urine. Unrecognized extraperitoneal urine collection may result in pelvic abscess formation; intraperitoneal urine accumulation may lead to respiratory compromise and/or peritonitis. Delayed diagnosis of extraperitoneal bladder rupture should be managed with a large-caliber catheter in the bladder and percutaneous drainage of a pelvic abscess. Undiagnosed intraperitoneal bladder rupture requires transperitoneal exploration, drainage of urine or abscess, and closure of the bladder.

Urethra

Although anatomically the male urethra is divided into four regions (penile, bulbous, membranous, and prostatic), for the purpose of trauma, the male urethra is classified as anterior (below the urogenital diaphragm, including the penile and bulbous urethra) or posterior (above the urogenital diaphragm, including the prostatic and membranous urethra). Female urethral injuries are less common than male urethral injuries, and the only anatomic distinction of importance is to recognize bladder neck versus urethral injuries.

Initial Evaluation

Posterior urethral injuries are nearly always associated with pelvic fractures (>90 percent of cases). Crush injuries or penetrating trauma may result in disruption of the attachments of the prostate to the pelvic floor or symphysis pubis.

Anterior urethral trauma is usually associated with blunt straddle injuries in which the bulbous urethra is crushed against the pelvic arch. A direct blow to the perineum also may

cause anterior urethral contusion or laceration. Uncommonly, vigorous sexual activity with fracture of the corpus cavernosa additionally may cause blunt urethral trauma. Penetrating anterior urethral injury may occur with gunshot or stab wounds or as a result of urethral instrumentation.

The patient with a ruptured urethra may be unable to void despite the sensation of feeling full. Most patients with a urethral injury will have blood present at the urethral meatus. In some cases, perineal or penile swelling and ecchymosis will be present. A high index of suspicion of urethral injury must be maintained in any patient with a pelvic fracture.

On physical examination, suprapubic tenderness from a pelvic fracture may be elicited, and perineal or penile contusion may be obvious. On digital rectal examination, the so-called high-riding prostate may be identified, or the prostate may appear to be absent; a large pelvic hematoma causes superior displacement of the prostate when the urethra is disrupted. If blood is seen at the urethral meatus or the prostate is absent or superiorly displaced, urethral catheterization should not be attempted. In the absence of these signs, a careful attempt may be made to pass a urethral catheter.

Injury to the anterior urethra may be associated with disruption of Buck's fascia that surrounds the corporeal bodies and the corpus spongiosum of the urethra. If the urethral injury is contained by an intact Buck's fascia, blood and urine dissect along the penile shaft, producing swelling and ecchymosis of the penis. If Buck's fascia is violated, blood and urine may extravasate diffusely into the perineum, interior thigh, and lower abdominal wall and are limited only by Colles' fascia. Accordingly, a characteristic butterfly hematoma of the perineum and scrotum may be seen (Fig. [38-49](#)).

Radiographic Evaluation

Retrograde urethrography is the radiographic imaging study of choice for evaluation the patient with suspected urethral injury. In general, this study should precede any attempts to pass a urethral catheter. A scout film is obtained with the patient turned 25 to 30 degrees in the oblique position. A #14 or #16 French Foley catheter is inserted a few centimeters into the urethra, and the balloon is inflated in the fossa navicularis with 1 to 2 mL of water. Water-soluble, iodinated contrast material (15 to 20 mL) is injected into the catheter to distend the urethra, and a film is taken. Alternatively, a catheter-tip syringe can replace the Foley catheter if the meatus is closed tightly around the tip of the syringe; however, this method risks unnecessary radiation exposure to the examiner's hand. If a partial tear is present, contrast material may pass into the prostatic urethra and bladder, and extravasation will be minor. In the case of a complete tear, significant extravasation into the retropubic space is seen, and no contrast material is seen in the prostatic urethra. If no extravasation is identified, the catheter balloon is deflated, and the catheter is advanced into the bladder and the balloon reinflated. A cystogram can then be performed to evaluate the bladder.

Posterior urethral injuries are classified into three types depending on the radiographic appearance of the retrograde urethrogram. In type I injuries, the urethra is stretched from the surrounding pelvic hematoma but remains intact. Type II injuries represent partial or complete rupture of the prostatomembranous urethra at the apex of the prostate just above the urogenital diaphragm. Extravasation is seen superiorly in the pelvis but is limited inferiorly by the urogenital diaphragm. Type III injuries are most common and represent injury to the prostatomembranous urethra as well as the urogenital diaphragm and bulbous urethra. Extravasation occurs both above and below the urogenital diaphragm (Fig. [38-50](#)). With penetrating trauma, extravasation may, in addition, track along the path of penetration.

Anterior urethral injuries are classified only as complete or incomplete, according to whether the urethra remains in continuity as revealed by the flow of contrast material into the bladder on retrograde urethrography. A straddle injury with hematuria but a normal urethrogram is assumed to represent a contusion.

Treatment

Posterior Urethral Injuries

No specific therapy is required for type I urethral injuries. However, a large pelvic hematoma may compress the urethra and prevent spontaneous voiding. Therefore, an indwelling urethral catheter may be placed until the patient is able to void spontaneously.

For partial urethral tears, a careful attempt at urethral catheter placement may be performed. The risk of this maneuver is conversion of a partial tear into a complete tear. If a catheter is placed successfully, drainage is continued for 2 to 3 weeks, followed by a voiding cystourethrogram to ensure adequate healing. A safer course of action is placement of a suprapubic tube for a few weeks, followed by a voiding cystourethrogram. After placement of the suprapubic tube, a formal cystogram should be performed to rule out a bladder injury.

The specific treatment for a complete urethral laceration is controversial and ultimately may influence the likelihood of urethral stricture formation, urinary incontinence, or impotence. Options for management include suprapubic cystostomy with delayed surgical repair versus immediate surgical realignment. Immediate attempts at suture repair of the urethral ends is fraught with difficulty and is ill-advised.

In the unstable patient or the patient with complex associated injuries, suprapubic cystostomy and delayed repair constitute the option of choice. Suprapubic diversion is continued for 3 to 9 months or until the pelvic hematoma has resolved. The injury is reassessed at the time with a simultaneous antegrade and retrograde urethrogram to assess the length and location of the injury.

In the hemodynamically stable patient undergoing exploration for associated injuries, immediate surgical realignment is a reasonable option, particularly in the patient with severe prostatic displacement, major bladder neck lacerations, or associated rectal or vascular injuries. Urethral realignment can be accomplished through a variety of techniques. The bladder is opened anteriorly, and a catheter is passed retrograde through the urethra; the catheter can be guided into the bladder by a hand in the wound, or it is passed into the prevesical space, where it is recovered. A second catheter passed antegrade from the bladder into the prevesical space is also retrieved. The two catheters are secured at their tips with a suture passed through their end holes; traction on the bladder catheter guides the urethral catheter into the bladder. The balloon is inflated, and the antegrade catheter is removed. Alternatively, retrograde flexible cystoscopy may reveal the true passageway into the bladder by endoscopic and fluoroscopic guidance. A Councill catheter is passed over the guidewire and positioned in the bladder. If necessary, simultaneous antegrade and retrograde flexible endoscopy may more clearly demonstrate the posterior urethra, thereby facilitating passage of a guidewire into the bladder. A number of investigators have advocated the use of magnetic catheters or sounds to establish continuity between the anterior and posterior urethra (Fig. [38-51](#)).

Advocates of delayed surgical repair versus primary realignment cite conflicting rates of posttraumatic impotence and incontinence. The incidence of impotence and incontinence is likely a function of the severity of the trauma and not the initial treatment method.

Regardless of the choice of initial management, delayed endoscopic repair has gained popularity for the definitive management of traumatic posterior urethral strictures, which occur in all patients managed with initial suprapubic diversion. A number of endoscopic techniques for reestablishment of urethral continuity have been proposed, including "cut to the light" techniques in which simultaneous antegrade and retrograde endoscopy is used to provide a target (the light from one endoscope) toward which the incision through the other endoscope is directed. Alternatively, a guidewire or needle may be passed through the cystoscope across the urethral scar using three-dimensional fluoroscopic imaging and/or digital rectal guidance to orient instrument passage with subsequent dilation or incision of the obliterated urethra. These techniques usually rely on a period of self-catheterization after removal of the urethral catheter in order to maintain an adequate channel. Advocates of endoscopic techniques cite favorable success rates (61 to 80 percent) and preservation of potency and continence. Detractors of endoscopic means of repair cite the frequent need for repeated procedures to maintain patency.

Despite the enthusiasm for endoscopic repair of posttraumatic posterior urethral strictures, open urethroplasty remains the "gold standard" for establishment of urethral continuity. A recent review of the experience at San Francisco General Hospital revealed a remarkable 97 percent long-term success rate after perineal anastomotic urethroplasty in 82 patients. A successful outcome depends on complete excision of periurethral scar and generous mobilization of the corpus spongiosum to reduce tension on the urethral anastomosis.

Anterior Urethral Injuries

Contusion of the anterior urethra, without urethral laceration, generally requires no specific treatment. Urethral catheterization is indicated only if bladder drainage is required. Partial urethral lacerations with minimal extravasation contained by Buck's fascia can be managed with suprapubic urinary diversion via a punch cystostomy tube for 7 to 10 days. In cases in which urethral continuity is fairly well maintained, careful passage of a urethral catheter is performed, and the catheter is left indwelling for a few days. For more extensive injuries with moderate to severe extravasation and violation of Buck's fascia, formal suprapubic tube placement is recommended.

Complete disruption of the anterior urethra with extensive soft tissue injury requires debridement of devitalized tissue and suprapubic urinary diversion. These injuries may be complicated by hematoma or abscess formation, and close observation with liberal short-term use of parenteral antibiotics is recommended.

Once the soft tissue edema, urinary extravasation, and hematoma have resolved with several months of suprapubic urinary diversion, a voiding cystourethrogram is obtained to determine the length and caliber of the resulting stricture. For short strictures, endoscopic visual internal urethrotomy may be highly successful. For more extensive strictures, primary repair via the perineal route may be accomplished using a free patch graft or a pedicled flap of penile skin to widen the anastomosis.

For penetrating injuries such as stab wounds or low-velocity gunshot wounds, debridement and direct end-to-end anastomosis are advisable. These injuries may be repaired over a catheter by reapproximating spatulated ends; a suprapubic tube is not necessary.

In females, urethral injuries are rare; however, most involve the bladder neck and vagina and require immediate reconstruction in order to maintain continence.

External Genitalia

Penis

Injury to the corpora cavernosa may result in extravasation of blood and urine within or outside of Buck's fascia depending on the extent of injury (Fig. [38-52](#)). These injuries may be caused by penetrating trauma or by blunt trauma sustained during vigorous sexual intercourse when the erect penis encounters the partner's pubic bone. These injuries must be explored, usually through a circumcising incision with degloving of the penis, and the corpora repaired. Rupture of the suspensory ligament of the penis caused by sudden downward flexion of the erect penis requires exploration and repair to maintain penile stability.

Degloving injuries of the penis, usually associated with industrial accidents, mandate surgical debridement and excision of the penile skin distal to the proximal extent of injury; failure to excise distal skin invariably results in severe, permanent penile edema. Skin grafting may be performed when the wound is clean and granulating. Traumatic penile amputation is a rare event, but penile salvage is possible with microsurgical repair of the dorsal penile arteries and vein and selective skin grafting.

Scrotum

Scrotal skin is quite elastic and can be mobilized to cover extensive scrotal defects. Debridement of necrotic tissue and foreign bodies is performed. Relatively clean wounds may

be closed in layers with absorbable suture; contaminated wounds are managed with dressing changes and subsequently closed primarily or with the use of split-thickness skin grafts. If the remaining scrotal tissue is insufficient to cover the testicles, they may be wrapped daily with warm saline-soaked gauze until granulation tissue is sufficient to allow application of a split-thickness skin graft. Alternatively, the testes may be temporarily placed in thigh pockets.

Testis

Rupture of the testis may occur from either blunt or penetrating trauma. On physical examination, the scrotum is swollen, and the testicle may be indistinct because of surrounding hematoma. Scrotal sonography may help distinguish an intact from a ruptured tunica albuginea, but false-negative studies have been reported. When in doubt, the scrotum should be explored and the testicle examined. Extruded seminiferous tubules are debrided, and the tunica albuginea is closed with absorbable suture. A drain is left in place for 24 hours. A completely shattered testis may require orchiectomy, and patients always should provide informed consent for orchiectomy when undergoing scrotal exploration.

PEDIATRIC UROLOGY

Prepuce

If a boy has not been circumcised, phimosis (constriction of the meatus of the prepuce) can occur, preventing retraction of the foreskin. If smegma accumulates beneath the prepuce, adhesions can develop, preventing retraction. Inflammation of the glans, *balanitis*, can occur; a contraction ring can develop, and the inflammation can extend to the shaft of the penis, *balanoposthitis*. This is reversed by circumcision and prevented by reducing the prepuce and cleansing the glans and coronal area on a regular basis.

Congenital Anomalies

Urethral Meatal Stenosis

This occurs almost exclusively in the male and is believed to be a congenital anomaly. The diagnosis is made by calibration of the meatus using a catheter. The normal neonatal meatus should accept a #10 French sound. Meatotomy is the treatment.

Urethral Valves

These are usually seen in boys and produce variable obstructive changes. There are four anomalies of the male urethra that are not related to intersex problems. These are megalourethra with ballooning of the distal penis, anterior urethral valves, urethral duplication, and posterior urethral valves. Three types of posterior urethral valves were originally described, and a fourth type, a wind-sock valve, was added later. More than 95 percent are type 1 valves, running from the verumontanum to the distal urethra (Fig. 38-53). These are one-way valves; a catheter can be passed easily, but the valves balloon and obstruct the ureter as the child tries to void. There is often a weak or dribbling stream, and the situation can lead to uremia and hypertension. The amount of cystic renal dysplasia is the life-limiting factor (Fig. 38-54). In type 3 valves, the valve acts as a diaphragm or wind sock with a small hole in it.

Diagnosis is made by voiding cystourethrogram and endoscopy, each of which clearly depicts the site of obstruction. Dilatation of the proximal urethra and bladder and marked hydroureteronephrosis are characteristic. Management consists of destruction of the valve(s). Preliminary cystostomy or catheter drainage might be required. The prognosis depends on the degree of secondary changes in the proximal collecting system and the renal function at the time of diagnosis.

Neurogenic Bladder

This is usually due to autonomic dysfunction accompanying meningomyelocele. Impaired sensation of filling leads to a large residual urine with overflow incontinence. The patient may present with symptoms of infection, diurnal incontinence, or impaired voiding. Diagnosis is established by excretory urography, voiding cystourethrogram, and cystometry. The latter study reveals a large residual urine with low- pressure, high-volume tracing and lack of sensation of filling until large volumes have been instilled.

Treatment for mild dysfunction can be manual assistance in voiding, coupled with therapy for infection. With moderately severe involvement, treatment is directed at eliminating the residual urine. This can be accomplished by transurethral resection of the bladder neck. If the children are reminded to void at 2- to 3-h intervals, many become continent. Children can be taught intermittent catheterization when they reach 6 to 8 years of age. For severely compromised urinary tracts with marked loss of renal function, urinary diversion is indicated.

Ectopic Ureteral Orifice

The ectopic ureteral orifice usually drains the upper part of a double kidney, with complete duplication of the ureters. The condition is four times more prevalent in females. In order to be manifest symptomatically, the ectopic ureteral orifice must empty outside the urethral sphincter in the urethra or vagina. The characteristic complaint in the female is urinary incontinence despite normal voiding habits. The ureteral segment drained by the ectopic ureter is frequently not visualized on the excretory pyelogram. Treatment is excision of the upper renal segment and ureter in some instances, but at times it is possible to reimplant the ureter into the bladder or into the other pelvis of the duplicated system.

Ureterocele

This is usually associated with a stenotic ureteral orifice. Adult or orthotopic ureterocele is rarely associated with duplication and frequently contains a small stone within its lumen. Infantile or ectopic ureterocele is often associated with duplication of the drainage system. As the intravesical ureter dilates, it forms a cystic mass that can become sufficiently large to obstruct the bladder neck during voiding. Occasionally, the ureterocele prolapses through the urethra (Fig. 38-55). Diagnosis is established by means of delayed films during intravenous pyelography (Fig. 38-56). The opaque-filled intravesical mass has a thin, nonopaque rim (the ureterocele wall) that contrasts with the opacity of the bladder content, giving the appearance of “cobra-head deformity.” The pyelogram may show a large filling defect in the bladder and the functioning portion of the duplicated collecting system being pushed laterally and downward (Fig. 38-57), occasionally causing confusion with the pyelographic displacement caused by a neuroblastoma.

Many ureteroceles cause sufficient ureteral obstruction to warrant surgical treatment to preserve renal function. Small lesions can be removed by endoscopic incision or excision. Prolapsing ureteroceles with urethral obstruction require transvesical excision. In children, ureterocele is often associated with ureteral duplication and may be ectopic. Ureteral reimplantation is often required if reflux is present following endoscopic incision of the ureterocele.

Vesicoureteral Reflux

Once urine is transported from the renal pelvis to the bladder, it should not back up into the ureters or kidneys. Vesicoureteral reflux (VUR) is always abnormal. Conditions associated with VUR include posterior urethral valves, prune-belly syndrome (Fig. 38-58), complete duplication of the collecting system, ectopic ureter, ectopic ureterocele, Ask-Upmark kidney (segmental renal hypoplasia), neurogenic bladder, bladder neck obstruction, tuberculosis and other bladder infections, in some instances caused by suprapubic and indwelling urethral catheters, and bladder-urethral dyssynergia.

In adults, often there are no adverse effects of the reflux, particularly when the ureter is not dilated. In children, a number of changes occur secondary to reflux into the ducts of Bellini (reflux nephropathy). VUR runs in families, and asymptomatic reflux is common. Therefore, preadolescent siblings should be screened. A grading system is presented in Fig. 38-59. There is no direct correlation between the severity of renal damage and the grade of VUR. The diagnosis is made by VCUG using radiopaque or, preferably, radioactive material. The appearance of the ureteral orifice on cystoscopy has been correlated with the presence of reflux. The appearance has been described as normal or horseshoe, stadium, or golf-hole shaped. The golf- hole type is the most severe anatomic variety and the most apt to require surgical intervention.

In most cases, VUR can be managed without an operation. Infection occurs secondary to stasis of urine. Residual urine can be minimized by having the patient double-void, i.e., void a second time 10 minutes after completing the initial voiding. Antibiotics are administered until the reflux disappears. Two consecutive VCUGs at least 12 months apart showing no reflux are necessary to conclude that the reflux has disappeared. Recently, submucosal injection of polytetrafluoroethylene or collagen has been shown to eliminate VUR. Reimplantation of the ureter into the bladder is required almost exclusively in young children.

Ureteropelvic Obstruction

Congenital ureteropelvic obstruction is quite common (Fig. 38-60). Trichrome stains have shown that what appears to be a normal ureteropelvic junction actually can have significant replacement of the normal smooth muscle by collagen. Another finding is an anomalous vessel from the aorta to the lower pole of the kidney. During diuresis, the pelvis prolapses over this vessel, tenting and obstructing the ureter. Division of the vessel, however, will not correct the situation because the intrinsic abnormality is in the ureter.

The lesion can present in the newborn as a mass and in the adult with flank pain on hydration, the so-called beer drinker's kidney. There are often associated gastrointestinal symptoms such as nausea and vomiting. Diagnosis is established by excretory urography with delayed films noting impairment of renal excretion of contrast material. When the kidney is forming urine at low volumes, the pyelogram can appear normal.

Because the basic pathologic lesion is a deficiency of longitudinal muscle at the ureteropelvic junction, the treatment is to resect the diseased area and anastomose the renal pelvis to the healthy ureter below. Preliminary nephrostomy is seldom required. If the renal function is markedly reduced and the other kidney is normal, nephrectomy can be the procedure of choice.

Cystic Disease of the Kidney

Although clinically recognizable simple renal cysts are seldom seen in childhood, cystic disease of the kidney occurs more commonly in children. At times, the kidney can be so severely deformed as to be nonfunctioning and present as a lobular flank mass (Fig. 38-61 A). In its most advanced form of cystic renal dysplasia, no glomeruli are noted on microscopic examination (see Fig. 38-61 B). In *infantile polycystic disease* there are many cysts throughout the kidney involving all segments of the nephron (Fig. 38-62). Despite the many cysts, the capsule is smooth, and the kidney retains a normal shape. These patients are often diagnosed shortly after birth. The disease is progressive, ultimately requiring dialysis and/or transplantation.

Adult polycystic disease of the kidney is inherited as a mendelian dominant trait and usually becomes manifest when the patient is over 30 years of age. The patients are “4-H Club members”—with headache, hypertension, hematuria, and heredity. The kidneys are large and nodular; many glomeruli are present. The patients often ultimately progress to renal failure (Fig. 38-63).

Scrotal Contents

Cryptorchidism (see also Chap. 37)

Even though it looks full, the scrotum should be palpated to be sure that a testis is present on both sides. About 30 percent of premature males have an undescended testicle; for full-term infants, the incidence is 4 percent. Spontaneous descent occurs in most of these by 1 year of age, when the incidence of undescended testicle is 0.8 percent. The

cryptorchid testis must be differentiated from the retractile testis, which can be brought into the scrotum, and from the ectopic testis, which has already migrated beyond the external ring and can be found in a superficial pouch or in the perineum. The cryptorchid testis often has an associated indirect hernia that should be repaired by high ligation of the sac at the time of orchiopexy.

Nonpalpable testes are frequently found beneath the inguinal ring, but peritoneoscopy or celiotomy may be required to locate them. Bilateral cryptorchidism after 1 year should be treated initially with gonadotropins, resorting to operation in refractory cases. Rudimentary testes (2 to 3 mm) should be removed, and a prosthesis can be substituted. There is an association between cryptorchidism and testicular malignancy, abdominal testes having a higher incidence than inguinal testes. The opposite descended testis also has a higher rate of malignancy.

Testicular Neoplasms

The most common testicular neoplasm in the adult, seminoma, is rarely seen before puberty. Embryonal carcinoma or its yolk-sac variant is the most common testicular tumor of childhood. Teratoma before puberty can be a benign tumor for which high inguinal orchiectomy is curative. By contrast, teratoma after puberty should be managed as a teratocarcinoma.

The interstitial cell (nongerminal) tumor behaves differently in the child. It has the ability to make testosterone as well as possessing 21-hydroxylase activity and has the ability to produce cortisol and 11-beta compounds. In the prepubertal boy it can cause virilization, advanced bone age, male muscular pattern, and evidence of adrenal cortical excess. The tumor is rarely malignant, and high inguinal orchiectomy is curative. Feminizing tumors causing gynecomastia; decreased libido and impotence are uncommon.

Gonadoblastoma occurs more commonly in childhood, usually in patients with gonadal dysgenesis, 80 percent of whom are phenotypic females. It can be bilateral and contain Leydig cells and Sertoli cells, which can exist in the pure form. In childhood, the testis is occasionally the last area cleared in a patient treated for leukemia, and the recurrence of a mass suggests relapse.

Torsion

Torsion of the spermatic cord is classified as extravaginal or intravaginal (see Fig. 38-6). Although torsion can occur at any age, it is more common before puberty, and the extravaginal type is often seen in newborns. Torsion of the appendix testis occurs in one-third of the cases of torsion in infants and small children and is often confused with torsion of the cord and testis. The isotope scan is helpful in making the diagnosis, but getting the scan should not delay surgical repair.

At operation, the tunica vaginalis must be opened to rule out intravaginal torsion. Testicular necrosis is evaluated and a decision made as to whether to remove the testis. After detorting the cord, a bilateral orchiopexy should be performed, suturing the tunica albuginea of the testes, scrotal septum, and tunica vaginalis.

Varicocele

This results from incompetent valves or obstruction of the gonadal vein. It has been noted in 4 percent of college freshmen and occurs more commonly on the left side. The sudden appearance of a varicocele mandates ruling out the presence of a retroperitoneal tumor. After puberty, there is an association between varicocele and subfertility, with the ejaculate demonstrating an increased percentage of abnormally formed sperm. The indications for repair include size, persistent scrotal pain, rapid growth, and subfertility with abnormal sperm in the ejaculate. An atrophic testis on the side of the varicocele is a late development and not a contraindication to repair. The repair consists of ligating the spermatic vein and its collaterals and removing the varicocele.

Hydrocele and Hematocele (see Fig. 38-5)

Hydrocele is an accumulation of clear fluid within the tunica vaginalis. As the processus vaginalis descends from the peritoneum into the scrotum, it usually seals off at its upper end, leaving a two-layered sac, the tunica vaginalis, which envelops the testis. If the sac does not fuse, fluid moves back and forth from the peritoneal cavity, and a communicating hydrocele develops. In this case, the child presents with an enlarged scrotal mass that decreases in size overnight. It is repaired as an indirect inguinal hernia. Noncommunicating hydroceles are excised.

A collection of blood within the tunica vaginalis can result from trauma or from rupture of the tunica or testis. This is known as *hematocele* and should be treated by aspiration or open drainage.

Ambiguous Genitalia (see Chap. 37)

Disorders of sexual differentiation are uncommon but challenging cases. The most common cause is the adrenogenital syndrome, which leads to virilization of a genotypic female. The most common form of adrenogenital syndrome is the 46,XX/21-hydroxylase variant that accounts for 90 percent of the cases diagnosed at birth. The anatomic variation ranges from hypospadias in a dysgenetic pseudohermaphrodite to a true hermaphrodite (Fig. 38-64). There can be minimal clitoral enlargement, or the clitoris can mimic a penis. Some of these patients have defects in steroid biogenesis, and adrenal crises have been noted requiring urgent treatment.

Urethral Abnormalities

Hypospadias

This is the most frequent fusion defect of the urethra, occurring in 1 in 300 male births. The abnormality consists of a dorsal hood (absent ventral foreskin), chordee (ventral curvature of the glans penis on the shaft), and proximal location of the urethral meatus. The urethral meatus may be located anywhere from the coronal sulcus to the perineum but is more frequently distal (Fig. 38-65). The more proximal the hypospadias, the more probable is the presence of chordee. Duplication of the urethral orifice may be present, but usually only one communicates with the urethra per se.

Hypospadias results in an abnormal urinary stream. The downward deformity of the glans penis causes the urine issuing from the urethral meatus to be deflected. Infertility is associated with a scrotal or perineal meatus. Hypospadias with the urethral meatus in the scrotal area is often accompanied by bilateral undescended testes and must be differentiated from the adrenogenital syndrome and pseudohermaphroditism. Before performing urethroplasty on such a patient, determination of the genetic sex by karyotyping and adrenal function by determination of 17- hydroxysteroids and 17-ketosteroids is indicated. Visualization of the lower genitourinary tract by means of retrograde opaque contrast studies can disclose the presence of a vagina, and abdominal exploration can reveal that the child's sex organs are female rather than male, in which case excision of the hypertrophied clitoris should be performed.

Hypospadias should be repaired as soon as adequate phallic tissue is present. The repair is adapted to the anomaly. More than 80 techniques have been described. Some of the popular ones are depicted in Fig. 38-66. Treatment of hypospadias can require staged operations. First the chordee is corrected, and then the urethra is reconstructed. In some cases, stenosis of the original urethral meatus requires meatotomy before the definitive operations are undertaken. One-stage hypospadias repairs are also performed, with simultaneous chordee repair and urethroplasty. Postoperative urethral fistula formation often requires further surgical correction. A minimum period of a few months should be allowed for adequate healing between stages or before a fistula is excised.

Epispadias

This is a dorsally cleft urethra that occurs in 1 in 35,000 births, usually in association with exstrophy of the bladder. It occurs without exstrophy in 1 in 200,000 births (Fig. 38-67). It is the consequence of failure of dorsal fusion. In its most marked form it involves the urethral sphincter and is associated with urinary incontinence. Repair consists of transferring the urethral tube to its normal ventral location, rolling the corpora dorsally into position, and covering the area with skin.

Exstrophy of the Bladder

This is a rare anomaly in which the open bladder is part of the anterior abdominal wall. There is a wide range of deformity. The least complex form is represented by female epispadias with bifid clitoris and a short urethra (Fig. 38-68 A). These patients present with urinary incontinence and hematuria at the onset of menstruation. Treatment consists of trigonal tubularization and ureteric reimplantation in the bladder. Another variant is classic exstrophy, with the open bladder serving as the lower abdominal wall, cryptorchidism in males, complete epispadias, and wide separation of the symphysis (see Fig. 38-68 B). The most severe form is the cloacal type, in which there is a large omphalocele, the bladder is split, and the bowel protrudes between the split halves, with only a few centimeters of large bowel present (see Fig. 38-68 C).

Early reconstruction is indicated, often in the first 24 hours after birth if the bladder capacity is sufficient. Continence-restoring operations are deferred until later in the male. Better continence results are anticipated in girls. Cloacal exstrophy is converted to classic exstrophy by closing the two halves of the bladder, returning the bowel to the abdomen, and performing a colostomy. Urinary diversion is required in those patients in whom closure is anatomically or functionally unsuccessful.

OPERATIONS ON GENITOURINARY ORGANS

Nephrectomy

The retroperitoneal approach to the kidney was the traditional route, because in the pre-antibiotic era a major proportion of renal operations were performed for pyogenic disease, and the avoidance of contamination of the peritoneum with urine was considered advantageous. Today, the flank or lumbar approach (Fig. 38-69 A) remains popular but is supplanted on many occasions by the transabdominal route when early access to the renal vessels is required, as in renal tumor, trauma, or renal vascular disease.

Indications for the lumbar approach include inflammatory renal disease, calculi, perinephric abscess, hydronephrosis, and renal cystic disease. The disadvantages are a limited exposure for abdominal exploration, the fact that the position may be poorly tolerated by the patient, and the fact that it precludes bilateral renal or adrenal exploration. In addition, the lumbar approach makes early control of the renal pedicle without manipulation of the kidney difficult, and therefore, it should not be applied for renal parenchymal tumors.

The lumbar approach may be carried out with the patient in the lateral flexed position, with the involved side on stretch. A modified flank position can be used with a subcostal incision and by remaining retroperitoneal throughout the procedure (see Fig. 38-69 B). The external and internal oblique muscles are divided in the line of the incision. The transversus muscle is separated, in the line of its fibers, the retroperitoneum is exposed, and the peritoneum is displaced medially. The fascia of Gerota is opened, exposing the kidney and the perinephric fatty tissue.

The important structures within the renal pedicle are located anteriorly (ventrally). During mobilization of the kidney, particular care is necessary in the region of the upper pole because of the adrenal vessels and in the region of the lower pole because of aberrant arteries. Usually the renal artery and vein can be dissected free of fatty areolar tissue so that individual ligation may be performed. The vein is anterior to the artery, and the artery is ligated first if nephrectomy is performed. Under adverse circumstances, a renal pedicle

clamp can be used to control bleeding and permit transection of the hilus. Suture ligation of the renal pedicle en masse may fail to control the bleeding and can lead to arteriovenous fistula. Absorbable sutures are often used when infection is present, and drainage is instituted. The incision is closed in layers with chromic catgut or Vicryl, with the exception of the skin, which is closed with nonabsorbable sutures or staples. Nonabsorbable sutures with increased tensile strength, such as Prolene, can be used in the fascia to increase the strength of closure in debilitated patients or in those in whom poor wound healing is expected.

Nephrectomy for a renal carcinoma is carried out through either a transperitoneal, thoracoabdominal, or flank approach with vascular ligation before mobilization and without opening Gerota's fascia, and insertion of a drain is not required.

Partial Nephrectomy (Fig. 38-70)

The exposure for this operation is the same as for nephrectomy and can be done in the flank position or through a subcostal incision with the patient in the supine position. The muscles are incised, and the kidney is exposed. The vasculature is secured first by reflecting the colon and defining the renal pedicle. The ureter can be followed superiorly to a point where it passes beneath the renal vein as a convenient way of locating the pedicle. The artery and vein are then dissected out separately, and a small vascular clamp is placed on the artery. The renal capsule is then incised, and the hub of the knife is used to identify arcuate arteries as the parenchyma is traversed. These arteries are ligated. Calyces that are opened are closed. Absorbable suture must be used for this purpose to avoid stone formation within the kidney. After closure of the calyces, the reflected capsule can be closed over the lower pole of the kidney; a piece of vascularized omentum is sewed to the lower end of the kidney to cover the raw surface. A JP drain is positioned near the site of the kidney repair.

Cutaneous Ureteroileostomy (Ileal Conduit)

Currently, the most popular method of supravescical urinary diversion is cutaneous ureteroileostomy (Fig. 38-71). The major indications for this form of diversion are (1) removal of the bladder and (2) impaired bladder detrusor function. The disadvantages of the procedure are the extent of the operation and the requirement of utilizing a meticulously applied external collection device. These disadvantages are offset by the advantages of the ileal conduit in providing a continuous unobstructed drainage of urine and its versatility in coping with varying anatomic situations. Colon is occasionally used instead of ileum and offers some advantages (colon conduit). Urinary diversion using bowel with modifications to produce continence is also applicable (Koch or Indiana pouch or orthotopic neobladder).

Preoperative preparation is an integral part of the procedure, and the blood volume and cardiac, pulmonary, and renal status should be optimal. Treatment of preexisting infections or metabolic disturbances should be instituted at this time. Mechanical and antibacterial bowel preparation is used preoperatively, but some surgeons use only a mechanical prep for the ileal loop operation. The urinary appliance should be fitted and worn by the patient preoperatively to determine tolerance to adhesive and to identify the optimal location of the stoma.

The ileal loop is best performed through a midline abdominal incision. Before making this incision, it is preferable to create the abdominal wall defect for the ileostomy (see Fig. 38-71 A). This results in the creation of a tract through muscles that have not been distorted by the incision and avoids the problem of mechanical obstruction of the ileostomy. The abdomen is then entered. Appendectomy is performed. A well-vascularized segment of the ileum approximately 20 cm in length is selected for the formation of the conduit. The ileum is divided at the selected sites, and the isolated segment is irrigated with saline solution. The continuity of the ileum is restored, and the isolated ileal segment is positioned posterior to the ileoileal anastomosis (see Fig. 38-71 B). The proximal end of the ileal segment to be used for the conduit is closed with a continuous inverting chromic catgut suture. The left ureter is brought through the sigmoid mesentery and mobilized to provide length and eliminate the possibility of kinking or obstruction. The full thickness of ureter is sutured to the ileal mucosa. If the left ureter is markedly dilated, an end-to-end ureteroileostomy is performed. Ureteral stents may be used to splint the anastomosis for a few days and to encourage good position of the ureters. The right ureter is anastomosed in a similar fashion over a catheter stent (see Fig. 38-71 C, D). The isolated segment of ileum and its mesentery are sutured to the posterior peritoneum to prevent an internal hernia. The distal end of the ileal conduit is brought to the previously prepared abdominal opening. The fascia is secured to the serosa of the ileum to prevent herniation (see Fig. 38-71 E). The ileostomy stoma is everted and sutured to the skin and itself (see Fig. 38-71 F).

A 1-cm protuberant cuff under no tension is considered optimal. The wound is closed with catgut, Vicryl, or Prolene for the fascial layers and silk or staples for the skin (see Fig. 38-71 G). A nasogastric tube is usually necessary postoperatively for 3 to 5 days. Another variation of the procedure involves the joining of both ureters together with one large stoma, which is then anastomosed to the proximal end of the ileal loop as a "cap." This can provide a larger anastomotic opening between the ureters and the segment of ileum and decrease the incidence of the complication of stricture at this vital site. The stoma is usually placed on the right side but can be placed on the left side of the abdomen when indicated.

Cystostomy, Cystolithotomy

Operation on the bladder (Fig. 38-72) is facilitated by filling the organ, because this elevates the peritoneum out of the line of the incision. A low transverse abdominal incision 4 cm above the pubis is carried through the skin (see Fig. 38-72 A), subcutaneous tissue, and rectus abdominis fascia (see Fig. 38-72 B). The recti are mobilized from their midline attachments and separated (see Fig. 38-72 C). The bladder is easily recognized by its characteristic muscular and vascular pattern, and the peritoneum is reflected off the bladder. The bladder is entered longitudinally (see Fig. 38-72 D), and if calculi are present, they can be removed digitally or with stone forceps (see Fig. 38-72 E). Thorough digital and visual exploration is indicated. Drainage can be provided by a Malecot cystostomy catheter brought out through a stab wound above the incision (see Fig. 38-72 F), or the bladder can be closed primarily. A prevesical tissue drain should be used and brought out through a separate stab wound below the incision (see Fig. 38-72 G). The bladder is closed with a one-layer, continuous 0 chromic catgut suture that incorporates the seromuscular wall of the bladder. The rectus fascia is approximated with interrupted 0 chromic catgut or Vicryl sutures. The subcutaneous fat may be approximated with 3-0 interrupted plain catgut sutures and the skin edges with nonabsorbable suture material, such as 4-0 silk, nylon, or staples.

Prostatectomy

Transurethral Prostatectomy

The most commonly performed operation for removal of prostatic obstruction is the endoscopic approach. Following adequate spinal or general anesthesia and dilatation of the urethra, the resectoscope sheath is inserted. Excellent visualization through the resectoscope allows for identification of the ureteral orifices, verumontanum, and external sphincter. With a cutting loop, the prostatic tissue is resected, and hemostasis is secured with electrocoagulation. A catheter is inserted at the end of the procedure, and further hemostasis may be obtained by pulling the catheter bag against the bladder neck and prostatic fossa. This procedure is well tolerated by the patient, with minimal postoperative discomfort and early removal of the catheter. Contraindications to the transurethral resection are inability to place the patient in the lithotomy position, an excessively enlarged prostate gland, and the presence of other intravesical pathologic conditions that warrant open exploration. The skilled operator is able to avoid the complications of excessive bleeding, ureteral orifice injury, perforation, and injury to the external sphincter. The results of this procedure are gratifying, and it can be performed on elderly, debilitated patients with minimal morbidity and mortality.

Suprapubic Prostatectomy (Fig. 38-73)

This is performed through the cystostomy approach previously described. The compressed normal prostate represents the pseudocapsule from which adenomatous tissue is mobilized and enucleated by finger dissection. The prostatic urethra and adenomas are removed together. The urethral attachment is transected, and the prostatic fossa is packed while the prostatic arteries that are located on either side of the posterior bladder neck are suture-ligated. Redundant mucosal fragments or small adenomas may be removed by sharp dissection. After the pack is removed, a #22 French, 30-mL urethral retention catheter is inserted. If oozing from the bladder neck or prostatic fossa occurs, gentle traction on the urethral catheter may effect hemostasis. A Malecot catheter and tissue drain are used in similar fashion to that described for suprapubic cystostomy. The suprapubic tube is usually removed in 2 to 3 days and the urethral catheter in 5 to 7 days.

Retropubic Prostatectomy

This is also performed through the incision described for cystotomy. After the bladder wall is identified, the bladder is decompressed, and a 5-cm transverse incision is made 1.5 cm distal to the bladder neck. An alternative approach, known as a *vesicocapsular prostatectomy*, uses a vertical incision in line with the urethra and into the bladder and has the advantages of easier closure and avoidance of the anterior prostatic venous plexus. After the prostatic capsule is incised, the adenomas are mobilized by finger dissection. The distal urethra and bladder neck attachments are transected, and hemostasis is effected by suture ligation. The prostatic capsule is closed with interrupted catgut sutures, and a watertight closure is created. The bladder is drained by a urethral retention catheter. Traction on the catheter following retropubic prostatectomy is not as effective for hemostasis as it is following transvesical prostatectomy and may interfere with closure of the capsule. If additional security against catheter obstruction is desired, a suprapubic cystostomy can be performed.

Radical Retropubic Prostatectomy (Fig. 38-74)

The modifications of the usual radical retropubic prostatectomy suggested by Walsh enjoy worldwide popularity. A slight Trendelenburg position is used so that the lower abdomen is parallel to the ceiling. The abdomen is opened through a midline incision. The vas deferens is identified on either side and divided between ligatures. The lymph nodes are removed by one of many techniques. Usually the lymphatics surrounding the external iliac vein and those surrounding the obturator nerve as far distally as the exit of the nerve into the obturator foramen and the crossing of the external iliac artery by the superficial iliac vein are removed.

After the nodes are free of tumor, the prostatectomy proceeds. An incision is made through the endopelvic fascia on either side, allowing the prostate to be mobilized and allowing the operator to palpate the apex of the prostate and feel the urethral catheter entering the apex of the prostate. The plexus of Santorini is then dissected off anterior to the urethra and doubly ligated. The prostatic side is suture-ligated. Division between two of these ligatures is then carried out, and one can immediately see the urethra. The neurovascular bundles are then mobilized off the apex of the prostate. The urethra is divided, and at this time sutures are put into the urethra distally. Usually, six 2-0 chromic sutures are placed. After complete division, fragments of rectourethralis muscle are divided, and the prostate is mobilized superiorly. Nutrient vessels to the prostate are ligated between silk ligatures, taking care to preserve the neurovascular bundle back to the base of the prostate.

When the prostate has been elevated to its base, an incision is made over Denonvilliers' fascia, exposing the seminal vesicles and vasa on either side. These are divided between ligatures, taking the nutrient vessel to the seminal vesicle at its tip. Attention is then directed to the prostatovesical junction. The catheter is grasped with a Babcock clamp in the bladder at the prostatovesical junction, helping define its demarcation. Sharp dissection is used to separate the prostate from the vesical neck. The prostate, seminal vesicles, and ampullary portions of the vasa are removed and the area inspected. Hemostasis is achieved by cautery and chromic catgut ligatures. The anastomosis is then completed over a #20 French Silastic catheter by passing the distal sutures (which had been placed previously at the time the urethra was divided). Once these distal sutures have been passed through the vesical neck, the balloon is inflated. When the ligatures are then tied down, one can easily push the bladder against the pelvic outlet. Irrigation is then performed, and the wound is closed in layers. A 10-mm JP drain is placed in the retroperitoneal space near, but not against, the anastomosis to allow for aspiration of any leakage of urine postoperatively.

Hydrocelectomy (Fig. 38-75)

A vertical scrotal incision is usually employed. It is carried through the skin, dartos, and overlying coverings of the hydrocele sac. It is desirable to avoid opening the hydrocele until the sac is totally mobilized. The tunica vaginalis is then opened and the excess tunica excised. Hemostasis is secured with multiple ligatures or electrocoagulation of the small bleeders in the cut edge. The edges of the tunica may be sutured together behind the testis with a continuous 3-0 chromic suture. If hemostasis is satisfactory, this may not be necessary. A drain can be brought out through a stab wound in a dependent portion of the scrotum if the hydrocele is very large, but in most instances a drain is not required. The incision is then closed in layers with absorbable sutures. A scrotal suspensory provides an excellent dressing in the postoperative period, and the use of an ice bag decreases the swelling. Hydrocelectomy in children is carried out through an inguinal incision so that an associated hernia can be repaired at the same time.

Inguinal Orchiectomy (Fig. 38-76)

This approach is used when a testicular tumor is suspected. It provides access to the spermatic vessels before manipulation of the testis. The incision is identical with that performed for repair of a hernia (see Fig. 38-76 A), and the spermatic cord is identified at the external inguinal ring. After a rubber-shod or bulldog clamp has been applied to the cord at this level (see Fig. 38-76 B), the testis is mobilized from its scrotal attachments and exteriorized through the inguinal incision. The tunica vaginalis is opened to permit examination of the testis, and if the lesion looks or feels neoplastic, the cord is ligated and transected (see Fig. 38-76 C). Any solid lesion within the testis is considered potentially malignant, and biopsy is seldom performed. A metal clip is placed in the end of the remaining cord as a marker for future surgery or radiation therapy localization. After the testis has been removed, the inguinal incision is closed. The incision is *not* drained.

Transseptal Orchiopexy

The testis is identified through an inguinal incision that permits mobilization of the spermatic cord and correction of an indirect hernia that usually accompanies the undescended testis. After sufficient length of cord structures is mobilized to permit placement of the testis in the scrotum, a small incision is made in the contralateral scrotum, and a subdartos pouch is prepared to receive the cryptorchid testis. The scrotal septum is incised over a clamp, and traction is applied to the gubernaculum so that the testis is brought down through the scrotal septum into the contralateral scrotum. The septal defect may require a partial closure to prevent the testis from pulling back but should not be closed too tightly because strangulation of the cord could occur. The testis can be brought down on the same side without using the transseptal method if there is absolutely no tension. The scrotal and inguinal incisions are closed, and the patient is usually able to leave the hospital the same day.

Bilateral Vasectomy (Fig. 38-77)

This male sterilization procedure usually is carried out under local anesthesia in an outpatient setting. Previous consultation with the patient and his wife has allowed for careful selection, instruction to the patient, and the signing of the required papers for sterilization and informed consent. The scrotum is shaved on the night before the procedure, and the patient brings a scrotal suspensory with him to the office. The procedure is carried out with the patient in the lithotomy or supine position. The vas on one side is isolated between thumb and index finger of the operator, and the overlying skin is infiltrated with 1% lidocaine or another local anesthetic agent. The underlying cord is also infiltrated. A towel clip is then used to isolate the vas. A small incision is made over the vas, and it is delivered into the incision. An Allis clamp is then used to pick up the vas and separate it from the surrounding cord structures. The vas is then mobilized for approximately 3 cm. A segment measuring approximately 1 to 2 cm is excised between clamps. The edges of the vas are then transfixed with sutures. Another suture is placed approximately 1 cm below the previous suture and tied. The end of that suture is then passed through the end of the vas and tied, producing a means of bending the vas back on itself so that the ends are no longer in close proximity. One of the ends also can be buried in the adjacent tissues to prevent recanalization. An alternative technique involving electrocoagulation of the lumen of each end is commonly used. After satisfactory hemostasis has been secured, the ends of the vas are dropped back into the incision. The edges of the skin and underlying dartos are then approximated. A similar procedure is then carried out on the other side. Incorporating the underlying dartos in the skin closure provides excellent hemostasis. A collodion dressing is then applied, and the patient uses the scrotal suspensory over a small gauze dressing.

He is instructed to use an ice bag on that area for the remainder of the day and returns to work wearing the scrotal suspensory the following day. He is advised to expect a minimal amount of discomfort, swelling, and ecchymosis. When he resumes intercourse, he continues contraception until an examination of his semen reveals no sperm. This usually requires a period of 6 weeks or at least 10 ejaculations to evacuate the remaining sperm that are distal to the anastomotic sites.

Vasovasostomy (Fig. 38-78)

This procedure is usually carried out in-hospital with the patient under regional, general, or spinal anesthesia. An incision is made in the scrotum, and the underlying cord structures are delivered into view. The previous site of vasectomy is usually palpable unless a large segment of the vas has been excised. Use of the operating microscope facilitates the operation and improves the results. Once the ends of the vas have been isolated, they are excised back to a normal-appearing vas deferens. The ends are then carefully approximated and sutured with 8-0 or 9-0 nylon sutures in a two-layer fashion. Care must be taken to prevent tension on the completed anastomosis. Bleeding is minimal, and the reconstituted vas is then dropped back into the scrotum and the skin and dartos approximated. A similar procedure is then carried out on the other side. The patient is often able to leave the hospital the same day or the following day.

Sperm can begin to appear in the ejaculate as early as 1 month after the procedure, but most often this does not occur until 2 or 3 months following the procedure, at which time the edema may subside enough to allow passage of sperm through the small lumen of the vas deferens. Satisfactory reconstitution with sperm in the ejaculate occurs in approximately 80 percent of cases. A higher percentage of success is reported when microsurgery is used. A satisfactory surgical result does not ensure that a pregnancy will result because other factors such as sperm quality and antisperm antibodies may play a role in persistent infertility.

LAPAROSCOPY IN UROLOGY

Early urologic applications of laparoscopy were confined to diagnostic maneuvers or simple ablative procedures. As the skill and instrumentation of the laparoscopist improved, complex ablative and reconstructive procedures became feasible. Currently, the biggest obstacle to be overcome by the laparoscopic surgeon is the long operative times, and therefore increased cost, that accompany most complex laparoscopic procedures. The recent development of an automated laparoscopic suturing device for tissue approximation has already decreased operative times for a variety of reconstructive procedures, such as laparoscopic pyeloplasty. With further innovations, operative times undoubtedly will decrease, thereby advancing many procedures from the realm of feasible into established clinical practice.

Cryptorchid Testes

One of the earliest urologic application of laparoscopy was in the localization of an intraabdominal cryptorchid testis. With minimal morbidity and an average operative time of about 15 minutes, this diagnostic procedure is highly accurate and remains the procedure of choice for localization of the impalpable testis.

If an intraabdominal testis is identified, a laparoscopic approach can be used to remove the testicle or to translocate it into the ipsilateral hemiscrotum. For the testis located just inside the internal ring or the testis that has easily mobilizable vessels, a single-stage orchiopexy may be performed. If the testicle cannot be sufficiently mobilized to reach the scrotum, a Fowler-Stevens staged orchidopexy may be performed in which the spermatic vessels are laparoscopically clipped or ligated, and the testicle is left in situ to allow formation of collateral blood supply. A staged procedure is performed 6 months later, at which time the testicle is translocated into the scrotum.

Laparoscopic orchiectomy also may be performed via a similar approach. After division of the gubernaculum, the spermatic vessels and vas deferens are mobilized and divided with clips or electrocautery. The testis is delivered to the abdominal wall through a laparoscopic port site.

Varicocelectomy

Varicocele refers to a chronic dilatation of the pampiniform plexus that comprises venous drainage of the testicle. Although varicoceles are often implicated in vague testicular pain, the more relevant clinical implication of a varicocele is its potential to cause male subfertility in adults and testicular growth retardation in adolescents.

Repair of a varicocele involves exposure and ligation of the spermatic veins. Laparoscopy is ideally suited to varix ligation because the spermatic cord is one of the more easily recognizable structures laparoscopically. Use of a laparoscopic Doppler probe may facilitate identification of the spermatic artery and help prevent its inadvertent ligation. The success rate, as assessed by physical examination and color Doppler ultrasonography, is more than 98 percent.

The role of laparoscopic varicocelectomy is controversial. Alternative approaches, including an open inguinal or retroperitoneal approach or percutaneous transvenous embolism, are also associated with a high success rate, low complication rate, and rapid convalescence. Indeed, the laparoscopic approach is more time-consuming than the alternative open approaches and is no less invasive than the transvenous approach. As such, the laparoscopic approach may be best suited to bilateral varicocelectomy, in which the time savings and rapid recovery are greatest.

Pelvic Lymphadenectomy

The use of laparoscopy to sample the obturator lymph nodes in patients with adenocarcinoma of the prostate marked an early application. Previous attempts using radiographic imaging with CT, MRI, or lymphangiography yielded unacceptably high false-negative rates. A minimally invasive means of assessing the pelvic lymph nodes that could preclude a potentially morbid major operation if metastatic disease was identified was an important advance.

The role of pelvic lymph node dissection is to identify patients with clinically localized prostate cancer who are considered for potentially curative treatment (i.e., radical retropubic or perineal prostatectomy, external-beam or interstitial radiation therapy, or possibly cryosurgery) and to exclude patients with metastatic disease (i.e., positive lymph nodes), in whom curative treatment is unlikely and ill-advised. Accordingly, the selection of patients who are suitable candidates for laparoscopic pelvic lymph node dissection (LPLND) is based on identifying patients at high risk for metastatic disease who will benefit from histopathologic staging. The likelihood of positive lymph nodes is based on algorithms that take into account the grade of the tumor, the serum PSA level, and the extent of local tumor as assessed by digital rectal examination and/or transrectal ultrasonography.

LPLND may be performed via a transperitoneal or an extraperitoneal approach. The extraperitoneal approach to laparoscopic pelvic lymphadenectomy offers the advantages of a reduced formation of intraabdominal adhesions, maintaining the integrity of the peritoneal cavity and reducing the risk of tumor spillage into the peritoneal cavity. The confined working area in the retropubic space renders dissection more challenging than in the more spacious peritoneal cavity. Also, the risk of lymphocele formation is greater in the extraperitoneal approach.

The laparoscopic approach yields a lymph node packet that is comparable in content and diagnostic yield with that obtained with open surgery. Compared with open lymph node dissection, however, the laparoscopic approach is associated with less blood loss, reduced pain medication requirements, a briefer hospitalization, and a shorter convalescence. Despite these advantages, the laparoscopic operative time is longer, and the procedure is more costly compared with open surgery.

Bladder Neck Suspension

Urinary leakage resulting from the loss of normal pelvic support mechanisms is termed *stress urinary incontinence*. The resulting urethral hypermobility allows the bladder neck to assume a dependent position in the pelvis, thereby facilitating transfer of intraabdominal pressure to the bladder; consequently, increases in intraabdominal pressure, such as occurs with coughing or exercise, are associated with urinary leakage.

A number of surgical techniques are available that attempt to replace the bladder neck into a high, fixed retropubic position using a vaginal or abdominal approach. The laparoscopic approach to bladder neck suspension encompasses the high success rates of the abdominal and vaginal approaches and the minimal morbidity of the vaginal approach.

A number of laparoscopic techniques, both intraperitoneal and extraperitoneal, have been described for bladder neck suspension. In most techniques, intracorporeal knot tying is necessary to secure the endopelvic fascia to Cooper's ligament; alternatively, an absorbable clip may be applied to the end of the suture, thereby obviating the need for cumbersome intracorporeal knot tying.

A prospective comparison of laparoscopic bladder neck suspension with an open vaginal bladder neck suspension (Raz procedure) demonstrated comparable 12-month continence rates in the two groups (78 versus 82 percent, respectively). The operative time was significantly longer in the laparoscopic group than in the open group, whereas the hospital stay, pain medication requirement, and length of catheterization were greater in the open group.

Simple Nephrectomy

The development of means to entrap and morcellate tissue enabled the first laparoscopic simple nephrectomy to be performed in 1990. Since then, the procedure has been performed worldwide for a variety of disorders including end-stage renal disease, renal artery stenosis, and nonfunctioning hydronephrotic kidneys.

The retroperitoneal approach was developed recently to avoid opening the peritoneum with its attendant potential complications. The retroperitoneal approach was hindered by the inability of a pneumoretroperitoneum to create sufficient working space to remove the kidney. The development of a balloon dissector, consisting of a fingerless surgeon's glove mounted on a red rubber catheter, facilitated dissection of the pararenal space and created a working space that could accommodate the pneumoretroperitoneum; the balloon dissector is guided into the retroperitoneal space and inflated with saline to bluntly separate the retroperitoneal tissues along natural cleavage planes.

Compared with traditional open nephrectomy, transperitoneal laparoscopic nephrectomy is associated with a shorter hospital stay, less need for pain medication, and shorter convalescence. However, operative time is significantly longer for the laparoscopic approach. An overall complication rate of 12 percent has been noted in a multi-institutional review, but the complication rate predictably decreased with increasing experience.

The indications for laparoscopic nephrectomy have now been expanded to include live-donor nephrectomy and radical nephrectomy for tumor. Live- donor nephrectomy has been performed using both a transperitoneal and a retroperitoneal approach; in both cases, one port site is extended to accommodate of the intact kidney. When removing a kidney for tumor, a transperitoneal approach is recommended because the limited working space in the retroperitoneal approach is too small to accommodate the large radical nephrectomy specimen.

Adrenalectomy

Laparoscopic adrenalectomy is indicated for aldosteronoma, Cushing's disease due to an adrenal adenoma, or a greater than 6-cm nonfunctioning adrenal mass that has increased in size during a period of observation. Although anecdotal reports of laparoscopic adrenalectomy for pheochromocytoma have appeared, at this time the procedure should be considered investigational for this indication. An appropriate endocrine work-up should be performed preoperatively, and hypertension and electrolyte abnormalities should be corrected.

Both transperitoneal and retroperitoneal laparoscopic adrenalectomies have been described, although experience with the transperitoneal approach is greater. Comparison of laparoscopic adrenalectomy with its open surgical counterpart has revealed a shorter hospital stay, decreased blood loss, and more rapid convalescence in the laparoscopic group. Operative time, however, remains shorter in the open group. Reported complications in several series have been few.

ENDOUROLOGIC MANAGEMENT OF UPPER URINARY TRACT OBSTRUCTION

Upper urinary tract obstruction is the end result of a variety of etiologic factors. Congenital anatomic problems, such as intrinsic narrowing or impingement by a crossing vessel or band of tissue, may contribute to impaired flow of urine from the kidney and are typical of ureteropelvic junction obstruction. Secondary obstruction from previous surgery or radiation or extrinsic compression by a mass is characteristic of ureteral obstruction. Less commonly, secondary causes of ureteropelvic junction obstruction occur, such as previous surgery for congenital ureteropelvic junction obstruction or stricture due to previous stone disease. The distinction is relevant in that the success rates of endourologic procedures differ according to the location of the obstruction and the underlying causes.

In most cases, the diagnosis of ureteropelvic junction obstruction is obvious. Typically, a patient complains of dull flank pain, exacerbated by increased fluid intake or alcohol. An IVU or sonogram demonstrates hydronephrosis with either nonvisualization or delayed visualization of a normal-caliber ureter (Fig. 38-79). However, in some cases, the diagnosis of obstruction is not clear-cut. Accordingly, a quantitative determination of renal drainage is indicated. The diuretic, or Lasix, washout renogram can distinguish between a normal collecting system, a dilated but nonobstructed system, and a dilated, obstructed system (see Fig. 38-9). A radionuclide such as Mag3 or ^{99m}Tc DTPA is administered intravenously. At the point of peak accumulation of radionuclide in the collecting system, Lasix is administered to enhance washout of tracer from the kidney. A gamma scintillation camera quantitates the accumulation and excretion of radiotracer from the collecting system. The time required for half the radionuclide to be excreted from the kidney (*T*_{1/2}) determines the presence or absence of obstruction; a *T*_{1/2} < 15 minutes indicates no obstruction, a *T*_{1/2} > 20 minutes is indicative of obstruction, and anything in between is equivocal. The renogram also determines the differential function of the two kidneys; a functionally significant obstruction may be evident as a loss of differential function from the affected side.

If the diuretic renogram is equivocal, a pressure/flow study of the kidney (Whitaker test) may be performed. Dilute contrast material is instilled at a constant rate into the kidney through a percutaneously placed needle, and the differential pressure between the kidney and the bladder is measured. The unobstructed kidney should accommodate an instillation rate of 10 mL/min with a pressure differential of less than 15 cmH₂O.

The “gold standard” for surgical repair of the obstructed ureteropelvic junction is the dismembered pyeloplasty, in which the narrowed ureteropelvic junction is excised, along with a segment of markedly redundant renal pelvis if necessary, and the ureter is reanastomosed to the remaining renal pelvis. The procedure is performed through a flank incision, and usually an internal ureteral stent and closed-suction drain are left in place. Success rates approach 95 to 100 percent.

Endopyelotomy

The endourologic management of ureteropelvic junction obstruction is based on the principle of the Davis intubated ureterotomy: A full-thickness incision is made through a narrowed segment, and a tube is placed across the incised segment. After approximately 6 weeks, the epithelial and muscular layers of the ureter regenerate, leaving a normal-caliber ureter. The percutaneous counterpart of the intubated ureterotomy (antegrade endopyelotomy) consists of antegrade passage of a nephroscope through a percutaneous tract into the kidney, with endoscopic incision of the narrowed segment using a “cold” or “hot” knife. A nephroureteral stent or an internal ureteral stent and a nephrostomy tube traverse the incised ureter and are left in place for approximately 6 weeks. Alternatively, a retrograde approach through the ureterscope may be performed, although the procedure is more technically demanding.

In an attempt to simplify the endopyelotomy procedure and avoid placement of a nephrostomy tube, a ureteral cutting balloon catheter was developed that enabled the incision to be performed entirely under fluoroscopic control (Fig. 38-80). The cutting balloon catheter consists of a balloon mounted on a ureteral catheter; a curved cutting wire affixed to either end of the balloon is activated by an electric current when the foot pedal is depressed. The catheter is advanced over a guidewire passed cystoscopically up the ureter into the kidney, and the cutting wire is straddled across the stricture (see Fig. 38-80 *B*) Inflation of the balloon with dilute contrast material while simultaneously activating the cutting wire carries the cutting wire into the narrowed ureter and creates a precise full- thickness incision. Extravasation of contrast material on fluoroscopy confirms an adequate incision (see Fig. 38-80 *C*). A ureteral stent is left in place for 6 weeks.

The results of endopyelotomy have been good but not quite comparable with open pyeloplasty. In the largest series of antegrade endopyelotomy performed to date (401 procedures), a success rate of 85 percent was achieved. The complication rate from endopyelotomy averages 13 percent; the most feared complication is bleeding, which usually results from transection of a crossing lower-pole vessel.

Areas of current investigation include attempts to identify patients in whom endopyelotomy is most likely to be successful, thereby increasing the success rates of endopyelotomy to those achieved with pyeloplasty. Some investigators have found that the presence of a crossing vessel in the vicinity of the ureteropelvic junction, massive hydronephrosis, and a marked decrease in differential function on the affected side are all negative prognostic factors that are associated with reduced success rates. Current debate has centered on the need for preoperative imaging studies to identify patients with a crossing vessel in the vicinity of the ureteropelvic junction in order to increase the chance of success or decrease the risk of bleeding due to inadvertent incision.

Endoureterotomy

The endourologic management of ureteral strictures is similar to the treatment of ureteropelvic junction obstruction. The cutting balloon catheter may be used to incise strictures in any location in the ureter (Fig. 38-81), although extreme caution must be used when incising a stricture overlying the iliac vessels; in this case, the cutting wire must be directed anteriorly away from the vessels that lie posterior to the ureter in this location. In addition, ureteral strictures may be incised under endoscopic vision (endoureterotomy) using a rigid or flexible ureterscope and an electrosurgical probe, “cold” knife, or laser. The least invasive means of treating a ureteral stricture is balloon dilation, in which a balloon positioned across the narrowed segment is inflated in order to rupture the fibrotic stricture. All procedures have in common the placement of a ureteral stent for 3 to 6 weeks. The success rate for endoureterotomy is similar to the success rate achieved with endopyelotomy. In one large series, 80 percent of benign ureteral strictures were treated successfully by endoscopic means. Strictures at the ureteroenteric anastomosis after urinary diversion are less successfully treated; although initial success rates approached 75 percent, they continue to drop with successive years after treatment to an average of 32 percent at 3 years.

(Bibliography omitted in Palm version)

CHAPTER 39 - Gynecology

Gregory P. Sutton
Robert E. Rogers
William W. Hurd

Anatomy

External Genitalia (Vulva)

The vulva is bounded by the symphysis pubis anteriorly, the anal sphincter posteriorly, and the ischial tuberosities laterally (Fig. 39-1). The *labia majora* form the cutaneous boundaries of the lateral vulva and represent the female homologue of the male scrotum. The labia majora are fatty folds covered by hair-bearing skin in the adult. They fuse anteriorly with the anterior prominence of the symphysis pubis, the *mons veneris*. Posteriorly, the labia majora meet in a structure that blends with the perineal body and is referred to as the *posterior commissure*. Adjacent and medial to the labia majora are the *labia minora*, smaller folds of connective tissue covered laterally by non-hair-bearing skin and medially by vaginal mucosa. The anterior fusion of the labia minora forms the *prepuce of the clitoris*; posteriorly, the labia minora fuse in the *fossa navicularis*, or posterior fourchette. The term *vestibule* refers to the area medial to the labia minora bounded by the fossa navicularis and the clitoris. Both the urethra and the vagina open into the vestibule. The clitoris lies superior to the urethral meatus; the male homologue is the penis.

Skene's glands lie lateral and inferior to the urethral meatus and occasionally harbor pathogens such as *Neisseria gonorrhoeae*. Cysts, abscesses, and neoplasms may arise in these glands.

Musculature of the Pelvic Floor

The *levator ani* muscles (Fig. 39-2) form the muscular floor of the pelvis. These muscles include, from anterior to posterior, bilaterally, the *pubococcygeus*, *puborectalis*, *iliococcygeus*, and *coccygeus* muscles. The first two of these contribute fibers to the fibromuscular perineal body. The *urogenital hiatus* is bounded laterally by the pubococcygeus muscles and anteriorly by the symphysis pubis. It is through this muscular defect that the urethra and vagina pass, and it is the focal point for the study of disorders of pelvic support such as cystocele, rectocele, and uterine prolapse.

Distal or caudad to the levator ani muscles, or *levator sling*, are the superficial muscles that constitute the *urogenital diaphragm*. This structure is bounded by the ischial tuberosities inferolaterally and by the pubic arch superiorly. The lateralmost muscular components of the urogenital diaphragm are the *ischiocavernosus* muscles. These structures parallel and are attached to the inferior rami of the symphysis pubis and, like the *bulbocavernosus* muscles, contain erectile tissue that becomes engorged during sexual arousal. The bulbocavernosus muscles arise in the inferoposterior border of the symphysis pubis and around the distal vagina before inserting into the perineal body.

The *transverse perinei* muscles arise from the inferior rami of the symphysis just anterior to the pubic tuberosities and insert medially into the perineal body, lending muscle fibers to this structure as well.

Internal Genitalia

Figure 39-3 is a view from above of the internal genitalia. The central uterus and cervix are suspended by the lateral fibrous cardinal, or *Mackenrodt's, ligaments*, which insert into the paracervical fascia medially and into the muscular sidewalls of the pelvis laterally. Posteriorly, the uterosacral ligaments provide support for the vagina and cervix as they course from the sacrum lateral to the rectum and insert into the paracervical or *endopelvic fascia*. The bilateral fallopian tubes arise from the upper lateral *cornua* of the uterus and course posterolaterally and anterior to the ovaries. Each widens in the distal third, or *ampulla*. The ovaries are attached to the uterine cornu by the *proper ovarian ligaments*. These fibrous bands are analogous to the gubernaculum testis in the male and continue laterally from the uterus as the *round ligaments*. These structures exit the pelvis through the internal inguinal ring and course through the inguinal canal and external inguinal ring to the subcutaneous tissue of the mons veneris. They insert into the connective tissue of the labia majora. The ovaries are seemingly suspended from the lateral pelvis by their vascular pedicles, the *infundibulopelvic ligaments*. The peritoneum enfolding the *adnexa* (tube, round ligament, and ovary) is referred to as the *broad ligament*, although it is no more ligamentous than the peritoneum overlying the ovarian artery and vein.

The peritoneal recesses in the pelvis anterior and posterior to the uterus are referred to as the *anterior* and *posterior cul de sacs*. The latter is also called the *pouch* or *cul de sac of Douglas*. On transverse section (Fig. 39-4), several avascular and therefore important surgical planes can be identified. These include the lateral paravesical and pararectal spaces and, from anterior to posterior, the retropubic or prevesical space of Retzius and the vesicovaginal, rectovaginal, and retrorectal, or presacral, spaces. The pelvic brim demarcates the obstetric, or true, from the false pelvis contained within the iliac crests.

The muscles of the pelvic sidewall (Fig. 39-5) include the iliacus, the psoas, and the obturator internus. The blood supply is demonstrated in Fig. 39-5; with the exception of the middle sacral artery, which originates at the aortic bifurcation, the blood supply arises from the internal iliac arteries. The internal iliac, or hypogastric, arteries divide into anterior and posterior branches. The latter supply lumbar and gluteal branches and give rise to the pudendal arteries. From the anterior division of the hypogastric arteries come the obturator, uterine, superior, and middle vesical arteries.

The nerve supply to the pelvis is composed of the sciatic, obturator, and femoral nerves (Fig. 39-6). Sympathetic fibers course along the major arteries and parasympathetics form the superior and inferior pelvic plexi.

The ureters enter the pelvis as they cross the distal common iliac arteries laterally and then course inferior to the ovarian arteries and veins until they cross under the uterine arteries just lateral to the cervix. After travelling around to the cervix, the ureters course downward and medially over the anterior surface of the vagina before entering the base of the bladder.

DIAGNOSIS

Gynecologic History

The gynecologic evaluation includes a general history with special emphasis on the function of the reproductive system. The history should include

Purpose of the visit

Present illness

Menstrual and reproductive history

Past medical, surgical, obstetric, emotional, social, family and sexual history

Medications

Allergies

Family planning

Systems review

The gynecologic history should include the patient's age, date of her last menstrual period (LMP), the number of pregnancies, the number of deliveries, and the number of abortions. Gravidity, parity, and abortions are frequently indicated as G-P-A. The patient's menstrual history should include her age at the onset of menses, menstrual interval (time from the beginning of one period until the beginning of the next), number of days of flow, and some description in regard to the amount of flow (light, moderate, or heavy). The examiner should inquire as to when the patient's last cervical cytology was obtained, and, in patients over 35, the date of the patient's last mammogram should be noted. A description of the patient's current and recent contraceptive methods should be listed. The age of the patient at the time of her last menstrual period is recorded in postmenopausal patients.

Physical Examination

The initial evaluation and the presurgical work-up of a patient should include a general physical examination to include a description of the patient's

Height

Weight

Nutritional status

Blood pressure

Head and neck, including thyroid

Heart

Lungs

Lymph nodes

The gynecologic portion of the examination should document an examination of the breasts, the abdomen, and the pelvis. The pelvic examination is performed on all female patients with a pelvic complaint or on the occasion of their annual gynecologic examination. The patient is examined with her legs comfortably placed in stirrups on an examination table. A good light is essential. Instruments should be warm. Vaginal specula of several sizes must be available.

The external genitalia are inspected, noting the distribution and condition of the pubic hair. The glans clitoris, labia, urethral meatus, and the vaginal introitus are evaluated. The condition of the perineum is noted. The anus and perianal area are inspected.

The speculum is inserted into the vagina, and the vaginal walls and cervix are studied. A cervical cytology is taken at this time. The speculum is removed, and a bimanual examination of the pelvis is performed (Fig. 39- 7). The Bartholin, urethral, and suburethral areas are palpated. The vaginal walls are palpated. The cervix is examined, and its consistency, shape, mobility, and tenderness to motion are noted. The uterus and adnexae are evaluated by pressing them between two fingers of the vaginal hand and a hand placed on the lower abdomen. The size, shape, mobility, and tenderness of these organs are noted. After the bimanual examination, a rectovaginal examination should be performed. The middle finger is inserted into the rectum while the index finger is inserted into the vagina. This important maneuver allows the physician to evaluate the posterior surface of the uterus as well as the rectovaginal septum and uterosacral ligaments (Fig. 39-8).

Diagnostic Procedures

Cervical Cytology

Cervical cytology (Pap smear) should be performed beginning at 18 years of age or sooner if the patient is sexually active. Most women should have a cervical cytologic evaluation yearly at the time of their annual pelvic examination. Except in emergency situations, no female in the reproductive age group should undergo a gynecologic surgical procedure in the absence of a recent cervical cytologic evaluation. After total hysterectomy, the Pap smear should be obtained annually in patients treated for cervical neoplasia. After hysterectomy for conditions that did not include cervical neoplasia, the vaginal apex may be screened cytologically every 3 to 5 years, although the cost-effectiveness of continued screening is controversial.

Cervical cytologic specimens are obtained at the time of pelvic examination. The cervix is exposed, and the external cervix is scraped with a suitable spatula. The material is placed on a slide expeditiously fixed with any of the fixatives favored by the cytology laboratory that serves the practice. A sample of endocervical cells is then collected, either with a cotton-tipped applicator or with one of the specialized cytologic brushes. This specimen is then placed on a slide and fixed in the manner of the previous specimen.

The practitioner should expect a report from the laboratory in the format of the Bethesda classification (Table 39-1) for cervical cytologic reporting. The Bethesda system for reporting cervical cytologic diagnoses was developed in 1988 and improved in 1991; it replaced the original Papanicolaou reporting system and provides a uniform format for cytopathology reports.

All cytologic reports must be studied carefully to determine whether further evaluation or treatment is indicated (Fig. 39-9). Atypical smears or smears with severe inflammation should be repeated generally in 3 months. Persistent (two or more consecutive) atypical smears should be evaluated with colposcopic examination. All smears that indicate dysplasia or neoplasia should be investigated with colposcopy.

Colposcopy is a specialized technique that allows evaluation of the cervix under magnification, enabling the practitioner to do directed biopsies of abnormal areas. In many cases the endocervical canal, which is not directly visible to the colposcopist, is biopsied with a small curette at the time of colposcopic evaluation. Colposcopic examination is important to define the severity and size of a cervical lesion. The colposcopic examination following abnormal cervical cytology will preempt cone biopsy and allow office treatment of cervical dysplasia in most patients. Colposcopy may find a lesion too large for the ablative procedure that was planned and indicate another approach. When an endocervical lesion is found, the biopsy indicates a lesser lesion than cytologic report, or the biopsy is indicative of microinvasion of the cervix, a cone biopsy is indicated.

Office Tissue Biopsy

Biopsy of suspicious lesions of the vulva, vagina, cervix, and uterus should be obtained in the office. Vulvar biopsy is obtained by infiltrating the biopsy site with a small amount of 1% lidocaine using a 27-gauge needle. Adequate biopsies can be obtained using a dermatologic skin punch to the vulvar skin and rotating it slightly. The biopsy then is separated from its base with thumb forceps and a pair of fine scissors. Any bleeding from the biopsy site is controlled with a silver nitrate stick.

Biopsy of vaginal lesions is accomplished under local anesthesia. A 25- gauge needle is helpful to reach lesions in the middle or upper vagina. Most lesions of the vaginal wall can be obtained through the use of specialized cervical biopsy forceps or laryngeal biopsy forceps.

Biopsy of the ectocervix does not require anesthesia. Specialized cervical biopsy punches, such as the Kevorkian or Tischler type, are used. The endocervical canal should be sampled with an endocervical curette such as the Kevorkian or Duncan endocervical curette.

Biopsy of the endometrial cavity is an office procedure. It is essential to be assured that the patient is not pregnant before performing this procedure. A number of instruments are currently available for this biopsy. The Novak endometrial biopsy curette has been replaced largely by sampling devices such as the Pipelle endometrial biopsy instrument or the Vabra suction instrument; these instruments have the advantages of being narrow in caliber, fitting more comfortably into the cervical os, and being entirely disposable.

Vaginal Discharge

The patient's complaint of abnormal vaginal discharge should be investigated. Vaginal secretions that appear abnormal or have a foul odor must be studied. The pH of the vagina, which is normally between 3.8 and 4.4, may be an aid to diagnosis. A vaginal pH of 4.9 or more indicates either a bacterial or protozoal infection. The pH is obtained by dipping a pH tape in the vaginal secretions collected in the vaginal speculum.

Vaginal fluid is collected for study using a cotton-tipped applicator and transferring the sections to a small test tube containing a few drops of saline. The “wet mount” is prepared by placing a small amount of the saline suspension on a microscopic slide with a cover slip and examining it under magnification. The examiner may note motile trichomonads, indicative of *Trichomonas vaginalis*; characteristic “clue cells,” indicative of bacterial vaginosis; or pus cells, which may be indicative of a variety of vaginal, cervical, and uterine problems such as gonorrhea, chlamydial or other bacterial infections.

After the initial microscopic examination, a drop of 10% potassium hydroxide is placed on the specimen, and the vaginal material is again evaluated. Potassium hydroxide has the ability to lyse cellular material and enable the practitioner to appreciate the presence of mycelia characteristic of *Candida* vaginitis.

Cultures

Vaginal and cervical cultures are most useful for the detection of sexually transmitted disease. While the diagnosis of gonorrhea might be suspected when gram-negative intracellular diplococci are found on a vaginal smear stained by Gram stain, culture should be obtained to prove the infection. Gonorrhea is cultured on a chocolate agar plate and incubated in a reduced oxygen atmosphere. Cultures are most conveniently collected on a Thayer- Martin medium in a bottle containing a carbon dioxide atmosphere.

Chlamydial infection is suggested by the finding of a characteristic thick yellow mucus (mucopus) in the cervical canal. Mucopus should be collected with a calcium alginate–tipped swab and sent to the laboratory in transport media specifically designated for *Chlamydia*. Some laboratories are now offering urine tests for gonorrhea and *Chlamydia* utilizing the ligase chain reaction (LCR). This test offers improved sensitivity and specificity for gonorrhea and *Chlamydia*.

Pregnancy Tests

A number of pregnancy tests are available for use in the office. These tests measure increased amounts of the beta subunit of *human chorionic gonadotropin (hCG)* in urine. These urine tests are very sensitive and specific, measuring hCG as low as 50 mIU/mL. Serum tests are even more accurate and sensitive, and they have the advantage that they can be quantitated to give an hCG level. Serial hCG levels are helpful in circumstances where it is important to determine that hCG levels are increasing or decreasing, such as in the management of threatened abortion, ectopic pregnancy, or trophoblastic disease.

Abnormal Bleeding

After the first menstrual period (menarche), cyclic bleeding is considered the norm but is subject to great variation. Menstrual interval varies from 21 to 45 days (time from the beginning of one menstrual period until the beginning of another). Menstrual duration varies from 1 to 7 days. The menstrual flow is a subjective assessment and varies from light to heavy. Some women experience bleeding at midcycle at the time of ovulation. Abnormal genital bleeding falls into six categories.

Bleeding Associated with Pregnancy

The availability of extremely sensitive pregnancy tests has made it possible to confirm pregnancy in the early days of gestation. Although bleeding can occur in up to 25 percent of all normally pregnant women, this symptom must be considered a threatened abortion until the bleeding is otherwise clarified. In the presence of threatened abortion, the pregnancy test is positive, the cervix is closed, and the uterus is generally consistent with the history of gestation. A threatened abortion is considered inevitable when the cervix is dilated and fetal tissue appears at the cervical os. Abortion is incomplete after a portion of the products of conception has been expelled; it is considered complete after all the products of conception have been expelled. Inevitable and incomplete abortion is generally treated by dilatation and curettage.

Ectopic pregnancy must be considered in any patient with a positive pregnancy test, pelvic pain, and abnormal uterine bleeding. Approximately 20 percent of patients with ectopic pregnancy have no bleeding, but others might complain of bleeding from spotting or hemorrhage.

Gestational trophoblastic disease also causes abnormal bleeding associated with a positive pregnancy test. Most gestational trophoblastic disease is represented by hydatidiform mole. Molar pregnancy is suggested when the uterus is larger than would be expected from the history of gestation, vaginal bleeding, and the passage of grapelike tissue from the vagina. Quantitative gonadotropin levels are almost always greater than expected for the age of gestation. Gestational trophoblastic disease must be differentiated from normal pregnancy. Ultrasound examinations and knowing quantitative gonadotropin levels are helpful in diagnosis.

Dysfunctional Uterine Bleeding

This type of bleeding abnormality is characterized by irregular menses with occasional extended intervals of amenorrhea. When bleeding does occur after one of these periods of amenorrhea, it tends to be extremely heavy. The combination of a period of amenorrhea and extremely heavy bleeding occasionally suggests spontaneous abortion. In the majority of instances, the problem is secondary to failure to ovulate. Evaluation of these patients should include a pregnancy test, which should be negative. Endometrial sampling reveals a nonsecretory or proliferative endometrium. In the presence of extremely heavy bleeding, dilatation and curettage is required on occasion, but in most instances the condition can be managed with cyclic estrogen/progesterone treatment.

Trauma

The bleeding associated with genital trauma may be diagnosed secondary to a history of rape or genital injury. In the premenarchial female, the vaginal canal should be examined carefully for foreign bodies. In the presence of genital bleeding secondary to trauma, the lesion must be evaluated carefully and repaired in the operating room under anesthesia if necessary.

Bleeding Secondary to Neoplasm

Tumors, both benign and malignant, involving the genital tract from the vulva to the ovary, can produce abnormal bleeding. The most important tool in diagnosis is a meticulous pelvic examination that includes visualization of the vulva, vagina, and cervix and careful bimanual examination of the uterus, tubes, and ovaries. Lesions of the vulva, vagina, and cervix that produce bleeding should be biopsied to exclude malignancy.

The most common cause of abnormal bleeding in the reproductive age group is leiomyomas (fibroids). Leiomyomas are almost always benign and are a common cause of menometrorrhagia. Dilatation and curettage is helpful in diagnosing submucous uterine tumor. Pelvic ultrasound and other forms of pelvic imaging are helpful in the diagnosis of uterine fundal tumors.

The bleeding associated with tumors of the fallopian tube and ovary is generally scanty and is almost always associated with a palpable pelvic mass.

Bleeding from Infection

Bleeding is an uncommon symptom of pelvic inflammation. It is associated most often with inflammatory conditions of the vulva, vagina, and cervix. On rare occasions, patients with endometriosis and acute pelvic inflammatory disease have vaginal bleeding.

Bleeding of Nongenital Etiology

Genital bleeding can be associated with coagulopathy secondary to the use of systemic anticoagulants, clotting disorders, or blood dyscrasias.

Pain

Pelvic pain and abdominal pain are common gynecologic complaints. Pain associated with menses is the most common office complaint. Cyclic pain limited to that period just before or with the onset of menses is referred to as *dysmenorrhea*. Pain occurring without a demonstrable pathologic lesion is referred to as *primary dysmenorrhea* and is a common feature of ovulatory menstrual cycles. This condition is usually treated satisfactorily with simple analgesics. In some cases producing periodic disability, the use of ovarian suppression with oral contraceptives may be considered. *Secondary dysmenorrhea* is commonly associated with endometriosis, cervical stenosis, and pelvic inflammation.

Acute pelvic pain must be studied carefully. It may have its origin in abnormal pregnancy, benign or malignant neoplasia, or a variety of nongynecologic diseases.

Pregnancy disorders include threatened abortion, inevitable abortion, incomplete abortion, and ectopic pregnancy.

Neoplasms cause acute pain through degeneration of a myoma or torsion of a myoma or ovarian neoplasm. The spontaneous rupture of an ovarian cyst can produce severe pelvic pain. Pain associated with pelvic malignancy is a late symptom and generally follows other opportunities to diagnose this condition.

Acute pain may be caused by salpingitis or endometriosis. Pain secondary to inflammatory conditions is associated with fever and other evidence of infection in most cases. Pelvic infection secondary to *Chlamydia trachomatis* is the exception to this rule. The possibility of a nongynecologic condition as the cause of pain must be always considered. Appendicitis and other acute gastrointestinal problems are also causes of acute pelvic and abdominal pain. Patients with severe abdominal and pelvic pain should be evaluated for urinary problems such as renal and ureteral stones as well as inflammatory conditions of the bladder.

In women in the reproductive age group, a differential diagnosis commonly involves appendicitis, ectopic pregnancy, and salpingitis. Readily available, rapidly performed, sensitive, and accurate pregnancy tests have made it possible to quickly exclude the possibility of pregnancy in many situations. Bilateral low abdominal pain increased by movement of the cervix and associated with fever and leukocytosis most often indicates acute pelvic inflammatory disease. Right abdominal pain and tenderness at McBurney's point, associated with a history of gastrointestinal symptoms, on most occasions will indicate appendicitis. In many cases it may not be possible to make a definitive diagnosis in the office, and some form of exploratory operation may be required. Direct visualization of the pelvis can be carried out with a laparoscope. Acute appendicitis, ruptured tuboovarian abscess, torsion of a tube or ovary, or an ectopic gestation may indicate celiotomy. The finding of pelvic inflammation in most cases will dictate medical treatment.

Pelvic Mass

The finding of a pelvic tumor is a common event in reproductive-age women. At one time pelvic examination was the only tool for detection and diagnosis of pelvic tumors, but with increasingly sensitive imaging devices, the surgeon is called to evaluate masses that in the past escaped detection. The clinician must be aware that several physiologic conditions cause enlargement of pelvic organs. Pregnancy should be considered in all cases of uterine enlargement in reproductive-age women. Ovarian enlargement, as a result of ovulation and corpus luteum hematomas, produces masses that are easily palpable and, that may, persist for several weeks. In addition to a carefully performed pelvic examination, abdominal and vaginal ultrasonography is a useful tool. No imaging method will distinguish between benign and malignant disease, however.

Pelvic ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI) all provide clues to the origin of pelvic tumors. Uterine enlargement may suggest pregnancy, uterine myomata, adenomyosis, or malignancy such as endometrial cancer or sarcoma. Tubal tumors may represent a tubal pregnancy, inflammatory conditions of the tube and hydrosalpinx formation, or a primary fallopian tumor. Ovarian enlargement may suggest endometriosis, ectopic pregnancy, tuboovarian abscess, or benign or malignant tumor of the ovary. The decision to operate is predicated on the patient's age, clinical presentation, and character and clinical course of the mass. If the differential diagnosis points to a strong possibility of ovarian malignancy, the patient should be explored under conditions that will allow for the treatment of a pelvic cancer.

INFECTIONS

Vulvar and Vaginal Infections

Vulvar, perineal, and perianal itching and burning are symptoms that may indicate an inflammatory condition. The area is subject to most of the infections that involve skin on any other part of the body. Vulvar infection may be primary on the vulva or can originate in the vagina.

Mycotic Infection

The most common cause of vulvar pruritus is candidal vulvovaginitis. The infection is most common in patients who are diabetic, pregnant, or on antibiotics. The majority of cases are caused by *Candida albicans*, although other species might be incriminated. The most prominent symptom is itching; burning of the skin, dysuria, and dyspareunia are also common. Diagnosis is confirmed by examination of the vaginal secretions and recognition of the characteristic pseudomycelia. The condition is treated by the topical application of any one of a number of imidazole preparations. Systemic treatment is possible through the oral use of fluconazole.

Parasitic Infections

Pin worms (*Enterobius vermicularis*), which are common in young girls, cause vulvitis. Diagnosis is made by finding the adult worms or recognizing the ova on microscopic examination of perianal material collected on adhesive tape. Mebendazole therapy is indicated.

Trichomonas vaginalis causes primarily a vaginal infection, but the copious vaginal discharge causes secondary vulvitis. The patient complains of heavy, foul-smelling discharge. Diagnosis is made by recognizing the motile flagellates on microscopic examination. Treatment consists of metronidazole 250 mg given three times daily for 7 days.

The vulvar skin is a frequent site for infestation by *Phthirus pubis* (crab lice) and *Sarcoptes scabiei* (scabies, itch mites). The primary symptom of both these infestations is severe pruritus. The adult and immature forms are recognized on close inspection of the skin. Treatment consists of lindane, available for medical use as Kwell.

Bacterial Infections

Many bacteria attack the vulvovaginal region; on occasion, bacteria considered normal inhabitants of the genital tract cause symptoms. The streptococci and staphylococci are the most common offenders.

Gardnerella vaginalis is the most common bacterial pathogen. The vaginal discharge found with this condition is not unlike that found with trichomonal vaginitis. The discharge is thin, gray-green in color. The patient complains of a foul, fishy, or "dead mouse" odor. Diagnosis is made by microscopic study of the vaginal secretions to identify characteristic "clue cells." The condition is treated with metronidazole 500 mg orally every 12h for 1 week.

Viral Infections

A number of viral infections affect the vulva and vagina, the most common of these being *condyloma acuminatum*. The causative organism is the human papillomavirus. This infection has increased dramatically in the past 20 years. The lesions are characteristic wartlike growths that begin as single lesions but can grow to huge confluent lesions that distort the normal structures. The lesions enlarge rapidly in pregnancy. Diagnosis is suspected on the basis of appearance and confirmed by biopsy. Treatment depends on the destruction of the lesions with caustic agents, cryocautery, laser ablation, or electrocautery. Some large lesions could require surgical removal.

Herpes simplex infection causes painful vesicles followed by ulceration of the vulva, vagina, or cervix. Initial infection is usually widespread, but recurrent infection usually involves a single lesion. Cytologic evaluation of lesions in the vagina is helpful; culture is confirmatory for herpes infection. Once a patient is infected, there is a tendency for the lesions to

recur at various intervals for the life of the patient. The attacks may be aborted and the interval between attacks lengthened through the use of acyclovir 200 mg orally five times daily. Active infection in pregnancy carries the risk of newborn infection if the patient delivers vaginally. Cesarean section is recommended in patients in labor with vulvar or vaginal ulceration as a result of herpes simplex infection.

Molluscum contagiosum causes groups of small pruritic nodules with an umbilicated center. The lesions are treated by ablation by cautery, curettage, or corrosive medication.

Pelvic Inflammatory Disease

While pelvic inflammatory disease is basically a medical problem, it has profound surgical implications. It is estimated that there are approximately 1.5 million cases of pelvic inflammatory disease in the United States each year. This condition produced approximately 350,000 hospital admissions and could be responsible for over a hundred thousand surgical procedures annually. The condition might produce infertility in 10 percent of the cases that occur; 3 percent or more of patients will have ectopic pregnancy, and chronic pain is a problem in many others.

Pelvic inflammatory disease is largely limited to sexually active females. Several factors have been recognized as placing the patient at risk: (1) age less than 20, (2) multiple sexual partners, (3) nulliparity, and (4) previous pelvic inflammatory disease.

Pelvic inflammatory disease is classified as *acute* or *chronic*. The most common organisms that produce the condition are *N. gonorrhoeae* and *Chlamydia*, but numerous other organisms have been incriminated. Diagnosis of pelvic inflammatory disease is based on clinical findings. The classic signs include fever, lower abdominal pain with pelvic tenderness, and purulent vaginal discharge. Some patients, however, will have minimal or absent symptomatology, particularly in the presence of a chlamydial infection. The lack of symptoms does not preclude pelvic inflammatory disease and tubal damage. Those patients who present with an acute illness must be studied thoroughly to rule out the possibility of acute appendicitis, ectopic pregnancy, gastrointestinal obstruction or perforation, and urinary stones.

In patients requiring further study, laparoscopy, pelvic ultrasonography, and pelvic CT scanning may be helpful in confirming a diagnosis. When pelvic inflammatory disease is present, laparoscopy will confirm it by finding tubal edema, erythema, and exudate. The presence of a tuboovarian abscess can be confirmed in this manner. Various imaging techniques such as ultrasound and CT scanning may also confirm a pelvic abscess.

Treatment

Women with pelvic inflammatory disease can be treated as inpatients or outpatients, depending on the severity of their disease. Patients with evidence of peritonitis, high fever, or suspected tuboovarian abscess should be admitted to the hospital for observation and intravenous antibiotics. Some specialists believe that all women with pelvic inflammatory disease should be admitted to the hospital for more intensive care, which might preserve their fertility.

The Centers for Disease Control and Prevention recommendations include one of the following outpatient therapy combinations: cefoxitin 2.0 g I.M. with oral probenecid, or ceftriaxone 250 mg I.M. or equivalent cephalosporin, plus doxycycline 100 mg orally two times daily for 10 to 14 days. The first part of the therapy is aimed at *N. gonorrhoeae*. Because cefoxitin is active against the penicillinase-producing gonorrhea, this agent should be used when such strains occur. The doxycycline is added to cover *Chlamydia* either as a single pathogen or as a coexisting agent with *N. gonorrhoeae*.

Follow-up of patients treated on an ambulatory basis should be carried out within 48 to 72 h. If there is no improvement in the patient, she should be admitted for intravenous antibiotics.

Recommendations from the Centers for Disease Control and Prevention for inpatient treatment include cefoxitin 2.0 g I.V. every 6 h plus a loading dose of gentamicin 2.0 mg/kg I.V., followed by a maintenance dose of 1.5 mg/kg I.V. every 8 h. This regimen is continued for at least 48 h after the patient shows clinical improvement. Doxycycline 100 mg orally twice daily is given after the patient is discharged from the hospital, to complete a total of 10 to 14 days of therapy. An alternative regimen is clindamycin 900 mg I.V. every 8 h plus a loading dose of gentamicin 2.0 mg/kg I.V., followed by a maintenance dose of 1.5 mg/kg I.V. every 8 h. This regimen is continued for at least 48 h after the patient improves, following which the patient is discharged on doxycycline 100 mg orally twice daily to complete a total of 10 to 14 days of therapy.

Gentamicin or tobramycin in the doses listed above is given provided the patient has a normal creatinine level. If the creatinine level is elevated, an adjusted dose of gentamicin or one of the other recommended regimens is advised. It is now known that the use of broad-spectrum antibiotics, which must include an antibiotic with anaerobic activity, will result in cures. Some patients may require surgery for persistent abscess or chronic pelvic pain.

Surgical Therapy

Surgery becomes necessary under the following conditions: (1) the intraperitoneal rupture of a tuboovarian abscess, (2) the persistence of a pelvic abscess despite antibiotic therapy, and (3) chronic pelvic pain.

At one time, total abdominal hysterectomy with bilateral salpingo-oophorectomy was considered the procedure of choice when surgery for pelvic inflammatory disease was required. The availability of good antibiotics and a better understanding of the pathophysiology of the disease allow less radical surgery. In young women whose reproductive goals have not been achieved, especially in the presence of unilateral disease, a unilateral salpingo-oophorectomy may be more appropriate than total hysterectomy with removal of both ovaries and fallopian tubes.

The rupture of a tuboovarian abscess is a true surgical emergency. Physical findings are frequently nonspecific. Rupture is most frequently associated with a sudden severe increase in abdominal pain. A shocklike state commonly accompanies rupture. Leukocyte counts are not necessarily increased, and some patients are afebrile. In the days before surgical intervention for this problem was common, mortality approached 100 percent. With prompt surgical intervention and intensive medical management, the mortality rate today is less than 5 percent.

The patient with a ruptured abscess must be explored promptly through a large midline incision. Hysterectomy and oophorectomy are commonly indicated. Operation may be technically difficult because of the distortion and edema secondary to the inflammatory process. Before the extirpation of any pelvic organ, adhesions must be lysed and normal structures, such as ureters and the large and small bowel, identified. At the conclusion of the procedure, the abdomen should be liberally irrigated. If the uterus is removed, the vaginal cuff should be left open for drainage. Patients should be treated with high-dose intravenous antibiotics. Because abdominal wound infection is extremely common in these patients, the rectus fascia should be closed securely with a mass closure of the Smead-Jones type. The skin and subcutaneous tissue can be closed but frequently are left open for later delayed closure.

ENDOMETRIOSIS

Endometriosis is one of the most common conditions encountered by the pelvic surgeon. It has been estimated that endometriosis will be demonstrated in approximately 20 percent of all laparotomies in women in the reproductive age group. Although the condition occurs in teenage women, it is found most often in the third and fourth decades. Endometriosis persists into the postreproductive years.

The exact cause of endometriosis is unknown, but the most common theory is that it is initiated by retrograde menstruation. The theory is supported by the fact that it is extremely common in women who have congenital anomalies of the lower reproductive tract that would favor menstrual reflux. The most common of these anomalies is an imperforate hymen.

The most common lesions of endometriosis can be recognized as bluish or black lesions, sometimes raised, sometimes puckered, giving them a “gunpowder burn” appearance. Some lesions are white or yellow, but these are less common. The disease is found most commonly on the ovary and in many cases will involve both ovaries. Other involved organs can include the uterosacral ligaments, the peritoneal surfaces of the deep pelvis, the fallopian tubes, rectosigmoid, and a number of distant sites, including the skin or even the lungs, diaphragm, and nasopharynx.

While many patients are asymptomatic even with widespread endometriosis, others have severe pain, particularly dysmenorrhea, and dyspareunia. Other signs and symptoms depend on the location and depth of endometriotic implants. Infertility and abnormal bleeding are common problems.

The complaint of pain is common and in most cases is characteristic of the disease. Pain is associated most often with the menstrual period, characteristically beginning before the flow starts and ending when bleeding is complete. Deep pelvic dyspareunia is commonly associated with this disease, particularly in those individuals with implants involving the uterosacral ligaments or the rectovaginal septum.

The finding of a pelvic mass and tender nodularity of the uterosacral ligament strongly suggests endometriosis. The mass usually represents an ovarian endometrioma, often referred to as a “chocolate cyst” because of its dark-brown fluid contents. Endometriomas are found in approximately a third of women with endometriosis and are often bilateral. Endometriotic involvement of the skin, mucous membranes, or peritoneum is characteristically a bluish discoloration, which will bleed or cause discomfort at the time of menstruation.

Although endometriosis may be suspected on the basis of clinical findings and the patient's history, the definitive diagnosis is made visually, usually with the aid of a laparoscope. Biopsy may be helpful in atypical cases. Medical management of this condition should not be started without a confirmed diagnosis. Laparoscopy offers the best diagnostic opportunity for this disease.

Treatment

Choices of treatment include expectant management only, medical management, and surgery. Patients with minimal endometriosis who are asymptomatic can be cared for through simple observation and management with cyclic oral contraceptives and simple analgesia. The medical management of this condition involves the use of a number of agents in several pharmacologic classes. Progestins have been used for the management of endometriosis for many years. Medroxyprogesterone acetate is given orally. The agent is used in doses of 10 mg two to three times daily and frequently provides symptomatic relief.

Pseudomenopause is currently the most common medical treatment for endometriosis. The most common medications used today for this purpose are the *gonadotropin-releasing hormone agonists (GnRH-a)*. These agents produce a suppression of ovarian function by suppression of both follicle-stimulating hormone and luteinizing hormone as a result of continuous stimulation of pituitary GnRH receptors. These agents have low toxicity, and while they reliably produce the hypoestrogenic effects of hot flashes and vaginal atrophy, these symptoms are generally well tolerated. They can be given by depot injection or daily nasal spray. Because bone loss is also a result of hypoestrogenism, it is recommended

that the treatment not be continued for more than 6 months.

In the past, danazol, a weak oral androgen, was used to create pseudomenopause. This agent suppresses pituitary gonadotropins by negative hypothalamic feedback. The resulting ovarian suppression produces endometrial atrophy and regression of ectopic endometrium. Along with vasomotor symptoms and vaginal atrophy, this medication has many other symptoms, including weight gain, muscle cramps, and signs of androgen excess, including oily skin, acne, and hirsutism. For this reason, it has been replaced largely by either GnRH-a or progestin therapy for the medical treatment of endometriosis.

All these medical therapies have been well documented to result in temporary relief in patients with symptomatic endometriosis. In some patients, the effects can be relatively long lasting, but complete, permanent regression of endometriosis is rare with medical therapy. Although these treatments are used widely to enhance fertility, there is little evidence that medical therapy actually increases pregnancy rates compared with expectant therapy.

Conservative surgical therapy for endometriosis has become much more common with the advancement of laparoscopic surgery. At the time of initial diagnosis, superficial endometrial implants can be ablated with electrocautery or laser, and ovarian endometriomas can be removed. This approach appears to result in short-term enhancement of fertility and may give substantial temporary pain relief. In some cases of severe pain, deep retroperitoneal endometriosis implants can be removed either by laparoscopy or laparotomy with good results. However, as with medical therapy, conservative surgical treatment for endometriosis is palliative rather than curative in most patients.

The approach to ovarian endometriomas deserves special consideration. These “chocolate cysts” cannot be treated effectively medically. In general, even large endometriomas can be drained and the cyst lining removed laparoscopically. Although it was recommended in the past to close the ovary with several layers of absorbable sutures, it appears that this approach tends to increase postoperative adhesion formation. For this reason, it is recommended that after hemostasis is achieved, the ovary should be left open to close spontaneously. Other methods to minimize adhesion formation include atraumatic handling of the tissues and the use of a cellulose adhesion barrier (Interceed) over the surgical site. Several series document pregnancy of about 50 percent rates following conservative operation.

Extirpative surgery is the only permanent treatment for symptomatic endometriosis. Patients with symptomatic endometriosis whose reproductive goals have been achieved may have no interest in preserving their reproductive potential. In these circumstances, extirpation of the endometriosis along with the patient's fallopian tubes, ovaries, and uterus may be the best choice. If extirpative surgery has been chosen, removal of all ovarian tissue has been advocated in the past to prevent the stimulation of residual endometriosis or the development of the residual ovary syndrome. In younger patients, a normal ovary may be spared in some cases. If total hysterectomy with bilateral salpingo-oophorectomy is required, replacement hormone therapy is indicated and recurrence is uncommon. To minimize the risk of recurrent endometriosis, it is recommended that replacement hormones include daily estrogen combined with a progestin such as medroxyprogesterone acetate, 2.5 mg given orally.

ECTOPIC PREGNANCY

Ectopic pregnancy affects a large number of women of reproductive age in this country. The incidence of this condition has increased dramatically in the last two decades. Because of improvements in diagnostic and therapeutic approaches, however, maternal mortality has declined over the same period of time.

Women in the reproductive age group have an increased risk of ectopic pregnancy as they age. Women in the last 10 years of their reproductive life have more than three times the risk of women in the first 10 years of reproductive life. Black and Hispanic women have a higher risk than white women. A history of salpingitis is common in women with ectopic pregnancy. Sterilization protects against ectopic pregnancy, but when sterilization methods fail, the risk of tubal implantation is increased.

The most common complaint of patients with ectopic pregnancy is pain, frequently associated with irregular vaginal bleeding. Approximately 80 percent of affected women will recall a missed menstrual period. Physical findings include abdominal tenderness on cervical motion and adnexal tenderness on bimanual pelvic examination. An adnexal mass may be palpated in approximately 50 percent of patients. As a result of the intraperitoneal bleeding, some patients present in shock.

The most helpful laboratory examination is measurement of the beta subunit of *hCG (beta-hCG)*. Today's test, with a sensitivity of 50 mIU/mL or less, enables the surgeon to confirm the pregnant state in almost all patients at risk for ectopic pregnancy. Once the physician is assured that the patient is pregnant, it must be determined that the pregnancy is in the uterus. Pelvic ultrasonography, particularly when performed with a vaginal transducer, is proving important in differentiating uterine gestations from ectopic gestations.

If the patient's condition is not emergent, the serum level of beta-hCG at 24- to 48-h intervals is followed. In a normally implanted pregnancy, hCG levels will double every 2 days in early pregnancy. The surgeon can separate normally implanted pregnancies from those with impending abortion or those located in an ectopic site.

Ultrasonic evaluation of the pelvis is increasingly important. The vaginal probe enables the clinician to determine whether the developing pregnancy is in the uterus or in the tube at a time when the hCG levels are barely more than 1000 mIU/mL. Significant intraperitoneal hemorrhage also can be visualized by vaginal ultrasound. However, culdocentesis remains an expedient means to determine the presence of hemoperitoneum in an emergency situation. In those patients who do not desire to continue the pregnancy, curettage of the uterus with examination of the tissue can be diagnostic. In the event that fetal tissue is not found, a diagnostic laparoscopy is usually required in the symptomatic patient for definitive diagnosis. In the presence of hemodynamic instability or significant intraperitoneal bleeding that precludes adequate visualization of the pelvis, immediate laparotomy is indicated.

Treatment

Once a diagnosis of ectopic pregnancy has been established, several choices are available for treatment.

Laparoscopic Procedures

The laparoscope has been an important diagnostic tool for the last several decades, but only recently has it become the standard approach for treatment. Linear salpingostomy is the treatment of choice for ectopic pregnancies less than 4 cm in diameter that occur in the distal third (ampullary) segment of the tube. To aid in hemostasis, the mesentery below the involved tubal segment is infiltrated with a dilute vasopressin solution. The tube may then be opened in its long axis along the antimesenteric side with either a laser or a unipolar cutting cautery. The conceptus is then aspirated, and any bleeding is electrocoagulated with bipolar cautery. Closing the tube is not necessary because the tube closes spontaneously in almost every case. If hemostasis cannot be achieved, coagulation of a portion of the mesosalpinx just below the segment may be required. Partial or total salpingectomy is indicated when the pregnancy is located in the isthmic portion of the tube. Bipolar electrocoagulation is used to desiccate a short segment of fallopian tube on either side of the pregnancy, and the pregnancy and tubal segment are removed together. Larger ectopic pregnancies are managed by total salpingectomy because adequate hemostasis is difficult to achieve without extensive tubal damage. For this procedure, the mesosalpinx is serially coagulated with bipolar cautery and transected with scissors. When the uterotubal junction is reached, the tube is desiccated with bipolar cautery, and the entire tube and pregnancy are removed with the aid of a specimen bag and a large port.

Abdominal Operation

In those cases in which the surgeon has elected to perform a laparotomy, the same treatment options exist that were available laparoscopically. In a patient who desires future pregnancy, every attempt should be made to preserve a functional fallopian tube. If linear salpingostomy cannot be performed, consideration should be given to midsegment resection. Midsegment resection invariably shortens the tube but preserves the fimbria, which allows later reanastomosis. Salpingectomy should be reserved for those patients who have completed their reproductive goals, those in whom salpingostomy has failed, or those whose tube has been so completely destroyed by the ectopic gestation that it cannot be salvaged.

Medical Therapy

A relatively new approach to ectopic pregnancy is the use of methotrexate. Conservative criteria for treatment of ectopic pregnancy with methotrexate include serum beta-hCG levels less than 3500 IU/L and vaginal ultrasound that reveals the tubal pregnancy to be less than 3.5 cm in diameter with no visible fetal cardiac motion and no sign of hemoperitoneum. In this situation, studies have shown that administration of intramuscular methotrexate will result in complete resolution of the ectopic pregnancy in 96 percent of the cases. Subsequent tubal patency on the affected side can be documented in approximately 85 percent of the patients so treated. The risk of rupture and intraperitoneal hemorrhage must be made clear to the patient. In these cases, surgical management can be lifesaving. To what degree methotrexate treatment of ectopic pregnancy will replace definitive surgery has yet to be established.

PELVIC SUPPORT DEFECTS

Pelvic support defects include uterine prolapse, cystocele, rectocele, enterocele, urethral detachment, and posthysterectomy vaginal prolapse. The various types of pelvic support defects seldom occur as isolated problems. The forces that produce these defects influence all the organs in the pelvis that are susceptible to injury. Pelvic support defects may be produced by one or more of the following conditions: obstetric injury, conditions that increase abdominal pressure, obesity, decreased estrogen levels, and inherent tissue weakness secondary to genetic or nutritional factors.

Uterine Prolapse

Uterine prolapse is abnormal descent of the uterus relative to the bony pelvis and vagina. If the cervix protrudes through the vaginal introitus, the prolapse is considered partial; if the entire uterus prolapses through the introitus, the condition is considered a total prolapse. To assess this condition, the patient should be examined at rest in a lithotomy position and while straining. The uterus should descend without the use of a tenaculum.

Cystocele and Rectocele

These conditions are due to herniation of the bladder and the rectum into the vaginal canal, generally through a widened vaginal introitus. There are many classifications of these conditions; the simplest is mild, moderate, or severe. A somewhat more useful one is as follows: first degree, a slight bulge in vaginal wall; second degree, a bulge of mucosa that brings the vaginal wall to the introitus; and third degree, the bulge prolapses through the introitus.

Enterocele

An enterocele, herniation of intraperitoneal organs generally at the vaginal apex, most often follows hysterectomy. The hernia sac is lined by peritoneum and can be filled with small bowel, sigmoid colon, and occasionally uterine adnexa. Enterocèles are frequently misdiagnosed as rectoceles. The two conditions can be differentiated by placing one finger in the rectum and another in the vagina, allowing the examiner to palpate small bowel in the hernia sac.

Urethral Detachment

At one time, urethral detachment was called *urethrocele*. When the female urethra loses its normal support, it bulges into the vaginal canal. In the normal state, there is a definite angle produced by the urethra and the posterior wall of the bladder. Urethral detachment results in the angulation approaching 180°. In most cases, the urethrocele coexists with a cystocele.

Stress Urinary Incontinence

Urinary incontinence affects almost 40 percent of all women over 60 years of age and is a common problem for younger women. Some forms of incontinence are surgically correctable. Those conditions improved by operation are frequently associated with loss of the posterior ureterovesical angle, i.e., that angle produced by the posterior wall of the urethra and the posterior wall of the bladder. Mobility in this area, particularly at times of increased abdominal stress such as coughing, laughing, or sneezing, produces involuntary loss of urine.

A patient complaining of the involuntary loss of urine secondary to coughing or sneezing should be evaluated carefully. The surgeon should be certain that the condition is due to distorted anatomy and not to conditions that produce an irritable bladder syndrome.

Incontinent patients are examined initially in the lithotomy position. On straining, the patient generally will exhibit a cystocele or the loss of posterior ureterovesical angle. When asked to cough, there may be a small loss of urine. Movement of the anterior vaginal wall may be observed by placing a Sims speculum or the posterior blade of a medium Graves speculum in the vagina and asking the patient to cough. On occasion, the patient will not demonstrate incontinence in a supine position, and she must be examined in the standing position.

Before considering operation, the patient should be evaluated with a cystometrogram obtained by placing known amounts of water in the bladder and measuring bladder pressure with a manometer. The volume at which the patient feels the first urge to void and the volume at which the patient is unable to tolerate additional fluid are recorded. The amount of fluid in the bladder is reduced by 250 mL, the catheter is removed, and the patient is asked to cough. Most patients with genuine urinary stress incontinence will demonstrate loss of urine. At completion of the study, the patient is asked to void, and residual urine is evaluated by recatheterizing the patient. Normal patients will have the urge to void at 150 to 200 mL. A strong urge to void should be experienced at between 400 and 500 mL. Residual postvoid urine should be less than 50 mL.

Urethral descent may be quantified through the use of a Q-Tip test. The periurethral area is cleansed, and a lubricated Q-Tip is inserted 3 cm into the urethra. The resting angle of the Q-Tip relative to the horizontal is measured. The patient is asked to strain, and the angle that represents the maximum excursion of the Q-Tip is again measured. An excursion of over 20° represents increased urethral mobility, frequently associated with genuine urinary stress incontinence.

Therapeutic Considerations

A decision whether to treat a patient expectantly or surgically depends on the patient's symptomatology. Minor asymptomatic support defects may be treated expectantly or by pubococcygeal exercises. Pubococcygeal exercises involve contracting and relaxing the levator muscle repetitively several times daily. This intervention will not change the anatomic abnormalities of support defects but might give the patient better urinary control. Symptoms that may require surgery include urinary stress incontinence, symptomatic prolapse of the uterus, bladder, or rectum, urinary retention, vaginal ulceration due to prolapse, and constipation secondary to rectal sacculation.

BENIGN TUMORS

Ovarian Tumors

Nonneoplastic Cysts

By definition, a cystic enlargement of the ovary should be at least 2.5 cm in diameter to be termed a cyst.

Follicular Cysts

These are unruptured, enlarged graafian follicles. They grossly resemble true cystomas. They can rupture, causing acute peritoneal irritation, undergo torsion and infarction of the ovary or infarction of the tube and ovary, or spontaneously regress.

Corpus Luteum Cysts

These cysts become as large as 10 to 11 cm. They can rupture and lead to severe hemorrhage and occasionally vascular collapse from blood loss. The symptoms and physical findings of these cysts mimic those of ectopic pregnancy, and they are occasionally associated with delayed menses and spotting.

Endometriomas

These account for most “chocolate cysts” and are cystic forms of endometriosis of the ovary.

Wolffian Duct Remnants

These are not ovarian cysts but often cannot be distinguished clinically from tumors of the ovary. They are small unilocular cysts. Occasionally, they enlarge and may twist and infarct. In most instances, they are incidental findings at laparotomy and cause no difficulties or symptoms.

Müllerian Duct Remnants

These can appear as paraovarian cysts or as small cystic swellings at the fimbriated end of the fallopian tube (hydatids of Morgagni).

Nonfunctioning Tumors Cystadenomas

Serous cystadenomas appear as cysts within translucent walls containing clear fluid and lined by simple ciliated epithelium. They frequently are on a pedicle and may undergo torsion leading to pain and infarction. When encountered surgically, they are adequately treated by simple salpingo-oophorectomy. Many fluid-containing cystic tumors of the ovary are also accompanied by papillary projections and are known as *papillary serous cystadenomas*. Because of epithelial variation in these tumors, it is often difficult to be sure where they fit in the spectrum of benign to malignant disease. A similar problem of malignant potential exists for the *mucinous cystadenoma*, which is a cystic tumor containing sticky, gelatinous material. These mucinous tumors are less likely to be malignant than the serous cystadenomas. About 20 percent of the serous tumors and 5 percent of the mucinous tumors are bilateral.

It is not always possible to be sure by gross inspection whether cystic tumors with solid components are benign or malignant. It is usually necessary to excise the involved ovary completely, even though there is no definite evidence of malignancy. The malignant potential of the cystadenoma is then determined by histologic examination. Some cystadenomas are classified as borderline tumors, or adenocarcinomas of low malignant potential. These (grade 0) carcinomas usually are associated with an excellent prognosis and, if they are unilateral, may be treated by unilateral adnexectomy for women in their reproductive years. Frozen-section examination of the tumor at the time of surgical intervention is necessary to determine the proper course of therapy for patients in the reproductive age group. The opposite ovary should be inspected.

Occasionally, a condition known as *pseudomyxoma peritonei* is encountered; this is a locally infiltrating tumor composed of multiple cysts containing thick mucin. These tumors arise either from ovarian mucinous cystadenomas or from mucoceles of the appendix, both of which commonly coexist. Histologically, they are benign, but by local spread and infiltration they compromise surrounding vital structures. Localized tumors should be excised completely, if possible. Both ovaries and the appendix are removed, even though they grossly appear to be normal.

Mature Teratoma

These germ cell tumors are thought to arise from the totipotent germ cells of the ovary. The tumors often contain calcified masses, and occasionally either teeth or pieces of bone can be seen on abdominal radiographs. Mature teratomas occur at any age but are more frequent in patients between 20 and 40 years. They are benign dermoid cysts. The occasional solid teratoma is usually malignant (immature teratoma).

If a teratoma (dermoid) is encountered in a young woman, it is preferable to shell it out from the ovarian stroma, preserving functioning tissue in the affected ovary. Usually these cysts contain ectodermal, mesodermal, and endodermal tissues, in addition to a thick, greasy, fatty material. If this material is spilled during surgery, a chemical peritonitis may result; therefore, it is important to remove these tumors intact. The opposite ovary should be inspected, but no further operative procedure is performed if the opposite ovary appears normal. In approximately 12 percent of patients, these tumors are bilateral. In patients of childbearing age, some functional ovarian tissue should be preserved. Immature teratomas are treated as a carcinoma of the ovary by hysterectomy and bilateral salpingo-oophorectomy.

Brenner Tumor

These are rare epithelial tumors that usually do not secrete hormones. Histologically, the epithelial elements are similar to Walthard rests and are believed to arise from these. These tumors occur primarily in later life and have a small malignant potential. Simple oophorectomy is usually sufficient therapy, and the prognosis is excellent.

Meig's Syndrome

This pertains to ascites with hydrothorax, seen in association with benign ovarian tumors with fibrous elements, usually fibromas. It is more common to see fluid accumulation with ovarian fibromas that are more than 6 cm in size. The cause of the condition is unknown, but the ascitic fluid may originate from the tumor, as a result of lymphatic obstruction of the ovary. Frequently, this clinical picture is encountered with other ovarian tumors, especially ovarian malignancies, which can produce a cytologically benign pleural effusion; in such cases it is termed a *pseudo-Meig's syndrome*. Meig's syndrome can be cured by excising the fibroma.

Functioning Tumors Granulosa Cell–Theca Cell Tumor

Pure theca cell tumors (thecomas) are benign, but those with granulosa cell elements may be malignant. It is often impossible to predict their behavior from the histologic features, and prolonged follow-up is necessary in order to judge the nature. Usually, granulosa cell tumors elaborate estrogen, but some of these tumors have no hormone production. In young girls they are characteristically manifested by isosexual precocity, and in elderly women they are sometimes associated with endometrial carcinoma. The tumor can occur at all ages from childhood to the postmenopausal period, but it is most common in later life, with maximal occurrence between the ages of 40 and 60. If the tumor is discovered in the reproductive years and confined to one ovary without signs of surface spread or dissemination, a simple oophorectomy may be sufficient therapy. If it is discovered in later life, removal of both ovaries with the uterus is indicated.

Sertoli-Leydig Cell Tumors (Arrhenoblastomas)

These rare but potentially malignant tumors are associated with androgen output and masculinization. Rarely, they elaborate estrogen. They usually occur in the reproductive age group and appear to contain tubular structures as well as Leydig-type cells. In young patients with a single involved ovary, unilateral oophorectomy is adequate therapy, provided there is no extension of the tumor. For older patients or for those with bilateral involvement, total hysterectomy and bilateral salpingo- oophorectomy are performed.

Struma Ovarii

This term refers to the presence of grossly detectable thyroid tissue in the ovary, usually as the predominant element in dermoid cysts. This tissue occasionally may produce the clinical picture of hyperthyroidism.

Uterine Tumors

Leiomyomas

Uterine leiomyomas are the most common benign tumor in the female pelvis. It is estimated that up to 50 percent of all women at some time in their life have one or more of these uterine tumors. The tumor is never seen before menarche, it grows during reproductive life, and it generally regresses following menopause. The tumors significantly complicate pregnancy by virtue of their rapid growth secondary to the response to pregnancy hormones.

Many leiomyomas are asymptomatic; when they do produce symptoms, they cause pain, abnormal uterine bleeding, infertility, ureteral obstruction, bladder distortion, and pressure symptoms secondary to the enlarged uterus.

Uterine leiomyomata are subject to a number of degenerative changes, including calcification, necrosis (occasionally with liquefaction), fatty degeneration, and occasionally, sarcomatous change. Malignant degeneration occurs in less than 1 percent of all tumors. Uterine myomas may be found in a number of locations within the uterus (Fig. 39-10). The most common location is intramural, but tumors frequently are found just below the peritoneum and occasionally as a pedunculated mass attached to the uterus. Other tumors grow into the endometrial cavity, where they are pedunculated on occasion, prolapsing through the cervix.

Treatment

Most symptomatic tumors can be managed expectantly. When symptoms indicate surgical treatment, consideration must be given to the age of the patient, the number of children she desires, the patient's age, and her reaction to possible loss of reproductive and menstrual function. Surgery should be fitted to the needs and desires of the patient. Therapeutic options might include myomectomy, total abdominal hysterectomy, or transvaginal hysterectomy.

Pedunculated myomas are the easiest to remove because the stalk of the tumor is simply ligated or coagulated and the tumor separated from the uterus. Pedunculated tumors in the endometrial cavity often can be removed with the operative hysteroscope. Similarly, tumors projecting from the external wall of the uterus into the peritoneal cavity can be removed laparoscopically by ligating the tumor pedicle and then morcellating the tumor intraperitoneally for removal or removing the tumor through a cul de sac incision.

Most myomectomies will be performed transabdominally. The most common indication for abdominal myomectomy is the presence of myomas that cause pain, bleeding, or infertility in a patient who continues to desire pregnancy. Before performing myomectomy, the patient must be evaluated completely and counseled about the risk of operation. Blood loss associated with myomectomy frequently exceeds that of hysterectomy. The patient should be advised of the possibility of hysterectomy in the event that myomectomy cannot be effectively performed. It is wise to set aside two or more units of autologous blood for possible operative use. The patient's tubal status should be evaluated with a hysterosalpingogram, and the patient's sexual partner should be evaluated with a semen analysis. In the presence of a very large tumor, chemoreduction of the tumor is currently being performed using a *GnRH-a*, such as nafarelin, given by nasal inhalation for a period of 3 months, or leuprolide acetate given intramuscularly every month for 2 months. These agents will cause the tumor to regress in size and in most cases provide for less blood loss at the time of operation.

Adenomyosis

Adenomyosis is a growth of endometrial tissue in the myometrium of the uterus and is sometimes referred to as *endometriosis of the uterine corpus*. The condition occurs primarily during reproductive years and leads to a thickening of the myometrial wall with subsequent uterine enlargement. Adenomyosis usually occurs in women who have had a number of pregnancies. Occasionally, patients with adenomyosis will complain of dysmenorrhea, and some present with increased uterine bleeding and heavy menstrual flow. However, a number of patients with adenomyosis in hysterectomy specimens have been asymptomatic. Therefore, the association of adenomyosis with heavy menstrual bleeding and dysmenorrhea is questionable.

Polyps

Endometrial polyps can occur at any time after puberty. A *polyp* is a local hyperplastic growth of endometrial tissue that usually causes postmenstrual or postmenopausal bleeding or staining, which is cured by polyp removal or curettage. The polyps are usually benign, but cases of adenocarcinoma of the endometrium arising in a polyp have been reported.

Cervical Lesions

Cervical polyps cause the same symptoms as endometrial polyps. Since they are often quite small and are visible at the external os, they often can be removed as an outpatient procedure followed by cauterization of the base of the polyp. Nabothian cysts are mucous inclusion cysts of the cervix. They are occasionally associated with chronic inflammation and can be removed easily with a cautery. They are harmless, usually asymptomatic, and generally do not require surgery.

During reproductive years, the portio of the cervix is covered primarily with glycogenated squamous epithelium, and columnar epithelium is normally found centrally near the external os in most women. This exposed columnar epithelium, termed *ectropion* or *eversion*, is usually bright red. Unless accompanied by inflammation and a purulent discharge (cervicitis), it requires no treatment. During adult life, the columnar epithelium is usually replaced by squamous metaplasia, and this physiologic process occurs in the transformation zone at the interface of squamous and columnar epithelium. After menopause, the squamous columnar junction is usually in the endocervical canal.

Vulvar Lesions

The term *leukoplakia* is often used for any white patch of the vulva; it is properly reserved for areas that show histologically atypical epithelial activity. These alterations may precede the development of malignant changes. In many instances, chronically irritated and itchy white areas of the vulva will show sclerosing atrophy of the skin (*lichen sclerosus*). Lichen sclerosus is a pruritic lesion that does not appear to be premalignant. Hyperplastic lesions termed *hypertrophic dystrophies* are found that may be benign (epithelial hyperplasia) or may show atypia, in which case dysplastic changes can be observed. The pruritic symptoms can be helped by topical application of corticosteroids. Testosterone also has been beneficial, especially for the atrophic changes of lichen sclerosus.

Noninvasive malignant change of the surface squamous epithelium of the vulva occurs in the same way as that described for the cervix. Carcinoma in situ of the vulva both histologically and clinically behaves like carcinoma in situ of the cervix. The changes are confined to the squamous elements of the vulva, and the condition is sometimes referred to as *Bowen's disease*. In certain instances, the apocrine glandular elements of the vulva are involved in association with an intensely pruritic area. Histologically, large, foamy Paget's cells are seen, similar to those noted in the breast, although invasive carcinoma occasionally can accompany Paget's cells. Bowen's disease and usually Paget's disease are considered part of the carcinoma in situ complex of the vulva; they are adequately treated by wide local surgical excision (simple vulvectomy). The laser also is used to treat these lesions locally.

MALIGNANT TUMORS

Ovarian Tumors

Ovarian Carcinoma

Ovarian carcinomas are divided histologically into epithelial, germ cell, and stromal malignancies. The majority of the 26,700 or more cases of ovarian cancer diagnosed annually in the United States are of the epithelial type. The median age at diagnosis for epithelial ovarian cancer is 61, and the overall 5- year survival rate for epithelial cancers is 37 percent. Approximately 14,800 women die of this disease in the United States annually.

Although the etiology of ovarian cancer is uncertain, approximately 5 percent of patients with epithelial tumors come from families where one or more first-degree relatives also have the disease. In such families, prophylactic oophorectomy may be considered at the completion of childbearing, especially if specific BRCA1 mutations are identified. Testing for BRCA1 mutations is now available at select centers in the United States. Primary peritoneal carcinomatosis has been reported in women who have undergone prophylactic surgery, however. Life-long screening with CA- 125 levels, pelvic examination, and vaginal ultrasonography of women from affected families bears consideration.

The FIGO (International Federation of Gynecology and Obstetrics) staging system for ovarian cancer is outlined in Table 39-2. Early lesions are largely asymptomatic, and advanced tumors may produce only nonspecific symptoms such as early satiety, abdominal distension, and vague gastrointestinal pains. Although an annual pelvic examination is valuable in detecting early ovarian cancer, efforts to establish other cost-effective screening programs using serum markers such as CA-125 and vaginal ultrasound examination are being developed. Vaginal ultrasound is a promising technology that is not presently cost-effective in screening programs. Currently, the majority of women with epithelial cancers have stage III tumors at the time of diagnosis. Widespread peritoneal dissemination, omental involvement, and ascites are the rule rather than the exception in these women.

Treatment

In general, therapy for epithelial ovarian cancer consists of surgical resection and appropriate staging followed by adjuvant radiation or chemotherapy. Women with low-grade early-stage (IA or IB) cancers who have undergone appropriate surgical staging may be treated with surgery without adjuvant therapy. If the lesion is bilateral (stage IB), abdominal hysterectomy and bilateral salpingo-oophorectomy are sufficient. It is in the limited group of patients with unilateral histologic grade 1 or 2 lesions that fertility can be preserved by performing adnexectomy and staging biopsies without removing the uterus or contralateral ovary and fallopian tube. In all other patients (stage IA, grade 3, and stage IB and above), appropriate initial surgery includes bilateral salpingo-oophorectomy, abdominal hysterectomy if the uterus has not been removed on a prior occasion, appropriate staging, and tumor resection.

Staging

Staging indicates surgical resection or biopsy of all potential areas of tumor spread. Thorough staging is imperative in determining appropriate treatment for patients with ovarian cancer. Among patients whose cancer is confined to one or both ovaries at the time of gross inspection, occult metastases can be identified by careful surgical staging in one-third. If staging is improperly performed and adjuvant therapy omitted in patients whose tumors are apparently confined to the ovary, 35 percent will suffer preventable relapse.

Epithelial ovarian cancers disseminate along peritoneal surfaces and by lymphatic channels. The first site of spread is the pelvic peritoneum. Later the abdominal peritoneal surfaces and diaphragms are involved. The omentum is a common site for metastases, as are both the paraaortic and pelvic lymph nodes. Because the abdominal cavity in its entirety is not accessible through a transverse pelvic incision, it is paramount that surgery for ovarian malignancies be performed through a full-length midline abdominal incision. After the peritoneal cavity is entered, the visceral and parietal surfaces are inspected for metastatic disease, and any suspicious areas are biopsied. If ascites is present, it should be aspirated and heparinized. Cytologic evaluation for metastatic cells or clusters is then performed. If no ascites is found, peritoneal washings with balanced salt solution or lactated Ringer's solution are obtained from the pelvis, abdominal gutters, and subdiaphragmatic areas and submitted for cytologic evaluation after centrifugation and fixation.

Patients with histologic grade 1 or grade 2 tumors confined to one or both ovaries (stage IA or IB) require no postoperative therapy if proper staging procedures have been performed. Five-year survival in this group of patients exceeds 90 percent.

Those patients who have stage I, grade 3 lesions, stage IC tumors (malignant peritoneal washings, rupture of tumor, surface excrescences, or ascites), or stage II cancers that are completely resected may be treated equally well with systemic chemotherapy, radiotherapy of the whole abdomen, or a single instillation of intraperitoneal radioactive chromic phosphate. Five-year survival approaches 75 percent in this group of patients.

Women with stage III and IV disease require systemic chemotherapy with cisplatin or carboplatin, generally in combination with paclitaxel or an alkylating agent. Survival at 5 years in such patients may exceed 20 percent, although this rate drops as low as 10 percent at 10 years.

Survival in advanced ovarian cancer is influenced by a number of factors, such as patient age, the histologic type and grade of the lesion, the presence or absence of ascites, and the type of chemotherapy employed. Of prime importance in advanced-stage disease, however, is the volume of tumor remaining after the initial surgical procedure. Many patients with stage III and IV ovarian cancer have diffuse peritoneal, retroperitoneal, diaphragmatic, and mesenteric metastases that resist complete surgical resection. It is often possible, however, to remove large amounts of peritoneal tumor by entering the retroperitoneal spaces and freeing the disease-laden surfaces from the underlying viscera.

It is widely accepted that patients in whom little or no residual disease remains after initial operation, on average, live longer than those in whom a great deal of tumor remains unresected. The terms *debulking* and *cytoreduction* have been introduced to indicate aggressive surgical removal of ovarian cancer. When disease remaining after surgical resection consists of nodules or plaques less than 1 to 2 cm in diameter, the surgical effort is termed *optimal*, and when a larger volume of residual disease remains, the surgical removal is termed *suboptimal*. Because of the survival advantage, every effort should be made to resect as much disease at the time of diagnostic laparotomy as is possible. Since many patients with advanced ovarian cancer are elderly and nutritionally depleted, surgical enthusiasm must be tempered by proper preoperative evaluation and support with appropriate central monitoring and hyperalimentation where indicated. Occasionally it is more prudent to obtain confirmation of the diagnosis, treat with systemic chemotherapy, and then perform definitive surgery when the tumor has diminished in size and the patient has been nutritionally resuscitated.

Resection of nodules involving the small or large bowel is warranted if the exercise results in complete removal of all observed disease. Such procedures are probably not indicated if tumor remains at other sites. After surgical extirpation of the tumor, patients with suboptimal ovarian cancers must be treated with chemotherapy. Approximately 80 percent of these tumors will respond to platinum-based combination therapy; 40 percent of all patients will experience a complete response, or complete resolution of tumor identified on physical examination or radiographic study.

Resection of Advanced Ovarian Cancer

When advanced ovarian carcinoma is discovered at the time of exploratory celiotomy, the first reaction is often one of resignation. There has been a tendency to perform a diagnostic biopsy and close the abdomen without further surgical intervention. In experienced hands, however, successful reduction of tumor volume to nodules 2 cm or less is possible in at least 50 percent of women with advanced ovarian cancer. If the primary surgeon is incapable of obtaining such results, the patient should be referred to one with sufficient expertise in this area. Survival following chemotherapy is inversely related to the volume of residual disease at the time of primary surgery.

Several techniques ensure adequate resection. First, most ovarian cancer is found on peritoneal surfaces and not invading viscera. A retroperitoneal approach thus facilitates mobilization of the involved mesothelium. The lateral aspects of the paracolic gutters may be incised and dissection carried medially to undermine tumor in these locations. The ovarian artery and vein should be identified at this point and securely ligated before division. Once the blood supply to the main body of the ovarian tumor is secured, the adnexa may be mobilized more easily. It is often useful to dissect the ureter from the underlying pelvic peritoneum and retract it laterally with a vessel loop or umbilical tape. This allows access to the lateral pelvic peritoneum. Tumor nodules on anterior and posterior cul de sac peritoneum may be resected by developing planes in the retroperitoneal spaces and isolating the disease from the underlying bladder, sigmoid colon, and ureters. Opening the pararectal and paravesical spaces facilitates this dissection and also allows access to the uterine vessels, which then may be clamped, ligated, and divided. When the hysterectomy and adnexectomy are complete, the omentum may be resected.

Disease on the right diaphragm may be resected by transecting the falciform ligament and retracting the liver inferiorly. If it serves to remove all remaining tumor, splenectomy may be performed. Resection of small and large bowel may be performed if the operation removes all residual disease. In recent years, use of the ultrasound aspirator has resulted in an increased ability to completely remove tumor, including that which is implanted on the serosal surfaces and mesentery of the bowel. With diligence it is often possible to remove all appreciable disease with this instrument. Another useful addition to the surgical armamentarium is the argon beam coagulator, which carries a stream of electrons into the tissue via an argon jet.

“Second Look” Operations

“Second Look” Laparotomy

Ovarian cancer often defies diagnosis because it does not produce symptoms and is detectable neither radiographically nor serologically even in relatively advanced stages. The assessment of ovarian cancer during and after therapy is similarly difficult. Although CT or MRI may identify masses as small as 2 to 3 cm in diameter, neither technique can reliably detect smaller masses, much less the miliary spread so often identified in advanced ovarian cancer.

CA-125 is more sensitive than radiographic or magnetic scanning but is also associated with a number of false-positive results and may not be elevated in patients with mucinous tumors. In addition, approximately half of patients with advanced ovarian cancer whose CA-125 levels normalize during chemotherapy harbor viable and clinically undetectable disease. Radiolabeled monoclonal antibodies raised against epithelial tumor surface antigens may be more sensitive than traditional methods but remain to be proven effective.

The practice of performing exploratory surgery following chemotherapy originated during a time when alkylating agents were used almost exclusively. Because acute nonmyelocytic leukemia is associated with prolonged administration of such agents, a “second look” operation was performed at an interval of 12 to 24 months following primary surgery so that treatment could be stopped in women with no disease. Presently, the duration of postoperative combination chemotherapy is often only 5 to 6 months, and the risk of leukemia is very low. In approximately 20 to 30 percent of patients who receive such treatment, no cancer will be identified at the time of a second operation. These patients have an excellent long-term prognosis. In women who have persistent microscopic disease, the prognosis is also favorable, and in those with persistent gross tumor, the prognosis is relatively poor.

“Second look” surgery is currently used primarily as a research tool. New treatment regimens can be evaluated quickly by performing a “second look” operation, since the findings at such an operation reflect the ultimate clinical outcome and hence the value of the treatment regimen. “Second look” surgery is also valuable in determining when therapy can be discontinued and when further treatment is indicated. If cancer is still present, secondary surgical resection has a beneficial effect on prognosis.

“Second look” laparotomy is performed through a midline abdominal incision. Peritoneal washings are obtained from both abdominal gutters, the diaphragms, and the pelvis. Since persistent cancer is most likely to be identified in sites where there was tumor at the conclusion of the primary operation, these areas are explored first. Any suspicious nodules, thickened peritoneal areas, or adhesions should be biopsied carefully and submitted for histologic evaluation. If no visible tumor is present, a formal staging operation should be performed. Liberal peritoneal biopsies from the above-mentioned sites are obtained, and then pelvic and paraaortic lymph nodes are sampled. Any residual omentum should be excised because it may harbor occult tumor. At the conclusion of “second look” surgery, some consideration should be given to subsequent therapy.

If tumor is identified and can be resected, a Tenckhoff catheter may be placed, through which intraperitoneal chemo- or immunotherapy may be given. If no visible disease is present, a Silastic drain can be inserted for the postoperative administration of radioactive chromic phosphate. Although the morbidity of “second look” surgery is very low, there is no place for this type of operation in patients who wish or who can physiologically tolerate no further treatment at the conclusion of primary combination chemotherapy. Nor is there any reason to reexplore such a patient without the ability or intention to perform a thorough, deliberate staging procedure that can guide subsequent therapy.

Other Secondary Operations

Surgical resection of tumor after chemotherapy or at the time of relapse is termed *secondary cytoreduction*. In the occasional patient who undergoes diagnostic biopsy only before the administration of chemotherapy, early reexploration may be termed *interval cytoreduction*. There is evidence that the surgical removal of extensive tumor is facilitated by the administration of one or two courses of combination chemotherapy. In patients with a massive tumor burden, this approach may not only be safer but also might result in a more successful tumor resection before the completion of chemotherapy. It also promotes the early administration of intraperitoneal chemotherapy.

The importance of secondary cytoreduction is not clearly established. In patients with relapsing ovarian cancer, the prognosis depends in part on the extent of tumor and in part on the type of response to previous therapy. Also important is the interval between primary therapy and relapse. In those who completely responded to platinum combination treatment and who have a disease-free period exceeding 2 years, resumption of platinum-based chemotherapy is very effective. Paclitaxel therapy may be effective in similar situations. It is in such patients that surgical removal of the recurrent tumor is likely to be the most beneficial.

Palliative Surgery

In most cases of advanced ovarian cancer, death is associated with bowel dysfunction or frank obstruction. Although invasion of the small bowel and colon is unusual, growth of the tumor adjacent to the bowel leads to mesenteric compromise and dysfunction usually heralded by distention, nausea, and vomiting. When bowel obstruction occurs early on in the clinical course of ovarian cancer, and particularly if it occurs before the administration of chemotherapy, surgical intervention is warranted and should be aggressive. Resection or bypass of the involved small bowel is indicated; colonic resection also may be indicated. It is important to perform adequate radiographic studies preoperatively so that obstructed small bowel is not decompressed into a compromised colon.

When bowel obstruction occurs after chemotherapy, the prognosis is unfavorable. Women who develop such difficulties have a limited survival following surgical correction. Surgery is often difficult to perform because of extensive tumor. Laparotomy may be complicated by enteric injury or fistula. Often the best approach in these patients is the use of a percutaneous or endoscopically positioned gastrostomy tube and intravenous fluids or conservative nutritional support. Such a procedure may limit the length of hospitalization and

allow the patient to remain in a supportive home environment for a greater period of time.

Laparoscopy in Ovarian Cancer

At present, our ability to resect large ovarian cancers successfully using laparoscopic equipment is limited. In the past, efforts to perform “second look” procedures through the laparoscope were ineffective when compared with laparotomy. However, with the advent of new equipment and techniques, the role of laparoscopy in the staging and treatment of ovarian malignancies is expanding. Several investigators have developed successful methods of performing both pelvic and paraaortic lymphadenectomies using endoscopic equipment. In addition, ultrasonographic and serologic criteria are evolving that will allow the surgeon to more successfully distinguish between benign and malignant neoplasms of the ovary.

Tumors of Low Malignant Potential

These are epithelial tumors of malignant potential intermediate between benign lesions and frank malignancies. Histologically, most are of the serous type. They are distinguished from invasive cancers microscopically by lack of stromal invasion. The median age of diagnosis is approximately 10 years younger than that of patients with epithelial cancers. The vast majority occur in stage I and have a favorable prognosis. Surgery should include abdominal hysterectomy and bilateral salpingo-oophorectomy unless fertility is to be preserved in patients with unilateral lesions. These patients may undergo unilateral salpingo-oophorectomy. Ovarian cystectomy or nonextirpative resections commonly result in recurrences.

Patients with stage III and IV lesions have 5-year survival rates that approach 85 percent after complete surgical resection. There is little evidence that either chemotherapy or radiotherapy administered after surgery improves survival; on the other hand, deaths from chemotherapy- induced leukemia are not uncommon.

Germ Cell Tumors

These tumors occur in women in the first three decades of life and typically grow rapidly, producing symptoms of distention and abdominal fullness. Torsion may occur, producing an acute abdomen. Most are unilateral, and all have a tendency to spread to the paraaortic lymph nodes as well as throughout the peritoneal cavity. Although they are similar in many ways to testicular cancer in the male, there are some differences.

Dysgerminoma, the female equivalent of testicular seminoma, is composed of pure, undifferentiated germ cells. It is bilateral in 10 percent of patients and is occasionally associated with elevated levels of *hCG* or *lactate dehydrogenase (LDH)* . It is the most common ovarian malignancy diagnosed during pregnancy. Patients bearing dysgerminomas should undergo appropriate staging at the time of the primary resection but need not undergo hysterectomy (if fertility is to be preserved) or removal of the opposite ovary if it is normal in appearance. Secondary operations solely for staging purposes are unwarranted. Adjuvant therapy is unnecessary unless there is evidence of extraovarian spread. Either radiotherapy encompassing the whole abdomen or systemic chemotherapy can be given to patients with metastases. This tumor is exquisitely sensitive to either type of treatment, and the cure rate exceeds 90 percent even in patients with metastases. Chemotherapy has the advantage of preserving ovarian function, whereas radiotherapy results in ovarian failure.

The other germ cell tumors in order of frequency are immature teratoma, endodermal sinus, or “yolk sac,” tumor, mixed tumors, embryonal carcinomas, and choriocarcinomas. The first may be associated with elevated levels of alpha-fetoprotein (AFP). Elevated AFP levels are found in all patients with endodermal sinus tumors and mixed tumors that contain this component. Embryonal carcinomas are associated with abnormal levels of both AFP and *hCG*, and *choriocarcinomas secrete hCG* .

These tumors are invariably unilateral but may spread by peritoneal, hematogenous, or lymphatic routes. Surgical therapy is much the same as that described for ovarian dysgerminomas. Except for those with completely resected stage I, grade 1 immature teratomas and those with stage I dysgerminoma, all patients with germ cell tumors require systemic chemotherapy. Three courses of a platinum and etoposide-containing combination suffice in those patients whose tumors are completely resected. Cure rates in these patients approach 90 percent. In women with incompletely resected nondysgerminomatous germ cell tumors, cure may still be expected in over 50 percent, but prolonged chemotherapy may be necessary. These tumors are *not* sensitive to radiotherapy.

Carcinoma of the Cervix

Carcinoma of the cervix accounts for about 16,000 cases and 5000 deaths annually in the United States. Risk factors include multiple sexual partners, early age at first intercourse, and early first pregnancy. DNA related to that found in the human papillomavirus has been identified in cervical dysplasia and carcinoma in situ, both precursor lesions, as well as in invasive cancers and lymph node metastases. Other etiologic agents considered in this disease are herpes simplex type II and cigarette smoking.

In no cancer has widespread screening had as profound an impact on mortality as it has in carcinoma of the cervix. Georges Papanicolaou devised the cytologic smear that bears his name in 1943. Since that time, screening programs have dramatically reduced the rate of invasive cervical cancer in countries where this test is widely available. Use of the Pap smear has shifted the frequency of cervical abnormalities toward the premalignant intraepithelial diseases, dysplasia, and carcinoma in situ. Although there are histologic grades of dysplasia leading to carcinoma in situ, all intraepithelial lesions are noninvasive and can be treated successfully using conservative methods.

Eighty percent of all cervical cancers are squamous cell in type and arise from the squamocolumnar junction of the cervix. This epithelial transition zone is found on the face of the cervix or ectocervix in adolescence and, through a process of squamous metaplasia, gradually moves into the endocervical canal as menopause is passed. Dysplasia represents a disordered metaplasia and gives rise to epithelial cells that contain increased mitotic rates and nuclear atypia and that lack appropriate maturation within the epithelium. Identification and eradication of intraepithelial lesions before invasion can occur are the goals of cervical cancer screening.

The remainder of cervical malignancies arise in the endocervical canal and are either adenocarcinomas or adenosquamous carcinomas. Although adenocarcinomas are very similar in their clinical behavior to squamous cancers, there is some evidence that adenosquamous cancers are more aggressive. Other rare histologic varieties associated with poor prognosis are neuroendocrine small cell carcinomas and clear cell cancers. The latter are frequently associated with maternal exposure to diethylstilbestrol.

Staging

Cervical cancers spread predominantly by lymphatic channels. The first lymph nodes involved are those in the tissues immediately lateral to the cervix. This region is referred to as the *paracervical* or *parametrial area* . The next lymph nodes to be involved, in order, are those in the obturator fossa, the internal and external iliac chain, the common iliac chain, and the paraaortic lymph nodes. Direct vaginal extension may occur. The lymph nodes in the presacral area may be involved in early stage lesions, and the supraclavicular lymph nodes are the most common site of distant nodal metastases.

FIGO staging for cervical cancer is based on clinical examination, intravenous pyelography, and chest radiography. CT or MRI findings do not affect the clinical stage. The FIGO staging system is illustrated in Table 39-3. Note that the presence of hydronephrosis connotes stage IIIB even if there is no clinical evidence of extracervical spread. Except for selected patients with stage IVA lesions and those with distant metastases, all patients with stage IIB cancer and above are treated primarily with radiotherapy in the United States.

Treatment

Intraepithelial or Preinvasive Disease

Abnormal Pap smears must be evaluated by colposcopy and biopsy. Colposcopy is the examination of the cervix with a low-power (10 to 50×) microscope after application of dilute acetic acid to the cervix. The acid solution is mucolytic and serves to desiccate the epithelium, a process that brings out subtle epithelial patterns referred to as white epithelium, punctation, mosaicism, and abnormal vasculature. Abnormal areas must undergo mechanical biopsy or wide excision with a wire loop electrode and are examined histologically. If loop excision is not performed, the endocervical canal should be curetted to exclude epithelial abnormalities in this area, which is difficult to visualize colposcopically.

Once the diagnosis of an intraepithelial process is made and stromal invasion excluded, local therapy can be performed. If there are abnormal cells on the endocervical curettage specimen, a diagnostic cone biopsy is indicated to exclude the possibility of an invasive or microinvasive lesion in the endocervical canal.

Cervical intraepithelial neoplasia is treated in a number of ways. In general, the larger the lesion and the higher the grade of dysplasia, the greater is the failure rate. Similarly, more aggressive therapy yields lower failure rates at increased risk of complications. The most definitive treatment for cervical intraepithelial neoplasia is vaginal or abdominal hysterectomy. This operation is associated with a rate of subsequent dysplasia at the vaginal apex of 1 to 2 percent. This major operation is usually reserved, however, for patients with extensive or high-grade lesions, those with recurrent disease after conservative treatment, those in whom adequate follow-up is unlikely, and those with other indications for hysterectomy such as prolapse, abnormal uterine bleeding, pain, or a pelvic mass. Cervical cone biopsy is curative in most cases of cervical intraepithelial neoplasia. In patients in whom the surgical margins of the cone specimen are uninvolved, the risk of recurrence is less than 5 percent. If the surgical margins are involved, half of such patients will develop recurrent disease. This is an outpatient procedure and associated with few serious risks. It may, however, require general anesthesia.

More conservative methods of treating cervical intraepithelial neoplasia include wire loop excision, laser vaporization, and cryosurgery. Loop excision can be done under local anesthesia (paracervical block) in the outpatient setting. The advantage of loop excision is that it removes the diseased area and provides a diagnostic biopsy specimen. The main disadvantage is the relatively large amount of cervical stroma that is taken with the involved epithelium. In cases of cervical intraepithelial neoplasia confined to the ectocervix, such deep excision is probably unnecessary.

Laser vaporization is usually performed with a carbon dioxide laser, but other laser instruments may be used. The ectocervical transformation zone is ablated to a depth of about 7 mm to ensure the removal of endocervical glandular epithelium. This is a convenient outpatient procedure that results in a clearly visible squamocolumnar junction at the site of treatment. Risks of bleeding and infection are small. Cryotherapy is an inexpensive outpatient procedure that produces a frostbite injury to the ectocervical epithelium. When the cervix reepithelializes, the dysplasia generally does not recur. This is a simple technique that should not be applied to patients with endocervical lesions. The main disadvantage of cryotherapy is obliteration of the squamocolumnar junction, making subsequent colposcopic examination somewhat difficult. In patients with very localized mild dysplasias or low-grade cervical intraepithelial neoplasia, local excision or electrocautery may be sufficient to eradicate the disease.

Microinvasive Cervical Cancer

FIGO (see Table 39-3) subdivides microinvasive cancers into those with “early” invasion (stage IA₁) and those in which the tumor measurements are less than 5 mm in thickness and 7 mm in lateral extent (stage IA₂). This aspect of the FIGO staging system for cervical cancer fails to distinguish adequately between stages IA₂ and IB, however, because both may have occult lymph node metastases requiring regional therapy.

Many prefer the original system of the Society of Gynecologic Oncologists, in which stage IA (microinvasive) tumors may invade to no greater than 3 mm and must lack capability of lymphatic space invasion. Stage IB includes all other cancers clinically confined to the cervix, even if they cannot be visualized on examination. The advantage of this system is that

it clearly divides stage I into two treatment groups. Few patients with stage IA cervical cancer have metastases to the lymph nodes. Simple, or extrafascial, hysterectomy without lymphadenectomy is therefore adequate therapy. Five-year survival rates approach 100 percent in these patients. In exceptional patients, cervical cone biopsy or electrosurgical excision may be sufficient treatment, provided close surveillance is possible.

Early Invasive Cervical Cancer (Stage IB and IIA)

Stage IB and IIA tumors are associated with a risk of pelvic lymph node metastases of 10 to 15 percent and a risk of spread to the paraaortic nodes of about 5 percent. Treatment must include the regional lymph nodes in these patients. Radical hysterectomy with pelvic lymphadenectomy or definitive radiotherapy is effective treatment in this stage cancer. Prognosis with either modality depends on the size of the primary lesion, the presence or absence of lymph-vascular space involvement, spread to the regional lymph nodes, and status of the surgical margins.

Women with stage IB₂ cervical cancers (exceeding 4 cm in diameter), especially those endocervical primaries that distend the cervix circumferentially, may require a combination of radiotherapy and surgery. These large endocervical tumors are referred to as “barrel” lesions and are refractory to surgery or radiotherapy alone. Isodose curves from cesium sources may not encompass the entire tumor (Fig. [39-11](#)). Cure rates with either treatment may be as low as 50 percent.

One current approach to these tumors is the administration of pelvic radiotherapy followed by a cesium implant and subsequent simple hysterectomy. This technique may reduce the number of patients who have persistent invasive cancer in the cervix after radiotherapy and consequently improve survival.

Stage IB₁ lesions and early stage IIA cancers may be treated successfully with radical hysterectomy and pelvic lymphadenectomy. This operation was pioneered by John Clark at Johns Hopkins Hospital in 1895. Radical surgery was transiently eclipsed by the first use of radium in the treatment of cervical cancer by Sjögren and Stenbeck in 1899 and subsequent establishment of the first radium hospital in Stockholm, Sweden, in 1910.

Radical surgery reemerged in the treatment of early carcinoma of the cervix with the advent of the Pap smear and increased diagnosis of early-stage tumors in young women. Because early cervical cancer so rarely spreads to the ovaries, radical hysterectomy need not include oophorectomy. Ovarian preservation is one of the strongest arguments for the use of surgery over radiotherapy, since the latter inevitably results in the premature loss of ovarian function.

Locally Advanced Carcinoma of the Cervix (Stages IIB to IVA)

These cancers are treated primarily with radiotherapy. Treatment consists of a combination of external therapy to the pelvis (teletherapy) from a high- energy source such as a linear accelerator and a local dose delivered to the cervix and parametrial tissue (brachytherapy) using a cesium applicator such as a Fletcher-Suite tandem and ovoids (Fig. [39-12](#)). Combination therapy is essential because doses adequate to control cervical tumors exceeding about 1 cm in diameter cannot be given using teletherapy alone. Bladder and rectal tolerances are approximately 6000 rads; higher doses can only be attained by combination therapy.

Cure rates for stage IIB cervical cancers approach 65 percent, and those for stage IIIB, 35 percent. Because the risk of pelvic sidewall lymph node involvement increases with advancing stage, the dose of radiotherapy to this area is advanced with increasing stage. When paraaortic metastases are present in either stage, survival is significantly impaired. Survival for patients with stage IIB carcinoma of the cervix and paraaortic metastases is poorer than that for those with stage IIIB disease and negative paraaortic lymph nodes. Gross paraaortic lymph node metastases may be detected by CT, MRI, or lymphangiography. Microscopic nodal metastases are best detected by retroperitoneal common iliac and paraaortic lymphadenectomy, a relatively simple procedure performed through a “hockey stick” or paramedian incision (Fig. [39-13](#)). The fascial layers are divided, sparing the peritoneum, which is reflected medially to expose the lymph node–bearing areas overlying the major blood vessels.

The finding of metastases in the common iliac or paraaortic chain indicates the need for extended-field radiotherapy encompassing these areas in addition to the pelvis. Even with such therapy, 5-year survival rates are low, seldom exceeding 20 percent. Many consider the presence of paraaortic lymph node metastases to be an indicator of systemic disease, although supraclavicular metastases are present in fewer than 25 percent of such patients.

Recurrent Cervical Cancer

As a rule, patients who develop local recurrences after preliminary surgical therapy are treated most effectively with external- and internal-beam radiotherapy. Although those with lymph node failures may not be curable in this setting, those with vaginal recurrences often can be saved with such an approach. Patients who suffer recurrences at sites distant from the pelvis may be treated with palliative local radiotherapy or chemotherapy with limited success.

Women who develop recurrent cancer following primary radiotherapy are generally not candidates for curative therapy. If, however, the recurrent lesion is small, the interval to failure is a year or more, and the lesion is unaccompanied by symptoms such as back or leg pain or edema, surgical resection may be possible. Since radiotherapy results in fibrosis of the connective tissues surrounding the cervix, radical hysterectomy is impractical. The risk of vesicovaginal or rectovaginal fistulas is excessive. In addition, surgical margins may be compromised by limited resection in such a situation.

Most gynecologic oncologists prefer to perform pelvic exenteration in such circumstances. Often, an anterior exenteration with en bloc removal of the bladder, cervix, uterus, and upper vagina is feasible. These operations require urinary diversion. Because of radiation exposure, however, an ileal conduit may be associated with urinary leakage from ureteroileal anastomoses. The preferred method of diversion in these patients is the creation of a sigmoid urostomy or transverse colon conduit. Other surgical options include a Koch pouch or the Indiana reservoir (Fig. [39-14](#)), both of which provide a means of urinary continence without an external appliance.

In the case of extensive local recurrences, sigmoid resection may be required in addition to removal of the bladder. A total pelvic exenteration is performed. The sigmoid colon may be brought to the skin as a colostomy or reanastomosed to the rectal stump. Pelvic exenterations may be subclassified as supralelevator or infralevator depending on whether this muscular diaphragm is broached (Fig. [39-15](#)). Supralelevator exenterations are generally associated with less operative morbidity. An infralevator exenteration is required if tumor involves the middle or lower third of the vagina or the vulva. Vaginal reconstruction in these extensive procedures with gracilis myocutaneous flaps is highly satisfactory (Figs. [39-16](#) and [39-17](#)).

In general, about half the patients thought to be candidates for pelvic exenteration are found to have intraperitoneal spread or nodal metastases at the time of exploratory laparotomy and, in most centers, do not undergo resection. Of the remaining patients in whom surgery is possible, 30 to 50 percent will develop a second, nearly always fatal, recurrence after surgery. This complex operation should thus be undertaken only in carefully selected patients.

Endometrial Cancer

Endometrial cancer is the most common female genital malignancy, accounting for 34,000 cases annually in the United States. It is a highly treatable cancer, and only 6000 deaths are reported each year. Recently, there has been a disturbing increase in the proportion of fatalities reported in this country.

Risk factors for endometrial cancer include obesity, diabetes mellitus, hypertension, low parity, early menarche, and late menopause. Excessive exposure to estrogens is implicated in the genesis of endometrial cancer and its precursor, endometrial hyperplasia. Women who take estrogens in the menopausal years are known to have a sixfold increase in the risk of endometrial cancer if progestational agents are not taken as well. There is also an increase in the incidence of endometrial lesions in women with a history of chronic anovulation (Stein-Leventhal syndrome) and in those with estrogen-producing ovarian stromal neoplasms such as granulosa-cell tumors.

Endometrial hyperplasia may be divided into simple and complex, depending on the microscopic architecture, and into those with or without atypia. These hyperplasias are thought to be estrogen-dependent. Atypical complex hyperplasias are most likely to give rise to frank adenocarcinomas. They occur in women at an average age that is 5 to 10 years younger than those with frank carcinomas. Simple hysterectomy is the preferred method of treatment for the hyperplasias. In women with underlying health problems that preclude surgical therapy, therapy with progestational agents such as megestrol or medroxyprogesterone acetate may be used with success. Careful monitoring with endometrial biopsy or curettage is required in these patients, however.

Both endometrial hyperplasia and carcinoma are often heralded by abnormal perimenopausal or postmenopausal uterine bleeding. This symptom accounts for the early detection and relative curability of these neoplasms.

Treatment

Endometrial cancer is staged according to the FIGO criteria detailed in Table [39-4](#). Many patients have stage I disease and can be managed successfully with abdominal hysterectomy and bilateral salpingo-oophorectomy. Adjuvant radiotherapy may be required, primarily to reduce the risk of vaginal recurrence. This can be given preoperatively with external therapy or a Fletcher-Suite implant or intrauterine packing (Heyman's or Simon's capsules). Some clinicians prefer to deliver radiotherapy postoperatively after the uterus has been evaluated thoroughly. Either external-beam therapy or vaginal cesium may be used.

Pelvic lymph node metastases occur in about 12 percent of patients with endometrial cancer apparently confined to the uterus. Lymph node metastases have a significant negative impact on survival. Risk factors associated with lymph node spread include high histologic grade (grade 2 or 3), low levels of progesterone receptor, deep myometrial or lymphatic channel invasion, spread to the adnexa, endocervical extension, and unusual histologic variants such as papillary serous or clear cell carcinomas. It is unnecessary to perform lymph node sampling in patients with grade 1 adenocarcinomas confined to the endometrium or inner one-third of myometrium. Other patients should have pelvic lymph nodes sampled at the time of hysterectomy. Therapeutic lymphadenectomy is not advocated; sampling of the external and internal iliac and obturator areas is sufficient for patients with endometrial cancer. Those with a high likelihood of spread to pelvic lymph nodes (grade 3, outer one-third myometrial or uterine serosal involvement, and those with high-risk histologic subtypes) should undergo sampling of the common iliac and paraaortic lymph nodes, since these areas lie outside the usual fields of pelvic radiotherapy. Patients with papillary serous tumors may develop metastases in the abdominal cavity or omentum much as those with ovarian epithelial tumors.

Another important element of staging endometrial cancer is the evaluation of peritoneal lavage fluid for the presence of malignant cells. About 12 percent of patients are found to have malignant peritoneal cytology. Half have other evidence of extrauterine spread of the disease, but the remainder have no other associated risk factors. Malignant peritoneal cytology increases the risk for intraabdominal failure and treatment for this finding deserves consideration. If external-beam radiotherapy is not used, intraperitoneal radioactive chromic phosphate may be of benefit.

Vaginal hysterectomy is occasionally useful in patients with early endometrial cancer when lymph node metastases are thought to be unlikely. This operation is particularly well suited for massively obese parous patients in whom an abdominal incision would be prohibitively difficult.

It is unnecessary to perform radical surgery in women with endometrial cancer even if there is spread to the cervix (stage II). Although lymphatic spread is important, these cancers also may be disseminated by hematogenous or peritoneal routes. Radical surgery has never been shown to improve survival in comparison with simple hysterectomy and adjuvant radiotherapy. Simple, extrafascial, or complete abdominal hysterectomy is demonstrated in Fig. 39-18 and contrasted with radical hysterectomy. It is critical to remove the ovaries in women undergoing surgery for endometrial cancer because 5 percent harbor occult metastases. Additionally, the source of estrogen secretion in premenopausal women with endometrial cancer should be removed.

In patients with large stage IIB and III lesions, consideration is generally given to preoperative pelvic radiotherapy, since surgery may be otherwise difficult or impossible. These tumors should receive appropriate surgical staging or thorough radiographic evaluation if primary radiotherapy is used.

Pelvic exenteration is rarely necessary in the treatment of patients with endometrial cancer unless it occurs following full irradiation for a preexisting cervical cancer. Such cases are rare, of course.

Radiotherapy alone may be the treatment of choice in patients at excessive risk for operative intervention. Radiotherapy alone produces results inferior to those of surgery or surgery and adjuvant radiotherapy, however, so patients treated without hysterectomy should be selected carefully. Advanced or recurrent endometrial cancer is responsive to progestin or tamoxifen therapy in 30 percent of unselected patients. Lesions that are well differentiated contain higher levels of progesterone receptor and respond more frequently. Only 10 percent of poorly differentiated cancers respond to hormonal treatment. Local radiotherapy or chemotherapy with paclitaxel, doxorubicin, platinum compounds, or combinations may be of benefit in some cases as well.

Vulvar Cancer

Vulvar cancer accounts for about 5 percent of all gynecologic cancers. Although uncommon histologic types such as malignant melanoma and adenocarcinoma of the Bartholin's gland occur, over 90 percent of vulvar malignancies are squamous carcinomas.

Epidemiologic risk factors include older age, smoking, previous intraepithelial or invasive squamous cancer of the cervix or vagina, chronic vulvar dystrophy (often associated with diabetes mellitus), and immunocompromise (organ transplant recipients, systemic lupus erythematosus). Human papillomavirus-like DNA has been identified in both preinvasive and invasive squamous carcinomas of the vulva. Although the etiology of this cancer is not well understood, it is likely that the human papillomavirus plays an important role.

Spread of squamous carcinoma of the vulva is primarily via the lymphatics of the vulva. Lesions arising in the anterior aspect of the vulva drain preferentially to the inguinal lymph nodes and posterior lesions may drain directly to the lymph nodes of the pelvis.

Stanley Way of Great Britain identified five main groups of lymphatic drainage of carcinoma of the vulva (Fig. 39-19): (1) the superficial inguinal lymph nodes, which lie in the subcutaneous tissue overlying the inguinal ligament, (2) the deep inguinal lymph nodes, which lie along the course of the round ligament in the inguinal canal, (3) the superficial femoral lymph nodes, grouped around the saphenous vein just superficial to the fossa ovalis, with efferents to the deep femoral lymph nodes, (4) the deep femoral lymph nodes, including the most cephalad lymph node of Cloquet or Rosenmuller, and (5) the external iliac lymph nodes.

Since the lymph node of Cloquet receives efferents from the inguinal region and the vulva and drains into the medial portion of the external iliac chain, it is an important sentinel in the route of spread of vulvar lesions. There are also direct lymphatic connections between the clitoris and Cloquet's node.

The 1988 FIGO staging system for vulvar cancer (Table 39-5) is currently accepted. This system requires surgical evaluation of the inguinal lymph nodes and provides a schema in which prognosis and therapy are closely linked with stage.

Treatment

Historically, the single-stage en bloc "extended" radical vulvectomy championed in Great Britain by Way and in the United States by Friedrich Taussig was used to treat all vulvar neoplasms (Fig. 39-20). In this operation, wide margins of skin and subcutaneous tissue around the primary tumor are removed together with underlying lymphatic structures in the groins and the labia majora and minora and clitoris in the vulva. Also removed are the proximal saphenous vein and its tributaries, the superficial circumflex iliac, superficial external pudendal, and superficial inferior epigastric veins.

The deep inguinal lymph nodes are removed by opening the external oblique fascia overlying the inguinal canal; most of the round ligament is removed at the same time.

Pelvic lymphadenectomy is easily performed by opening the transversalis fascia below the inguinal ligament and exposing the external iliac vessels in the retroperitoneal space medial to the psoas muscle. Pelvic lymphadenectomy is probably not indicated in vulvar cancer except in those patients found to have grossly enlarged pelvic lymph nodes on preoperative CT or MRI. Patients with inguinal node metastases are best treated with inguinal and pelvic radiotherapy following resection of the inguinal lymph nodes. In the case of large vulvar primaries or suspicious inguinal lymph nodes, this approach yields better survival rates than those obtained when pelvic lymphadenectomy alone is performed.

Because extended radical vulvectomy is associated with long hospital stays and significant morbidity from wound breakdown and infectious complications, there has been a long-standing interest in more conservative surgery for early vulvar cancer. The first efforts to this end were made in the 1960s. Several investigators introduced the concept of radical vulvectomy and inguinal lymphadenectomy through separate incisions. This approach not only reduces hospital time but also results in fewer major wound complications. This approach, illustrated in Fig. 39-21, has been widely embraced by gynecologic oncologists. Since inguinal node metastases are the result of an embolic process rather than infiltration or direct extension, the approach is rational. Early concerns regarding recurrence in the skin bridge between the vulvar and groin incisions have been largely allayed by experience with this approach. Recurrence in the skin bridge is usually associated with preexisting large inguinal metastases.

Another area of progress in the surgical management of vulvar carcinoma has been the use of conservative surgery for early lesions of the vulva. Although specific criteria differ slightly, most investigators recognize that squamous cancers of the vulva less than 2 cm in diameter and no more than 1 mm thick, and that are of histologic grade 1 or 2, are associated with a very small risk of inguinal metastases. Such lesions are adequately treated with deep, wide excision, provided skin margins of 1 cm are obtained and the dissection is carried to the level of the superficial transverse perineal muscles. Inguinal lymphadenectomy can be omitted in such patients.

In patients with intermediate lesions located on the labium minus or majus that do not cross the midline or involve midline structures such as the clitoris, perineal body, or perianal area, modified hemivulvectomy and ipsilateral inguinal lymphadenectomy have been used successfully. This approach should be considered if the primary lesion is less than 2 cm in diameter and 5 mm or less in thickness. Lymph node metastases are uncommon in this group of patients and maybe evaluated by frozen section at the time of surgery. The lymph nodes superficial to Camper's fascia (Fig. 39-22) in the groin are removed through a conservative inguinal incision (Fig. 39-23). If these "sentinel" lymph nodes are free of tumor, the risk of involvement of the deep inguinal or femoral lymph nodes is small enough that further dissection may be avoided. The outlines of the modified radical hemivulvectomy are depicted in Fig. 39-24. This excision site may be closed primarily with good results.

Another controversial area in the management of squamous carcinomas of the vulva is that of the patient with locally advanced disease. When extensive vulvar cancer involves more than the distal urethra, the vagina or rectovaginal septum, or the anal musculature, ultraradical surgery may be required. Anterior, posterior, or total pelvic exenteration may be necessary to resect such lesions successfully. The presence of fixed, matted, or ulcerating inguinal lymph nodes presents another problem that may require extensive surgical excision. Following such extirpative procedures, reconstruction of the vulva and groins is accomplished using myocutaneous flaps based on the gracilis, sartorius, or tensor fasciae latae muscles. Approximately 50 percent of patients are cured by such surgical procedures.

In recent years, such locally advanced lesions of the vulva also have been treated successfully with external-beam radiotherapy combined with radiosensitizing drugs such as cisplatin and 5-fluorouracil. At the completion of combination therapy, the areas of involvement are excised widely or biopsied. This approach is associated with results as good as or better than those achieved with ultraradical surgery and generally results in less morbidity. The need for urinary and fecal diversion is also obviated.

Uncommon Vulvar Tumors Melanoma

Traditional surgical therapy for malignant melanoma of the vulva has included en bloc radical vulvectomy and inguinofemoral lymphadenectomy. It is now known that lesions less than 1 mm thick or Clark level II lesions may be treated conservatively with wide local excision. The value of inguinofemoral lymphadenectomy is controversial in lesions of greater depth, although primary surgical cure is occasionally achieved in patients with microscopic nodal metastases. Melanomas of the urethra or vagina are usually diagnosed in advanced stages and may require pelvic exenteration for successful management.

Intraepithelial Disease

Intraepithelial disease (Bowen's disease, bowenoid papulosis, vulvar intraepithelial neoplasia, carcinoma in situ) may be treated successfully by removing the involved epithelium. This is characteristically a raised, velvety lesion with sharply demarcated borders that may contain gray, brown, or red pigmentation. Removal is accomplished by simple vulvectomy, where the plane of dissection is limited to the epithelium, or by wide excision. In the case of diffuse intraepithelial disease, a so-called skinning vulvectomy and split-thickness skin graft may be required. This approach is associated with prolonged hospital stays, however, and should be reserved for exceptional cases. Also effective in the treatment of intraepithelial disease are the carbon dioxide laser and the electrosurgical loop.

Paget's disease is an unusual epithelial or invasive process characterized by the presence of distinct "Paget cells" in the involved epithelium. Grossly, the lesion is confluent, raised, red, and waxy in appearance. This lesion also can be excised widely, although the microscopic extent of the disease may exceed the visible margins. Intravenous fluorescein dye and ultraviolet light highlight areas that cannot be detected by the naked eye, and this assists in excision. Frozen-section examination of the surgical margins is also helpful but time-consuming. Paget's disease is occasionally associated with an underlying invasive adenocarcinoma; careful pelvic examination and proctoscopy are indicated in patients with this process.

Bartholin's gland carcinoma represents less than 1 percent of all vulvar malignancies and may be squamous carcinoma, adenocarcinoma, or adenosquamous or adenoid cystic carcinoma. Hemivulvectomy with dissection of the ischiorectal fossa and resection of involved contiguous structures is indicated. Because of the risk of inguinofemoral metastases, groin lymphadenectomy should accompany the vulvar operation.

GYNECOLOGIC OPERATIONS

Dilatation and Curettage (D&C)

At one time dilatation of the cervix and curettage of the endometrial cavity was among the most common surgical procedures performed in this country. Simple office biopsy and medical means of dealing with abnormal bleeding have largely replaced the need for diagnostic dilatation and curettage.

In some cases curettage is necessary for the relief of profuse uterine hemorrhage. It is indicated for removal of endometrial polyp or therapeutic termination of pregnancy and for retained placental tissue following abortion or obstetric delivery.

The patient is placed on the operating table in a lithotomy position, and the vagina and cervix are prepared as for any vaginal operation. The cervix is grasped on the anterior lip with a tenaculum. The cervix is gently pulled toward the outlet of the vagina. Some traction on the cervix is necessary to reduce the angulation between the cervical canal and the uterine cavity. A sound is inserted into the uterine cavity, and the depth of the uterus is noted. The cervical canal is then systematically dilated beginning with a small cervical dilator. Most operations can be performed after the cervix is dilated to accommodate a number 8 or 9 Hegar dilator or its equivalent. Dilatation is accomplished by firm, constant pressure with a dilator directed in the axis of the uterus (Fig. 39-25).

After the cervix is dilated to admit the curette, the endocervical canal should be curetted and the sample submitted separate from the endometrial curettings. The endometrial cavity is then systematically scraped with a uterine curette. The curettings are collected on a small piece of gauze or telfa. The curettings are then placed in the fixative. After the uterus has been thoroughly curetted, a ureteral stone forceps may be used to explore the endometrial cavity, searching for polyps or pedunculated neoplasms within the endometrial cavity. When the procedure is complete, the tenaculum is removed; the tenaculum site is evaluated for bleeding, and, if the puncture sites are bleeding, they are treated with a small amount of silver nitrate.

The major complication of D&C is perforation of the uterus. Perforation is diagnosed when the operator finds no resistance to a dilator or curette at a point where he or she normally would expect it. Perforation generally is treated in an expectant manner. The patient should be watched for several hours for signs of hemorrhage and be warned of the possibility of pelvic infection. A falling hematocrit and other signs of intraperitoneal bleeding indicate the need for laparotomy and control of the bleeding site. Any infection following D&C should be treated with antibiotics.

Uterine curettage is often required for incomplete abortion in the first or second trimester of pregnancy. Dilatation of the cervix in these cases is invariably present. Curettage of the postabortal uterus must be approached carefully because the uterus is extremely soft and perforation can occur with very little warning. Using the largest curette available is a safer choice than a small curette, which tends to cause perforation with less pressure. In the postabortal uterus, the endometrial cavity must be scraped thoroughly until the distinctive gritty feeling of curette against muscle is felt.

In recent years, suction curettage for incomplete abortion, hydatid mole, and therapeutic abortion has become popular. Suction machines fitted with cannulas that vary from 4 to 12 mm in diameter evacuate the uterus in less time and save blood loss. Most of these procedures are accomplished under sedation and paracervical block. Following the curettage, the uterine cavity is explored with a placental forceps or sponge forceps to remove any loose tissue within the cavity. Uterine perforation continues to be a concern when curettage is carried out for incomplete abortion. Perforation of the puerperal uterus is a much more serious problem because the organ is much more vascular than the nonpregnant uterus.

Postoperative bleeding should be modest if the curettage has been complete. Some operators control bleeding with the use of uterotonic agents such as vasopressin and prostaglandin-17-alpha.

Endoscopic Surgery

Endoscopic surgery, including both laparoscopy and hysteroscopy, has assumed a major role in gynecology. Laparoscopy, once used almost exclusively for diagnostic purposes and for tubal ligation, is now being applied to almost every kind of gynecologic procedure. Hysteroscopy has found an expanded role from purely diagnostic to removal of intrauterine pathology and ablation of endometrium for abnormal uterine bleeding. Although the limits of what is possible continue to be defined, the relative safety of some of these techniques in general use remains uncertain.

Laparoscopy

Laparoscopy was developed more than 20 years ago as a diagnostic tool and was soon adapted to perform tubal sterilization techniques. From the beginning, a few intrepid gynecologists used this approach for much more, including lysis of pelvic adhesions, treatment of endometriosis, and removal of ectopic tubal pregnancies. Slowly the role of the laparoscope expanded for conservative surgery and for removal of diseased tissue. More recently, laparoscopic approaches were developed for hysterectomy and gynecologic oncologic procedures. Ongoing research and experience continue to establish which of these approaches represent real advantages to the patients.

General Techniques for Laparoscopy

Placement of the Veress Needle and Primary Trocar

The standard method for laparoscopy remains the serial placement through the umbilicus of a retractable-pointed Veress needle for insufflation, followed by a sharp 10-mm primary trocar and sleeve. Although both reusable and disposable instruments are being used, there has been some movement back to sharp reusable instruments because of both cost and safety concerns.

When placing the Veress needle through the anterior abdominal wall, the goal is to minimize the risk of preperitoneal placement while avoiding retroperitoneal vessel injury by the use of proper placement techniques. In patients who are of normal weight or who are overweight but not obese (i.e., less than 200lb), instruments are placed through the umbilicus toward the pelvis. After the sacral prominence is palpated, the abdominal wall is elevated by grasping the skin and subcutaneous tissue midway between the symphysis pubis and the umbilicus in an effort to maximize the distance between the umbilicus and major vessels. The Veress needle is inserted through the base of the umbilicus at 45° from horizontal.

In the obese patient (>200 lb), the thickness of the abdominal wall requires an alteration of the approach for inserting the Veress needle. If the needle is placed through the base of the umbilicus at 45°, it may not reach the peritoneal cavity (Fig. 39-26). For this reason, it has been suggested that the Veress needle be placed at near 90° from horizontal. To minimize the risk of vascular injury, the umbilicus should be elevated (i.e., supported to avoid depression), and a standard length Veress needle should be used and checked for location as described previously. Following insufflation, a primary 10-mm trocar is also inserted at near 90° from horizontal.

Alternatively, open laparoscopic techniques continue to be refined and more widely applied. For these techniques, the anterior rectus fascia is incised with a scalpel, the peritoneal cavity is entered bluntly, and a blunt-tipped trocar is placed into the peritoneal cavity. Pneumoperitoneum is maintained at the site of entry either by sutures or by mechanisms built into the sleeves such as balloons or fascial threads. Although once reserved for patients with previous surgery, some laparoscopists use these techniques exclusively to avoid the risk of major retroperitoneal injury and to minimize the risk of bowel injury associated with closed techniques.

Placement of Secondary Trocars

As laparoscopic techniques advanced, the need for secondary trocars increased dramatically. After transillumination to locate the superficial vessels, an attempt is made to laparoscopically locate the inferior epigastric vessels. Secondary trocars are placed under laparoscopic visualization either 3 cm above the symphysis pubis in the midline or 8 cm above the symphysis pubis approximately 8 cm lateral to the midline. At the end of the procedure, the sleeves are removed and the sites observed for signs of hemorrhage. A full-thickness closure of any trocar site more than 5mm is performed to prevent herniation through the defect.

Power Instruments

Scissors and sutures have long been used for laparoscopic dissections and vessel ligation. However, because of the limitations of laparoscopy, including decreased depth perception and limited field of vision, innovative instruments have been developed for laparoscopic use both for tissue cutting and for vessel occlusion. Initially, unipolar electrosurgery was the only power instrument available, but because of concern about inadvertent damage to adjacent organs, other techniques were developed. Bipolar electrosurgery is excellent for hemostasis but has limited cutting ability. Laser, which can be aimed or placed in the proper location before activation, offers precision and some degree of hemostasis. Recently, an ultrasonic scalpel has been developed that avoids both the smoke and char associated with other power techniques.

Methods for Large-Vessel Occlusion

As techniques to remove tissue with significant blood supplies (e.g., adnexa, uterus) were developed, methods to effectively divide and occlude major vessels also were developed. Laparoscopic suture ligation, using either intra- or extracorporal knot tying, is relatively slow and technically difficult. For this reason, three alternative approaches have been widely applied. Pretied loops, linear stapling devices, and bipolar electrocautery have all been found to be relatively expeditious and effective, although rare cases of delayed bleeding have been reported with all three of these techniques.

Laparoscopic Procedures

Diagnostic Laparoscopy

This common procedure involves the placement of a 10-mm lens through an intraumbilical port, often with a 5-mm port placed above the symphysis for manipulation. Pelvic organs are closely inspected in a systematic fashion for signs of disease, and if tubal patency is an issue, a dilute dye solution is injected transcervically (chromopertubation). Biopsies can be obtained if malignancy is suspected.

Tubal Sterilization Procedures

As in diagnostic laparoscopy, a one- or two-port technique can be used. Tubes are occluded in the midisthmic section (approximately 3 cm from the cornua) using clips, elastic bands, or bipolar electrosurgery. With electrosurgery, approximately 2 cm of tube should be desiccated. Pregnancy rates after any of these techniques have been reported in the range of 3 per 1000 women.

Lysis of Adhesions

Pelvic adhesions usually are related to previous infection, either genital (i.e., pelvic inflammatory disease) or extragenital (e.g., ruptured appendix) in origin, or to endometriosis. Adhesions can be associated with decreased fertility or pain and can be lysed mechanically with scissors or any of the power techniques discussed above. Some degree of adhesion re-formation is unavoidable, and residual intrinsic tubal damage continues to interfere with fertility in most patients.

Adhesion re-formation can be minimized by achieving good hemostasis using discrete application of electrosurgery. Postoperatively, intraperitoneal solutions are used commonly to "hydrofloat" the adnexal structures, but controlled studies of their efficacy have been disappointing. Barrier methods have been shown to decrease adhesion formation in both animal and human studies but have not been demonstrated to improve outcome in terms of either subsequent pregnancies or pain relief.

Fulgarization of Endometriosis

Conservative laparoscopic treatment of endometriosis increases fertility and often helps with pelvic pain as well. This condition and the various approaches to treatment were considered earlier in this chapter (see Endometriosis).

Treatment of Ectopic Pregnancy

Laparoscopy has established itself as the primary treatment approach for ectopic pregnancies, a condition considered earlier in this chapter (see Ectopic Pregnancy, Treatment).

Ovarian Cystectomy

The laparoscopic removal of ovarian cysts less than 6 cm in diameter in premenopausal women has become common. Using a multiple-port technique, the peritoneal cavity is inspected for signs of malignancy, including ascites, peritoneal or diaphragmatic implants, and liver involvement. In the absence of signs of malignancy, pelvic washings are obtained, and the ovarian capsule is excised with scissors or a power instrument. The cyst is shelled out carefully and placed in a bag, intact if possible. The bag opening is brought through the lower port incision along with the 10-mm port. The cyst is then drained and the cyst wall removed. Hemostasis of the ovary is achieved with bipolar electrosurgery, but the ovary is usually not closed, since this may increase postoperative adhesion formation. Except in the obvious cases of simple cysts, endometriomas, or dermoid cysts, the cyst wall should be sent for frozen section to verify the absence of the malignancy. If malignancy is detected, immediate definitive surgery, usually by laparotomy, is recommended. All cyst walls are sent for permanent section and pathologic diagnosis.

In many cases the cyst will rupture prior to removal. This is always the case with an endometrioma that contains “chocolate” fluid. On rupture, the cyst contents are thoroughly aspirated, and the cyst wall is removed and sent for pathologic evaluation. The peritoneal cavity is copiously rinsed with Ringer's lactate solution. This is especially important when a dermoid cyst is ruptured, since the sebaceous material can cause a chemical peritonitis unless all the visible oily substance is carefully removed.

Although malignancies are not commonly encountered using these guidelines, there is concern that rupture may worsen the patient's prognosis. Data are accumulating suggesting that cyst rupture may not alter prognosis. Based on conventional wisdom, every effort should be made to remove ovarian cysts without intraperitoneal spillage.

Ovarian cysts larger than 6 cm or those discovered in postmenopausal women also can be removed laparoscopically. Because of the increased risk of malignancy associated with these situations, laparotomy is more commonly used. Laparoscopy may be a reasonable alternative in select patients if standard methods for staging are used in conjunction with appropriate frozen-section evaluation and expedient definitive therapy when indicated.

Removal of Adnexa

Occasionally, all or part of an adnexa must be removed. This may be the case with a large tubal pregnancy, a large hydrosalpinx, or when a small but growing cyst is found in a postmenopausal woman. Using a multiple- port technique, the vascular supply to the tissue is first desiccated with bipolar cautery and then divided with scissors. Alternatively, the ovarian vessels in the infundibulopelvic ligament can first be occluded with one of the techniques described (see Methods for Large Vessel Occlusion). Special care should be taken to identify and avoid the ureter, which lies retroperitoneally as it crosses the ovarian vessels and courses along the ovarian fossa (see Fig. [39-3](#)).

Once the adnexa has been excised and hemostasis is achieved, attention is turned to removing the tissue from the peritoneal cavity. Small specimens can be removed using a retrieval bag via a 12-mm port. The port is removed with the sack, and the fascial incision is enlarged, if required.

For larger specimens, a culpotomy may expedite this procedure. A 12-mm port is placed through the posterior cul de sac under direct visualization. A retrieval sack is placed through the port, and the port and specimen in the sack are removed together. The distensible peritoneum and vaginal wall will allow the removal of a large specimen through a relatively small defect, which can then be closed with a running suture vaginally. Prophylactic antibiotics may decrease the risk of infection.

Myomectomy

Uterine leiomyomas are often approachable via the laparoscope. Hemostasis is assisted by intrauterine injection of dilute vasopressin (10 units in 50 mL) at the site of incision. Pedunculated leiomyomas can be excised at the base using scissors or a power instrument. Intramural leiomyomas require deep dissection into the uterine tissue, which must be closed subsequently with laparoscopic suturing techniques. Because myomectomies are associated with considerable postoperative adhesion formation, barrier techniques are used to decrease adhesion formation.

Removing the specimen can be difficult. In general, morcellation is required, and power morcellators have been developed that significantly expedite this technique. Although leiomyomas of any size or location technically can be removed laparoscopically, it is yet to be proven that either menorrhagia or infertility, the two most common indications for myomectomy, are as effectively treated laparoscopically as they are by laparotomy.

Hysterectomy

Laparoscopy was first used to restore normal anatomy prior to vaginal hysterectomy. More recently, laparoscopy has been used to perform some or all of the actual hysterectomy to avoid laparotomy in patients with known pelvic adhesions, endometriosis, or in whom the uterus is enlarged by leiomyoma. Although multiple variations in technique exist, there are three basic laparoscopic approaches for hysterectomy: laparoscopic-assisted vaginal hysterectomy, laparoscopic hysterectomy, and laparoscopic supracervical hysterectomy. While basic techniques for each of these methods have become somewhat standardized, the indications and relative risk for each remain controversial.

The technically simplest, and probably the most widely applied, is the laparoscopic-assisted vaginal hysterectomy (LAVH). For this procedure, a multiple-port approach is used to survey the peritoneal cavity, and any pelvic adhesions are lysed. The round ligaments are then occluded and divided, and the uterovesicle peritoneum is incised. Next, the proximal uterine blood supply is occluded and divided. When the ovaries are removed, the infundibulopelvic ligaments (containing the ovarian vessels) are divided. If the ovaries are conserved, the uteroovarian ligament and blood vessels are divided and occluded. In some cases, the posterior cul de sac is incised laparoscopically as well. The remainder of the case is performed vaginally, including dissection of the bladder from the anterior uterus, ligation of the uterine vessels, removal of the specimen, and closure of the vaginal cuff.

A laparoscopic hysterectomy (LH) differs from an LAVH in that almost the entire hysterectomy is performed laparoscopically. This procedure is used for the indications listed above and also when lack of uterine descent makes the vaginal approach impossible.

LH is begun in a manner identical to LAVH. But after the proximal uterine blood supply is divided, the bladder is dissected from the anterior uterus. This is followed by a retroperitoneal dissection in which the ureter is identified along its entire pelvic course and the uterine vessels are selectively occluded and divided. The uterosacral ligaments are likewise divided and the posterior cul de sac incised. The specimen is removed vaginally, and the vaginal cuff is closed.

The third common laparoscopic approach is the laparoscopic supracervical hysterectomy (LSH)> This procedure has been advocated for all benign indications for hysterectomy. Technically, it is begun in a manner identical to the first two approaches. But after the proximal vessels are divided and the bladder is dissected from the anterior uterus, the ascending branches of the uterine arteries are occluded and the entire uterine fundus is removed from the cervix. The endocervix is either cauterized or cored out with a special instrument. The fundus is then morcellated and removed through a 12-mm abdominal port or through a special transcervical morcellator. The end result is an intact cervix and cuff, with no surgical dissection performed near the uterine artery and adjacent ureter. This approach avoids both a large abdominal incision and a vaginal incision. According to its advocates, this approach minimizes operating time, recovery time, and risk of both infection and ureteral injury. LH has yet to be widely applied, in part out of concern for the subsequent risk of developing cancer in the residual cervical stump.

Oncologic Procedures

As techniques developed, it became apparent that laparoscopy could be applied to oncologic procedures as well. In addition to the treatment of potentially malignant ovarian cysts, the laparoscopic approach also has been used for “second look” and staging procedures, including peritoneal washes and biopsy, partial omentectomy, and pelvis and periaortic lymphadenectomy. If positive nodes are discovered, treatment options often do not include laparotomy, and thus major surgery can be avoided without compromising patient prognosis. Laparoscopic approaches have been developed for definitive procedures, notably laparoscopically assisted radical vaginal hysterectomy.

A guiding principle is that the same care must be rendered laparoscopically that would be performed by laparotomy with the same or less risk of complications. Until the relative risk of complications and effect on prognosis have been established for these approaches compared with laparotomy, application of the laparoscopic approach in gynecologic oncology will remain highly controversial.

Risks of Laparoscopy

The many unique aspects of laparoscopy contribute to the distinctive complications associated with this approach.

Gas Embolism

Because pressurized CO₂ is used routinely to insufflate the abdomen for laparoscopy, gas embolization continues to be a rare but serious complication. This is related most commonly to misplacement of the Veress needle used for insufflation prior to primary trocar insertion. If the tip of the needle inadvertently enters the aorta or its branches, temporary distal arterial occlusion will result. Since CO₂ is quickly absorbed, no serious sequelae have been reported as a result of embolization, but arterial bleeding from this injury can be serious.

In contrast, inadvertent insufflation of the inferior vena cava or any of its venous branches can be fatal. Massive CO₂ embolism can result in partial or complete pulmonary arterial obstruction. This serious complication can be avoided in most cases by careful determination of Veress needle location prior to insufflation. Techniques for this include (1) ensuring lateral mobility of the needle, since retroperitoneal penetration will prevent this, (2) aspirating through the needle with a syringe, to rule out intravascular placement, and (3) the “hanging drop test,” in which a drop of saline placed on the hub of the needle is pulled into the hub of the needle when the abdominal wall is elevated to verify intraperitoneal placement. If any of these tests are not reassuring, the needle should be removed and replaced into the peritoneal cavity.

When CO₂ embolism is encountered, swift recognition and treatment can be lifesaving. Removal of the needle and placement of the patient in left lateral decubitus position are the first steps. In the presence of extreme hypotension, external cardiac massage has been suggested to break up large bubbles. Definitive treatment of this condition is central line placement and aspiration of the gas from the right-sided heart chambers and pulmonary vasculature.

Injury to Abdominal Wall Vessels

Abdominal wall vessel injuries have become more common with the development of more complicated operative laparoscopic procedures that use lateral trocar placement and larger trocars. These vessels include the inferior (“deep”) epigastric vessels, the superficial epigastric vessels, and the superficial circumflex iliac vessels (Fig. 39-27). Injury to the inferior epigastric artery can result in life-threatening hemorrhage. Injury to these or other vessels can result in significant hematoma or postoperative blood loss if unrecognized.

The primary methods to avoid vessel injury are knowledge of the vessels at risk and visualization of them prior to trocar placement when possible. The superficial vessels often can be seen and avoided by transillumination of the abdominal wall with the laparoscope. This is especially true in light-skinned and thin women. In contrast, the larger inferior epigastric vessels cannot be seen by transillumination because of their deeper location. But these vessels often can be seen laparoscopically and avoided as they course along the peritoneum between the lateral umbilical fold of the bladder and the insertion of the round ligament into the inguinal canal.

Because the vessels may not be visible in some patients either by transillumination or laparoscopically, it is important to know their most likely location and place lateral trocars accordingly. Although the traditional location used for lateral trocar placement was approximately 5 cm from the midline, a safer location may be 8 cm or more above the symphysis pubis and 8 cm from the midline, since both the superficial and inferior epigastric arteries are located approximately 5.5 cm from the midline (see Fig. 39-27).

Anatomic variation and anastomoses between vessels make it impossible to know the exact location of all the abdominal wall vessels. For this reason, other strategies also should be used to avoid vessel injury, including the use of trocars with conical tips rather than pyramid tips and the use of the smallest trocars possible lateral to the midline.

Injury to Retroperitoneal Major Vessels

Injury to major retroperitoneal vessels is one of the least common but most serious complications of the closed laparoscopic technique, occurring in approximately 3 per 10,000 laparoscopies. This includes both vessel perforation by the Veress needle with intravascular insufflation and vessel laceration by the 10-mm primary trocar.

Theoretically, the blind placement of sharp instruments through the umbilicus aimed toward the pelvis should rarely, if ever, result in vessel injury because both the aorta and the inferior vena cava bifurcate near level of the umbilicus. Unfortunately, in many patients the aortic bifurcation is at or below the level of the umbilicus, and in most patients the left common iliac vein crosses the midline below the umbilicus. The margin of error may be small, especially in thin patients, where the anteroposterior distance from the umbilicus to the retroperitoneal vessels may be as little as 2 to 3 cm (see Fig. 39-26).

The primary strategy to minimize the risk of vessel injury is to vary the angle of insertion based on the weight of the patient (see Primary Trocar Placement). An alternative strategy to avoid the risk is the exclusive use of an open technique. Although techniques for open laparoscopy have been available for years, the majority of gynecologic laparoscopic procedures are performed using a closed rather than an open technique. Because of the potential advantage of this technique in terms of patient safety, new techniques and instruments continue to be developed and evaluated.

Intestinal Injury

Another potentially serious complication of laparoscopic surgery is injury to either small or large intestines. An unrecognized bowel injury may occur at the time of trocar insertion, especially if the patient has had previous abdominal procedures that often result in bowel adhesions to the anterior abdominal wall peritoneum. In order to minimize the risk of bowel injury in patients who have undergone previous laparotomy, many gynecologic surgeons recommend the use of an open technique.

Another factor may be the type of primary trocar used. In recent years, disposable laparoscopic trocars have replaced reusable trocars at many hospitals. The extreme sharpness of disposable trocars that make insertion through the anterior abdominal wall easier also may have the potential to increase the risk of bowel injury. In one series, the risk of bowel injury with disposable trocars was approximately three times that previously reported for reusable trocars. Although the majority of these trocars have automatically extending “safety” shields designed to decrease the risk of inadvertent injury to these structures, no study has examined the relative safety of disposable versus reusable trocars.

Another laparoscopic risk to bowel is thermal injuries that may occur when power sources are used, such as electrocautery or laser. Regardless of the cause, major bowel injuries usually become obvious during surgery. Because of the limited field of view, some bowel injuries may not be seen during surgery. These injuries usually manifest 1 to 3 days after surgery, well after the patient has been released following these primarily outpatient procedures.

Urologic Injuries

Bladder injury is an uncommon laparoscopic injury, most commonly occurring as a result of retroperitoneal perforation during lower trocar placement or during sharp dissection of the bladder from the lower uterine segment during hysterectomy. The latter of these two situations is usually recognized intraoperatively; the first sign of the former may be postoperative hematuria or lower-port incisional drainage. Once diagnosed, large defects require layered closure, whereas smaller defects usually close spontaneously within days or weeks with the aid of transurethral catheter drainage.

Ureteral injury may occur as a result of any procedure that requires dissection or ligation of sidewall vessels, such as removal of an adnexa, since the ureter is adjacent to the pelvic peritoneum in the area of the ovarian fossa (see Fig. 39-3). This complication also has been reported after fulgarization of endometriosis on the pelvic sidewall.

Ureteral injuries, including complete ligation, partial resection, or thermal injuries, usually will manifest within hours to days of surgery. Complete obstruction most often manifests as flank pain, whereas the first sign of transection may be symptoms of intraabdominal irritation caused by urine leakage. Transperitoneal thermal injuries resulting from fulgarization of endometriosis may be similar to those after transection, but the appearance of symptoms may be delayed several days until tissue necrosis occurs.

Incisional Herniation

Incisional hernias after laparoscopy were rare prior to the use of large secondary ports (>5 mm) lateral to the midline. In recent times, incisional hernias have become a well-appreciated problem. A small peritoneal defect below the rectus abdominis muscle can allow bowel to become entrapped beneath the anterior rectus abdominis fascia. For this reason, closure of all layers of the incision is recommended whenever trocars greater than 5 mm are used at sites other than the umbilicus, and special needles have been developed for this purpose.

Although umbilical incision hernias also have been reported, fascial closure at the umbilicus is not commonly performed except when a large fascial incision is used for open laparoscopy. This is so both because the risk of herniation at this site is extremely low and because blind closure after removal of the laparoscope may risk bowel injury. Having the patient avoid heavy straining or lifting for 30 days after surgery may minimize the risk of umbilical herniation.

Hysteroscopy

Hysteroscopy, like laparoscopy, has gained widespread support as a very useful technique for both diagnosis and treatment of intrauterine pathology and for ablation of the endometrium as an alternative to hysterectomy for the treatment of abnormal uterine bleeding.

General Hysteroscopic Techniques

Type of instruments

Hysteroscopes can be divided into diagnostic, operative, and hysteroresectoscope. The lens for all three is identical. This is usually a fiberoptic lens and light source with an outside diameter of 3 mm and an objective lens that is offset up to 30° from the long axis of the instrument. In contrast, the sleeves for the three types of hysteroscopes vary considerably. The diagnostic sleeve usually has an external diameter of 5 mm and a single direction flow. Because outflow is limited, bleeding may impede a clear intrauterine view.

The operative sleeve, with an external diameter usually less than 10 mm, has a flow-through design with separate channels for input and outflow of distention media. A separate channel is available for placement of fine operating instruments.

The final type of sleeve is the hysteroresectoscope. This is also of a flow-through design and has an integral unipolar resecting loop identical to a urologic resectoscope. The loop can be replaced with a roller bar for endometrial ablation.

Distention Pumps and Media

Because of the small volume of the uterus coupled with its relative nondistensibility and the marked vascularity of the uterine wall, specifically designed pumps are required for hysteroscopy that are relatively high pressure (80–90 mmHg) and low flow. Several companies make CO₂ hysteroinsufflators that have these features and are very useful for diagnostic hysteroscopy. For operative hysteroscopy or for after cervical dilatation, a fluid medium pump is required. Preferably, this should be pressure regulated to approximately 80 mmHg to prevent excess intravasation of distention media. Although gravity alone has been advocated by some, the controllability of this technique is less predictable.

Several distention media have found widespread use. For diagnosis, CO₂ has been found to give excellent clarity. Although it has been found to be extremely safe in general use, fatal gas embolisms have been reported when CO₂ was used after cervical dilatation or intrauterine surgery. This suggests that the use of CO₂ should be avoided after cervical dilatation or any uterine instrumentation.

One of the first operative media used was 32% dextran 70 in dextrose. This syrup-like substance is usually introduced by hand with a large syringe. The advantage is simplicity and low cost. The view is excellent in the absence of bleeding. The disadvantage is the difficulty in completely removing the substance from the instruments. If this solution is allowed to dry in critical movable points, the instrument may “freeze up,” and it is very difficult to remove. In addition, intravascular intravasation can result in pulmonary edema.

More recently, aqueous solutions with pressure-controlled pumps have been used. For operative hysteroscopy, where electrosurgery is not being used, it is safest to use a balanced salt solution, such as Ringer’s lactate. Moderate fluid intravasation will be of no consequence in a healthy individual. However, intravasation of larger volumes can result in fluid overload, especially in a patient with any cardiac compromise.

When electrosurgery is used for hysteroresectoscope excision of leiomyomas or roller-blade endometrial ablation, a nonconducting solution such as glycine must be used. Significant vascular intravasation can cause hyponatremia, potentially resulting in cerebral edema, coma, or even death. For this reason, protocols must be followed rigorously to detect and treat significant intravasation whenever these solutions are used. Intraoperatively, differences in distention medium input and output should be calculated every 15 min. If the difference is greater than 500 mL, a diuretic should be given. If the difference is greater than 1000 mL, the procedure should be terminated as well. Whenever significant intravasation is suspected, serum sodium level should be checked immediately postoperatively and a few hours later because later hyponatremia, presumably due to transperitoneal absorption, has been reported.

Hysteroscopic Procedures

Diagnostic Hysteroscopy

This common procedure is often performed prior to uterine curettage to identify any focal abnormalities such as an endometrial polyp or a malignancy. This procedure is usually performed in the operating room with either general or regional anesthesia, although it has been performed by some as an office procedure with minimal analgesia.

After determining the position of the uterus, the anterior cervix is grasped with a tenaculum and traction placed to straighten the cervical canal. The lens and diagnostic sleeve are placed into the cervix, and distention medium is introduced with a pressure of 80 to 90 mmHg. The hysteroscope is slowly and carefully advanced toward the fundus, using tactile and visual cues to avoid perforation. The entire uterine cavity is inspected, and any abnormal anatomy is documented. As the hysteroscope is withdrawn, the uterocervical junction and the endocervix are examined.

Directed Endometrial Biopsy

If a focal abnormality of the endometrium is observed, directed biopsy may be more accurate than a simple uterine curettage. The cervix is dilated to allow passage of an 8- to 10-mm flow-through operating hysteroscope, and a balanced salt solution is used for distention. Once the hysteroscope is positioned in the uterine cavity, the area of interest is biopsied under direct visualization.

Polypectomy

If an intrauterine polyp is discovered, the base of the polyp is incised with hysteroscopic scissors, and the polyp is grasped with grasping forceps. The hysteroscope, sleeve, and polyp are removed simultaneously, since most polyps will not fit through the operating channel. Extremely large polyps may have to be removed piecemeal. Any residual base of the polyp may be removed with biopsy forceps.

Uterine Septum Resection

A septum may be resected with scissors, electrosurgery, or laser. Scissors are used most commonly in light of the minimal vascularity of septa and the decreased potential for bowel injury should inadvertent uterine perforation occur. An operating hysteroscope is placed into the uterine cavity, which will appear to be two tubular structures rather than the broad uterine fundus usually encountered. The septum is then evenly divided across the fundus. If scissors are used, rather than a power cutting instrument, the presence of bleeding indicates that the level of resection is shifting from the avascular septum to the vascular myometrium. After surgery, no special device is placed in the uterus because intrauterine synechiae formation is uncommon.

Removal of Intrauterine Synechiae

Intrauterine synechiae are almost always associated with previous uterine curettage, especially when performed in the immediate postpartum period. These synechiae may result in amenorrhea or infertility. The removal of synechiae is performed in a manner similar to that described above for a uterine septum, with some differences. The first is that the anatomy, and thus the visual cues for location of normal uterine wall, is completely unpredictable from patient to patient. Preoperative hysterosalpingography is usually very helpful. Findings can vary from a few small synechiae to complete obliteration of the cavity.

In difficult cases, simultaneous transabdominal ultrasound is extremely helpful in guiding the direction and limits of hysteroscopic resection. Standby laparoscopy should be available in the event of perforation, which is a significant risk in these patients. However, once pneumoperitoneum is achieved, abdominal ultrasound is no longer possible.

Following surgery, some type of intrauterine splint, such as an intrauterine device or a balloon catheter, is often placed to avoid synechia re-formation. Patients are usually placed on estrogen supplementation for a month and prophylactic antibiotics until the intrauterine splint is removed 1 to 2 weeks later.

Intrauterine Myomectomy

Pedunculated or submucosal leiomyoma can be removed safely hysteroscopically with subsequent improvement in both abnormal uterine bleeding and infertility. Because myoma tissue is relatively dense, a power cutting instrument is required. The choices are either laser or, more commonly, electrosurgery. For argon or Nd:YAG laser, a fiber is placed through the operating channel of the operating hysteroscope, and a balanced salt solution is used for distention. When electrosurgery is used via a hysteroresectoscope, an electrolyte-free solution, such as glycine or sorbitol, must be used because a balanced salt solution will dissipate the current and prevent cutting. Use of electrolyte-free solution requires a thorough understanding of the potential risk and prevention of hyponatremia, since fatal complications have been reported with its use (see Fluid Overload and Hyponatremia).

Both pedunculated and submucosal fibroids are shaved into small pieces with either the laser fiber or the hysteroresectoscope. In the case of a pedunculated fibroid, the urge to simply transect the stalk as a first step should be resisted unless the fibroid is 10 mm or less in size. Fibroids that are larger than this are difficult to remove in one piece without excessive cervical dilatation. Morcellation is much easier when the stalk is still attached for stability.

Since the field of view is obscured by multiple pieces of tissue, the hysteroresectoscope is removed and the tissue collected in the urologic pouch. The hysteroscope is replaced in the uterus, and the procedure is repeated until the pedunculated fibroid and its stalk are completely removed, or the submucosal fibroid is shaved flush to the adjacent wall of the uterine cavity. After surgery, some gynecologists will treat the patient with estrogen or place an intrauterine splint as described above (see Removal of Intrauterine Synechiae).

Endometrial Ablation

A recent treatment for abnormal uterine bleeding in the absence of endometrial hyperplasia is ablation of the endometrium using either a laser fiber or, more commonly, an electrosurgical "roller barrel." As described previously for myoma resection, an operative hysteroscope with a balanced salt solution is used for laser resection, and a hysteroresectoscope with an electrolyte-free solution is used for electrosurgery. For both techniques, the endometrium is destroyed down to the myometrium in a systematic fashion starting at the cornua and ending in the lower uterine segment. Electrosurgery has been used for resection of the endometrium with a loop electrode as well as for ablation. Both loop resection and laser ablation may have a somewhat greater chance of subsequent amenorrhea, but both appear to be technically more difficult with a greater risk of perforation than the more widely applied roller-barrel electrosurgical ablation. In general, these techniques result in amenorrhea in half the patients and decreased menstruation in another third over the first year of therapy. However, a large portion of these patients subsequently will require another ablative procedure for bleeding or a hysterectomy for residual bleeding or dysmenorrhea.

A theoretical risk of this approach is the delay of vaginal bleeding if the patient subsequently develops an endometrial malignancy. After several years of using this technique, this has not manifested as a significant risk. Nevertheless, the long-term risk of this problem remains uncertain, and patients undergoing this procedure should be aware of this.

Risks of Hysteroscopy

Gas Embolism

Gas embolism have been reported when using CO₂ for distention after intrauterine surgery. It is recommended that CO₂ not be used for any operative procedure or after significant dilation of the cervix. If symptoms of massive gas embolism occur during diagnostic hysteroscopy, the procedure should be stopped and the patient treated as described above (see Risks of Laparoscopy: Gas Embolism).

Fluid Overload and Hyponatremia

During operative hysteroscopy, significant intravasation of distention medium can occur through venous channels opened during surgery or transperitoneally as a result of any fluid forced through the tubes. Symptomatic fluid overload has been reported with all fluid distention media, including 32% dextran 70 in dextrose. The volume of distention medium introduced through the operating hysteroscope or hysteroresectoscope should always be compared with the volume retrieved using a urologic collection drape. When using a balanced salt solution (e.g., Ringer's lactate), symptomatic fluid overload is treated effectively with diuretics.

When electrolyte-free solutions are used for electrosurgery, the potential exists for serious and even fatal hyponatremia, even without significant fluid overload. Electrolyte-free solutions should not be used for hysteroscopy when electrosurgery is not required. When these solutions are used, careful monitoring of fluid balance should be performed every 15 min to detect intravasation.

Uterine Perforation and Bowel Injury

Uterine perforation is a common risk of uterine dilation prior to hysteroscopy. If it is not possible to distend the uterine cavity when the hysteroscope is placed in the uterus, perforation should be suspected. If no sharp instrument or power source has been placed through the defect, expectant outpatient management is appropriate.

Occasionally, perforation will occur during resection of a septum or leiomyoma or other operative procedures. If any chance of bowel injury exists, laparoscopy to evaluate contiguous bowel for injury is a reasonable precaution.

Intrauterine Synechia

The formation of adhesions between the anterior and posterior uterine walls, referred to as *synechiae*, is an uncommon complication after intrauterine surgery.

Although intrauterine devices, intrauterine catheters, and high-dose estrogen therapy have been advocated to decrease the risk of this complication, the efficacy of these treatments remain uncertain.

Abdominal Procedures

Incisions

The pelvic surgeon must consider a number of factors before beginning a pelvic operation. The most important requirement is that the incision provide adequate exposure for the anticipated procedure. Pelvic surgery is performed through vertical and transverse incisions. The majority of the incisions used in pelvic surgery are shown in Fig. 39-28. The midline incision is the most useful. It is simple and tends to bleed less than incisions made off the midline. The midline incision provides excellent exposure of the pelvis, and, when necessary, the entire abdomen is accessible for operation. This incision is more susceptible to hernia formation and is somewhat more uncomfortable than the transverse incision. The resulting scar occasionally is thicker than incisions made along Langer's lines, resulting in a less desirable cosmetic result.

Transverse incisions are used more often by a pelvic surgeon because the entire incision is centered over the area of operative interest. The incisions are more comfortable postoperatively and heal with a lower incidence of dehiscence or hernia formation. The most common transverse incision is the Pfannenstiel incision (Fig. 39-29). The skin is incised transversely approximately 2 cm above the symphysis pubis, and the incision is taken down to the rectus fascia, which is entered transversely. The rectus fascia is dissected bluntly away from the underlying rectus muscles in both a superior and inferior direction. The rectus muscles are separated in the midline, and the peritoneum is opened in the vertical midline.

The Maylard incision carries with it the advantages of a transverse incision but affords more exposure of the pelvis than that provided by the Pfannenstiel. The skin is incised transversely approximately 2 cm above the symphysis pubis, and the rectus fascia is opened transversely but not separated from the underlying rectus muscles. The rectus muscles are cut directly under the fascial incision, and several small bleeders in the body of the rectus muscles are clamped and coagulated. The epigastric artery and vein located just below the lateral edge of the rectus muscles are ligated and cut; the peritoneum is then opened transversely to afford good visualization and access to the entire pelvis.

The Cherney incision provides the advantages of a transverse incision and all the visibility provided by the Maylard incision. The incision is made in the transverse direction in the lower abdomen approximately 2 cm above the symphysis, the rectus fascia is opened transversely, the lower portion of the rectus sheath is dissected free of the rectus muscle, and the insertion of the rectus muscles on the symphysis pubis is visualized. The tendon of the rectus muscle is then cut free of the symphysis pubis, and the muscle is allowed to retract upward. The peritoneum is opened transversely. This incision is repaired by simply sewing the rectus tendon to the lower aspect of the rectus sheath just above the symphysis before closing the rectus sheath at the completion of the operation.

All the incisions used in pelvic surgery have advantages as well as disadvantages. The pelvic surgeon should anticipate the need for surgical exposure in the upper abdomen and in such cases choose a vertical incision. The Pfannenstiel incision is suitable for most operations for benign disease in the pelvis, but if wide pelvic or upper abdominal exposure is needed, the Maylard or Cherney incisions provide better operative exposure.

Hysterectomy for Benign Disease

The abdomen is entered through an appropriate incision. The upper abdomen is examined for evidence of extrapelvic disease, and a suitable retractor is placed in the abdominal wound. The self-retaining bowel is packed out of the pelvis and held in place with a retractor. The uterus is grasped at either cornu with Kocher clamps and pulled up into the wound (Fig. 39-30 A). The round ligament is identified and suture ligated and cut (see Fig. 39-30 B). If the ovaries are to be removed, the peritoneal incision is extended from the round ligament lateral to the infundibulopelvic ligament for approximately 2.5 cm. The retroperitoneal space is bluntly opened. The ureter is identified on the medial leaf of the broad ligament. The infundibulopelvic ligament is isolated, clamped, cut, and suture ligated. A similar procedure is carried out on the opposite side.

In the event that the ovaries are not to be removed, after ligating the round ligament, an avascular area in the broad ligament is chosen and the broad ligament bluntly fractured with a finger, producing an opening below the ovarian ligament and fallopian tube (see Fig. 39-30 C). The fallopian tube and ovarian ligament are clamped, cut, and ligated.

Upward traction is placed on the uterus. The peritoneum in the anterior cul de sac is opened between the ligated round ligaments. The bladder is mobilized by sharply dissecting it free of the anterior surface of the uterus and cervix (see Fig. 39-30 D). The uterine vessels are skeletonized by transilluminating the fold of the broad ligament and dissecting the avascular tissue off the uterine vessels. The peritoneum on the posterior surface of the uterus is dissected free of the uterus and then cut. Clamps are placed on the uterine vessels at the cervicouterine junction. The vessels are cut and the clamps replaced with suture ligatures (see Fig. 39-30 E).

The bladder is again examined to ensure that it has been mobilized sufficiently from the vagina near the cervix. The cardinal ligaments are clamped, cut, and ligated (see Fig. 39-30 F). Following division of the cardinal ligaments, the uterus is elevated and the vagina entered with scissors or a knife (see Fig. 39-30 G). The uterus and cervix are cut free of the vagina. Sutures are placed at each lateral angle of the vagina, and the cardinal ligament is sutured to either lateral vaginal angle (see Fig. 39-30 H). The central portion of the vagina is left open after repairing it with a running absorbable suture (see Fig. 39-30 I). Pelvic reperitonealization is not necessary. The rectosigmoid colon is allowed to return to the pelvis. The pelvic packs are removed and the small bowel allowed to return to the pelvis. The omentum is placed over the bowel and under the abdominal wound. The abdominal wound is closed in an appropriate manner. In some circumstances, uterine myomata interfere with the operative procedure and myomectomy or supracervical hysterectomy might be accomplished before removing the cervix.

Myomectomy

Myomectomy should be performed through an incision that will allow good visibility of the pelvis. Hemostasis for the procedure is aided by the placement of a Penrose drain (Fig. 39-31 A) around the base of the uterus pulled through small perforations in the broad ligament lateral to the uterine blood supply on either side. This “uterine tourniquet” is held in place with a clamp.

Further hemostasis may be obtained by placing bulldog or rubber-shod clamps on the infundibulopelvic ligament in order to control the uteroovarian blood supply. When possible, the uterine incision should be made in the anterior surface of the uterus in order to reduce the incidence of postoperative adhesions. An incision is made through the uterine musculature (see Fig. 39-31 B) into the myoma. The pseudocapsule surrounding the tumor is identified and the tumor is bluntly dissected out with scissors, a knife handle, or a finger. After the tumor is freed of its lateral attachments, it can be twisted to expose a pedicle that frequently contains its major blood supply (see Fig. 39-31 C). On occasion, several myomas may be removed through a single incision (see Fig. 39-31 D). The uterine wounds are closed with absorbable sutures to obliterate the deadspace and provide hemostasis (see Fig. 39-31 E). The uterine serosa is closed with a 000 absorbable suture placed subserosally if possible. A patch of Interceed to cover the uterine incision may prevent adhesion formation (Fig. 39-31 F).

Radical Hysterectomy (Modified from Okabayashi)

The patient is placed in a modified lithotomy position with legs in obstetric stirrups, hips abducted 45° and flexed 15°. The peritoneal cavity is entered through a Maylard incision (Fig. 39-32 A) after ligating and dividing the inferior epigastric vessels. The Maylard incision permits unequal exposure of structures on the lateral pelvic sidewall. Access to the retroperitoneum is obtained by ligating and dividing the round ligaments (see Fig. 39-32 C). A U-shaped incision is carried from one lateral abdominal gutter to the other, including the peritoneum of the bladder reflection (see Fig. 39-32 D). The pararectal and paravesical spaces are opened using blunt digital or instrument dissection, and narrow rigid retractors are placed to maintain exposure (see Fig. 39-32 E). Pelvic lymphadenectomy is performed by removing lymph nodes from the external, internal, and common iliac vessels (see Fig. 39-32 F) as well as the obturator fossa (see Fig. 39-32 G). If there are no pelvic lymph node metastases, paraaortic lymph node sampling is unnecessary. Isolation of the superior vesicle artery by lateral retraction brings the uterine artery into view; this vessel is skeletonized and clipped at its origin from the anterior division of the internal iliac artery (see Fig. 39-32 H). The branches of the posterior division are generally not visualized at the time of radical hysterectomy. Next, the structures inferior to the uterine artery in the cardinal ligament are clamped and divided, freeing the cervix and upper vagina from the lateral pelvic sidewall.

At this point, the proper ovarian ligaments and the proximal fallopian tubes may be transected between clamps (see Fig. 39-32 J). After the ovarian vessels are mobilized, the ovaries may be marked with vascular clips and suspended in the lateral abdominal gutters above the pelvic brim. This measure protects the ovaries if postoperative pelvic radiotherapy is to be given. The ureters are carefully detached from the posterior leaves of the broad ligament for a short distance and retracted laterally before the posterior cul de sac is entered and the rectovaginal space developed bluntly (see Fig. 39-32 L); the uterosacral ligaments are divided. Upward traction on the uterus facilitates dissection of the bladder inferiorly away from the underlying cervix and upper vagina. The ureters are freed from their investment in the paracervical tissue, allowing the bladder and ureters to be displaced inferolaterally, exposing the upper vagina and paravaginal tissues. The tissues are clamped and cut, taking care to remove a 3- to 4- cm “cuff” of vagina with the cervix (see Fig. 39-32 M). The vagina is closed, a suprapubic catheter is inserted, and the abdominal incision is repaired.

Radical hysterectomy is associated with 85 to 90 percent cure rates in patients without lymph node metastases and 65 to 70 percent rates in those with spread to the regional nodes. The primary morbidity is bladder denervation, which occurs to some extent in almost all women undergoing this procedure. Generally, loss of bladder sensation is the only deficit, although inability to void is not uncommon in the immediate postoperative period. Ureterovaginal fistulas occur in about 1 percent of all patients in recent studies.

Postoperative external-beam radiotherapy may be elected if nodal metastases, positive surgical margins, or parametrial tissue involvement is found. Because bladder and ureteral complications are more common in women undergoing postoperative radiotherapy, surgical candidates must be chosen with care.

Resection of Ovarian Cancer

Radical or modified radical hysterectomy is indicated in the treatment of epithelial ovarian cancer only if peritoneal tumor nodules obliterate the posterior cul de sac or extend to the retroperitoneal spaces. Generally, extrafascial (simple or conservative) hysterectomy (see Fig. 39-18) suffices in the resection of these tumors.

When hysterectomy and salpingo-oophorectomy are completed, the infracolic omentum should be removed by reflecting the fatty organ superiorly, isolating and dividing the right and left gastroepiploic vessels, and dissecting through the avascular posterior leaf before isolating and dividing the vessels in the anterior leaf of the omentum. If the omentum contains a large amount of disease, the gastrocolic omentum should be removed by isolating and dividing the short gastric vessels along the greater curvature of the stomach. In cases of extensive omental involvement, care must be taken not to injure the spleen, stomach, or transverse colon. Generous peritoneal biopsies should be obtained from the right hemidiaphragm, both lateral abdominal gutters, and the anterior and posterior peritoneum of the pelvis.

If gross peritoneal tumor is completely removed, the lymph nodes should be evaluated. The paraaortic lymph nodes may be exposed by reflecting the colon medially. These lymph nodes should be liberally sampled, keeping in mind that the primary venous drainage of the left ovary is the left renal vein and that of the right ovary is the inferior vena cava at the level of the renal vein.

Pelvic lymph node sampling is also an important aspect of surgical staging in ovarian cancer and is completed by removing lymph nodes from the distribution of the external and internal iliac vessels and obturator space above the level of the obturator nerve. This part of the staging procedure is facilitated by first opening the paravesical and pararectal spaces as described for radical hysterectomy above. Lymph node sampling is primarily a diagnostic procedure in the management of early ovarian cancers. There is little evidence that complete lymphadenectomy is therapeutic in patients with advanced and unresected disease.

Vaginal Procedures

Hysterectomy

The removal of the uterus through the vagina is preferred in many cases of myoma, uterine prolapse, intraepithelial neoplasia, and uterine bleeding disorders. Patients are more comfortable, and operative time, hospital stay, and recovery time are shorter than in cases of abdominal operation. Vaginal hysterectomy is an acceptable approach in those

patients in whom the uterus descends, the bony pelvis allows vaginal operation, the uterine tumors are small enough to permit vaginal removal, and the patient is amenable to vaginal operation. In the presence of large myomas, pretreatment with a *GnRH-a* will allow vaginal operation that would have been impossible previously.

The patient is placed in a high lithotomy position, and the pelvis is examined under anesthesia. This examination should confirm previous findings and provide assurance that the operation is possible through the vaginal route. The bladder is not catheterized before operation unless it is greatly distended. A weighted vaginal speculum is placed in the posterior vagina, and the cervix is grasped with a tenaculum and pulled in the axis of the vagina (Fig. 39-33 A). The posterior cul de sac is identified and entered with scissors (see Fig. 39-33 B). Mayo scissors are used to circumcise the cervix, and the mucosa is cut down to the pubocervical-vesical fascia (see Fig. 39-33 C). The vaginal mucosa and the bladder are sharply and bluntly dissected free of the cervix and the lower portion of the uterus. Care must be taken not to injure the bladder.

When the peritoneum of the anterior cul de sac is identified, it is entered with the scissors, and a retractor is placed in the defect (see Fig. 39-33D). The uterosacral ligaments are identified, doubly clamped, cut, and doubly ligated (see Fig. 39-33 E). The second ligature is held long. Serial clamps are placed on the parametrial structures above the uterosacral ligament; these pedicles are cut, and the clamps are replaced with ties (see Fig. 39- 33 F). At the cornu of the uterus, the tube, round ligament, and suspensory ligament of the ovary are doubly clamped and cut (see Fig. 39- 33 G). The procedure is carried out on the opposite side, and the uterus is removed. The first clamp is replaced with a free tie; the second clamp is replaced with a suture ligature that is transfixed. The second suture ligature is held long. The pelvis is inspected for hemostasis; all bleeding must be meticulously controlled at this point.

The pelvic peritoneum is closed with a running purse-string suture incorporating those pedicles which were held (see Fig. 39-33 H). This exteriorizes those areas which might tend to bleed. The sutures attached to the ovarian pedicles are cut. The vagina may be closed with interrupted mattress stitches, incorporating the uterosacral ligaments into the corner of the vagina with each lateral stitch (see Fig. 39-33 I). The vaginal cuff is inspected again for hemostasis. In most cases no vaginal packing is required. A catheter is left in the bladder until the patient has fully awakened and is ambulatory.

On occasion, the uterus, which is initially too large to remove vaginally, may be reduced in size by morcellation (Fig. 39-34). After the uterine vessels have been clamped and ligated, serial wedges are taken from the central portion of the uterus in order to reduce the uterine mass. This procedure will allow the vaginal delivery of even very large uterine leiomyomas.

Pelvic Support Repairs

Most pelvic support defects should be treated by the vaginal route. Vaginal operation allows repair of all support defects and generally allows an opportunity for the use of support structures that will prevent recurrence of prolapse. Uterovaginal prolapse is treated by vaginal hysterectomy. If the ligaments retain their strength, they may be used in a culdoplasty to provide good postoperative vaginal support. In the absence of good uterosacral ligament support, the operator can choose to support the vagina by suturing the vaginal apex to one or the other uterosacral ligament. On rare occasion, a partial colectomy with colpocleisis may be the surgical method of choice.

Cystocele repair generally follows vaginal hysterectomy. The anterior edges of the vaginal cuff are grasped with Allis clamps, and the vaginal mucosa is opened from the vaginal cuff to the base of the bladder. The bladder is sharply separated from the vaginal mucosa to completely expose the cystocele. A Foley catheter is placed in the bladder, and the bladder neck is identified by palpation. Serial interrupted mattress stitches are used to reduce the bladder by plicating the posterior bladder wall. Cystocele repair is accomplished by placing one or more layers of 2-0 absorbable suture in the bladder wall. After the bladder plication is complete, the vaginal mucosal flaps are trimmed, and the vaginal mucosa is closed with interrupted 2-0 absorbable sutures. After the cystocele repair is complete the vaginal apex may be closed with interrupted 2-0 absorbable sutures.

If a rectocele repair is required, Allis clamps are placed at the hymenal ring posteriorly, and the posterior vaginal mucosa is opened to a point above the apex of the rectocele. The vaginal mucosa is separated from the underlying rectum by sharp and blunt dissection. The vaginal mucosa is trimmed to reduce the vaginal luminal diameter, judged to be adequate if two inserted fingers fit the vaginal lumen with ease. The anterior rectal wall is imbricated with either interrupted or running stitches. The imbricating stitches pull the rectum in on itself, eliminating the rectal hernia when the posterior vaginal mucosa is reapproximated with a running stitch of 2-0 absorbable sutures.

Enterocele

These hernias can exist before hysterectomy, and when they do, they must be recognized and obliterated at the time of uterine removal.

After an enterocele has been diagnosed, it should be repaired in concert with the repair of other pelvic defects. If diagnosis is made at the time of pelvic laparotomy, the defect can be repaired by the abdominal route. The hernia sac can be obliterated using a series of vertical stitches to approximate the anterior rectal wall to the posterior vaginal wall in a technique described by Halban. Moschcowitz described a technique for the treatment of rectal prolapse that serves to obliterate the cul de sac very nicely. The Moschcowitz technique uses concentric purse-string sutures starting at the bottom of the cul de sac to obliterate the hernia.

Enterocèles frequently cause inversion of the vagina. By the time this has progressed far enough to produce symptoms, most of the normal supporting structures have been totally lost. This condition is treated by opening the vaginal mucosa over the enterocele sac, dissecting out the hernia sac, and removing it. The peritoneum is closed at the neck of the hernia with two or more absorbable sutures that approximate the anterior rectal wall to the posterior bladder wall. Anterior and posterior vaginal repairs are performed if appropriate. In most cases, sacrospinous fixation of the vagina will provide a more permanent solution to the vaginal prolapse.

An alternative approach might be an abdominal sacral colpopexy, in which the vagina is supported by a fabric graft placed between the anterior surface of the sacrum and the vaginal apex. If the abdominal route is chosen, either a Moschcowitz or Halban enterocele repair should be performed in association with the surgery.

Surgical Correction of Incontinence

Urethral detachment that causes stress urinary incontinence can be treated with a variety of repairs. A purely vaginal approach to this problem is the Kelly stitch, which involves plication of the periurethral structures at the bladder neck and tends to restore the posterior ureterovesical angle. The failure rate with this procedure is relatively high, and most operators prefer to choose some type of needle procedure such as the Peyrera, Raz, or Stamey procedure when operating vaginally.

The abdominal repair of stress incontinence is accomplished by the Marshall-Marchetti-Krantz procedure or the Birch procedure. The Marshall- Marchetti-Krantz procedure sutures the connective tissue on either side of the bladder neck to the periosteum of the symphysis. The Birch procedure pulls periurethral tissue up to Cooper's ligament if possible. The common denominator of both procedures is the elevation and angulation of the urethra at the bladder neck. Both the Marshall-Marchetti-Krantz and the Birch procedures are most valuable in patients who require repair of only a urethral detachment in the treatment of their urinary stress incontinence.

Pessaries

The vaginal pessary has been used since ancient times. It involves placing a plastic or rubber device into the vagina to support the vaginal apex and the vaginal walls. The device is finding little use in modern gynecology. When pessaries are used, they should be removed at least every 4 weeks. They produce discomfort secondary to pelvic pressure, and vaginal ulceration secondary to pressure necrosis is common.

Injuries Associated with Pelvic Surgery

Intestinal Injury

Adhesion formation is the most common antecedent to enterotomy at the time of pelvic operation. When lysis of adhesions is carefully carried out, this complication should occur rarely. Injuries to the small bowel range from serosal tears to through-and-through lacerations. Serosal tears are usually oversewn with a 4-0 nonabsorbable suture. Injuries to the lumen of the bowel are generally repaired using single- or double-layer closure; a single layer of 4-0 absorbable suture followed by an outer layer of interrupted nonabsorbable suture is effective. Lacerations that are extensive or involve multiple areas should be treated by resecting the injured segment.

Injury to the colon may occur during an operation to remove a left adnexal mass. When this injury occurs in a prepared colon, it is primarily repaired in a two-layer closure similar to that used in a small bowel repair. Large injuries may require resection of a segment of the colon.

Injury to the unprepared large bowel is a significant problem. A small wound that is promptly recognized may be treated primarily with a single- or double-layer closure. The patient should be treated with intraoperative and postoperative antibiotics to cover both anaerobic and aerobic organisms. A major injury to unprepared large bowel will require colostomy in most cases.

Ureteral Injury

Ureteral injury at the time of pelvic surgery is uncommon and largely preventable. The ureters are at highest risk when the infundibulopelvic ligament is clamped for removal of an ovary and when the uterine arteries or cardinal ligaments are clamped during the course of hysterectomy. In most cases, visualization or palpation of the ureter before placing a hemostatic clamp will avoid this complication. The ureter is at highest risk in the course of pelvic operation for endometriosis, pelvic inflammatory disease, and pelvic neoplasia. The surgery for most of these conditions allows the dissection of the ureter beginning at the pelvic brim with meticulous exposure down to and through the operative field. Transection of the ureter high in the pelvis may allow primary reanastomosis of the ureter. Ureteral injuries near the bladder generally require ureteral reimplantation into the bladder.

Bladder Injury

The pelvic surgeon must be constantly aware of the proximity of the bladder to the cervix and the anterior portion of the uterine fundus. The bladder must be dissected carefully free of the cervix at the time of hysterectomy. Hemostasis should be meticulous, and large clamps and ligatures should be avoided to prevent devitalization of the bladder wall. Operative entry into the bladder should be recognized immediately. If there is any question about the integrity of the bladder at the time of hysterectomy, a dilute solution of methylene blue will detect the defect and allow for closure.

Bladder injuries must be fully visualized in preparation for repair irrespective of whether they were produced by vaginal or abdominal operation. The surgeon must ascertain that the wound does not involve the ureter and that the resulting repair will not compromise ureteral function. 4-0 polyglycolic acid sutures are placed at either extent of the bladder laceration and held for retraction. The suture does not enter the lumen of the bladder. Once the bladder injury is delineated by the two initial sutures, the wound is closed with 4-0 interrupted polyglycolic acid suture in a running mattress stitch in order to strengthen the first layer and to remove any tension produced by the initial closure. An indwelling catheter should remain in the bladder for 5 to 7 days or until microscopic hematuria has disappeared. Some pelvic surgeons test their repair with a small amount of sterile milk. Milk is preferred to methylene blue or indigo carmine because it will not stain the tissue and therefore allows subsequent testing.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 40 - Neurosurgery

Julian T. Hoff
Michael F. Boland

GENERAL CONSIDERATIONS

A detailed history and a physical examination are the foundation of neurosurgical diagnosis. Headache, altered consciousness, memory impairment, speech difficulty, visual disturbance, weakness, paresthesia, and incoordination are some symptoms suggestive of central nervous system disease. Historical details suggest a cause (e.g., traumatic, neoplastic, vascular, infectious, degenerative, or metabolic) while neurologic examination permits anatomic localization of the lesion. For patients with nervous system disorders, an accurate history needs to be taken once, but the examination must be repeated and recorded often to gauge the course of the illness and to judge the urgency of other diagnostic steps.

Diagnostic Studies

Once a differential diagnosis is formulated using the information gathered from the history and the examination, diagnostic studies are used to confirm the definitive diagnosis. Common studies include plain film radiography, myelography, arteriography, computed tomography (CT), magnetic resonance imaging (MRI), *ultrasonography (USN)*, *electromyography and nerve conduction velocity testing (EMG/NCV)*, evoked potentials (visual, auditory, and somatosensory), *positron emission tomography (PET)*, electroencephalography (EEG), and cerebrospinal fluid (CSF) analysis.

Plain films, especially of the spine, are useful in trauma and degenerative disorders. Myelography, often combined with CT, is useful for evaluating spinal nerve roots and the spinal cord in trauma, tumor, and degenerative spine disease. CT and MRI provide detailed imaging of both cranial and spinal contents. They are useful in combination for visualization of bone and soft tissues. Arteriography provides detailed information regarding aneurysms, vascular malformations, and atherosclerotic disease. It is an essential tool in the evaluation of cerebral hemorrhage, embolic and thrombotic stroke, and preoperative planning for tumor surgery. *USN* is an important adjunct in the operating room, providing visualization of tumors, cysts, vascular malformations, and congenital anomalies lying beneath the exposed surface. *USN* is also used to visualize the brain and spinal cord in newborns.

EMG/NCV helps to assess peripheral nerve and nerve root lesions. It is also used to monitor brachial plexus and peripheral nerve recovery following traumatic injury. Visual and auditory evoked potentials can be employed in comatose patients to monitor the severity of head injury. Auditory potentials are useful in evaluating cranial nerve and brainstem function. Somatosensory potentials continually monitor spinal cord (dorsal column) function and are used often during surgical manipulation of spinal fractures and tumors. PET scans play an important role in epilepsy surgery, preembolization assessment of skull base tumors involving the cavernous sinus, and surgically untreatable vascular anomalies. EEG helps in delineating structural and metabolic disorders and can be employed during cerebrovascular surgery to monitor adequacy of cerebral perfusion. CSF analysis remains essential for many diagnoses, such as bacterial or viral meningitis.

SPECIAL SITUATIONS

Seizures

Seizures are useful clinical signs because the aura, onset, and type of seizure and the postictal state may localize the lesion. Seizures are particularly common in patients with neoplasms. Repetitive seizures should be treated vigorously. Phenytoin (Dilantin) is the drug of choice; 750 to 1000 mg may be given I.V. over 1 h as a loading dose. Supplemental doses usually consist of 100 mg given three or four times a day. Phenobarbital is also useful (32 to 65 mg, three or four times a day), but larger doses may depress consciousness. Diazepam (Valium) given intravenously in divided doses (10 to 50 mg) is highly effective in the control of status epilepticus but is not a good long-term anticonvulsant.

Raised Intracranial Pressure

Almost any space-occupying intracranial lesion can raise intracranial pressure (ICP). Clinical indications of elevated ICP are headache, stupor, diplopia, nausea, vomiting, and neck stiffness. Altered blood pressure and heart rate are late signs; typically the blood pressure is increased and the heart rate is decreased. Apnea may occur if ICP is very high. Raised ICP may be prevented and treated by the following measures.

Hyperventilation

The P_{aCO_2} should be monitored and maintained at about 35 mmHg. Brain swelling develops when P_{aCO_2} rises above normal (40 mmHg). A good airway is essential to maintain control of P_{aCO_2} .

Hypothermia

Because fever causes the brain to swell, the patient's temperature should be controlled by alcohol sponging, antipyretics (aspirin or acetaminophen), and hypothermia blankets. Chlorpromazine (Thorazine, 5 mg I.V. every 3 to 4 h) minimizes shivering during these maneuvers.

Osmotic Diuretics

Mannitol (1.5 g/kg/24 h) causes shrinkage of the brain and reduction of ICP. It should be used only when the clinical situation is desperate; its beneficial effect is transient, and the drug can severely alter serum electrolytes and osmolarity. Urea and glycerol are other osmotic diuretics that may be effective.

Steroids

Dexamethasone (Decadron, 4 to 6 mg every 4 h I.V. or orally) or methylprednisolone (Solu-Medrol, 125 to 250 mg I.M. or I.V. twice daily) lowers ICP by reducing brain edema. The drug probably acts to stabilize the blood-brain barrier.

Infections

CNS infections include meningitis, subdural empyema, brain abscess, and epidural abscess. Antibiotics penetrate the normal blood-brain barrier variably; however, in the presence of meningitis antibiotics are able to penetrate the blood-brain barrier better. Specific antibiotics should be given for specific infections. During the 1 to 2 days required to identify the pathogenic organism, broad-spectrum antibiotics should be given if the infection is life-threatening. Vancomycin (1.0 g I.V. every 12 h) and gentamicin (75 mg I.V. every 8 h) are the drugs of choice while awaiting culture results.

Antibiotics for prevention of CNS infection are rarely indicated. In patients with persistent CSF leakage, a broad-spectrum antibiotic is sometimes used (ampicillin or a cephalosporin).

Fluid Balance

Neurosurgical patients should have low-normal to normal intake and output of fluids (2000 to 2500 mL/24 h for an adult). Free water (5% dextrose in water) should be avoided because it causes brain swelling. The preferred I.V. solution is 5% dextrose in 0.5 normal saline with potassium supplements (40 mEq/day). Fluid balance should be monitored by daily weights and periodic measurements of serum electrolytes. Feedings by gavage may be started early (up to 3 days after injury or operation) provided gastrointestinal function is normal.

Some neurosurgical patients have profound disturbances of fluid balance. Inappropriate antidiuretic hormone (ADH) secretion, which most commonly occurs after head trauma, causes retention of free water, resulting in low serum sodium, high urinary sodium, low serum osmolarity, and high urinary specific gravity. Seizures and coma may be the first clinical signs of inappropriate ADH secretion. The best treatment is fluid restriction.

Diabetes insipidus is also common in patients with neurosurgical illnesses. Urine volume is high and specific gravity is low; serum sodium and osmolarity are high. Seizures and stupor may appear. The condition is treated by administration of I.V. or oral fluids. Sometimes antidiuretic hormone (vasopressin, Pitressin) is required to control urinary loss of water.

Patients who are given steroids, osmotic diuretics, anticonvulsants, and hyperosmolar feedings (typical neurosurgical patients) are prone to developing a hyperosmolar state, sometimes leading to hyperglycemic nonketotic coma. Consequently, careful monitoring of fluid and electrolyte status of the neurosurgical patient is essential.

Coma

Coma is a loss of consciousness from which the patient cannot be aroused by any stimulus. *Stupor* implies that the patient can be partially aroused by loud command or painful stimulus but promptly lapses into unconsciousness when the stimulus is withdrawn.

Diagnosis

The diagnosis of coma requires a careful history (generally from friends or relatives of the patient) and a complete physical examination, with specific attention to the neurologic examination. Laboratory and radiologic tests of value include serum electrolytes, blood glucose, and blood urea nitrogen assays, skull x-rays, and CSF examination via lumbar puncture. Medical consultation should be sought promptly.

Coma may be caused by poisoning, such as alcohol, barbiturate, and narcotic overdose; cerebral lesions, such as those caused by trauma, vascular accidents, tumors, infections, and epilepsy; metabolic disorders, such as diabetes mellitus, hypoglycemia, Addison's disease, uremia, hepatic disease, and eclampsia; and other stresses, such as severe infection, shock, asphyxia, heatstroke, and hypoxia. Diagnostic features of common types of coma are as follows.

Acute Alcoholic Intoxication

A history of alcohol abuse, alcoholic breath, flushed face, slow and stertorous respirations, diminished reflexes, and a blood alcohol level above 400 mg/dL establish this diagnosis.

Narcotic Poisoning

Even small doses of narcotics may cause respiratory depression and coma in patients with liver insufficiency, myxedema, emphysema, or head injuries and in debilitated or elderly patients. Findings include cold, clammy, cyanotic skin, pinpoint pupils, respiratory depression, and a feeble and often irregular pulse.

Acute toxicity caused by an overdose of a self-administered narcotic occurs commonly in some localities. The type and purity of the drug are difficult to determine, although a

companion or acquaintance may know the patient's drug habits. The examiner should look for needle marks in the arms and legs. Laboratory tests are valuable to determine barbiturate, alcohol, or narcotic levels.

Diabetic Coma

Coma may be precipitated in a diabetic patient by infection or by failure to regulate insulin dosage carefully. Diagnostic features include the following: history of diabetes, gradual onset, with blurred vision and thirst, air hunger or Kussmaul respiration, dehydration (soft eyeballs), acetone breath ("fruity" odor on breath), glycosuria, acetonuria, hyperglycemia, ketonemia, and low plasma bicarbonate level.

Hypoglycemia

Hypoglycemic reactions in diabetics may be precipitated by failure to eat, by vigorous exercise, or by insulin overdose. Mental confusion and bizarre behavior precede coma and convulsions. Tachycardia, sweating, tremors, and vomiting are other manifestations. Low blood glucose level confirms the diagnosis.

Management

Emergency Measures

Identify and treat any life-threatening condition immediately. Establish and maintain an airway to provide oxygenation. Insert an endotracheal tube if the respiratory rate is less than 10/min, if the P_{aO_2} is below 70 mmHg, or if the P_{aO_2} is greater than 50 mmHg with the patient breathing oxygen through a mask. Monitor arterial blood gases frequently and treat the patient for shock. When no cause for coma is immediately obvious, obtain blood for glucose determination and toxicologic analysis and then administer each of the following: 50 mL of 50% dextrose in water for possible hypoglycemia, 1 mL (0.4 mg) naloxone for possible narcotic overdose, and 100 mg thiamine I.V. for possible Wernicke's (alcoholic) encephalopathy.

General Measures

Observe the patient frequently, record neurologic and vital signs at regular intervals, and change the patient's position often to avoid postural pneumonia and decubitus ulcers. A lateral and slightly head-down position is best for patients who are likely to vomit. A suction machine and an alert attendant near the bedside are essential. Maintain ventilation and monitor urinary output through an indwelling catheter. Maintain fluid, electrolyte, and caloric intake. Nasogastric feeding should be started if the coma lasts more than 2 to 3 days. Avoid administering narcotics, sedatives, and other medications until the diagnosis is established; agitation can then be treated best by administration of parenteral diazepam (2 to 5 mg I.V. every 2 h as needed).

TRAUMA

Trauma is the single most common cause of death in children, adolescents, and young adults. The majority of accidents involving motor vehicles and falls include injury to the brain, spinal cord, and their supporting structures.

Scalp Injury

Scalp injury may cause hemorrhage and subsequent shock if not promptly treated. Bleeding can usually be controlled by a pressure dressing or by clamps applied to the galea aponeurotica. Scalp wounds should be closed as soon as possible. Lacerations that overlie a depressed fracture or a penetrating wound of the skull require debridement and closure in the operating room.

Simple scalp lacerations should be debrided, copiously irrigated, and closed primarily, taking care to approximate both the galea and the skin. A good galeal closure provides excellent hemostasis. Scalp avulsions typically include all layers of the scalp, sparing the underlying periosteum. If the avulsion is small, closure can often be accomplished primarily. Replacement of the avulsed part, employing microsurgical technique, is the preferred method of repair for large scalp avulsions, provided the tissue has been preserved adequately and surgery is not delayed.

If the injured scalp is not viable but the periosteum is intact, split-thickness skin grafts can be used to close the defect. The periosteum must be kept moist before operation. When the periosteum is absent or desiccated, closure is more difficult, since it is through the periosteal layer that the outer skull table receives its blood supply. In this instance, closure can be accomplished using free myocutaneous flaps attached by microsurgical vascular anastomoses.

Skull Fracture

Skull fractures are classified according to whether the skin overlying the fracture is intact (closed) or disrupted (open or compound), whether there is a single fracture line (linear), several fractures radiating from a central point (stellate), or fragmentation of bone (comminuted), and/or whether the edges of the fracture line have been driven inward below the level of the surrounding bone (depressed) or not (nondepressed).

Simple skull fractures (linear, stellate, or comminuted nondepressed) require no specific treatment. They are, however, potentially serious if they cross vascular channels in the skull, such as the middle meningeal artery or the dural venous sinuses. If these structures are torn, an epidural or subdural hematoma may form. A simple skull fracture that extends into the accessory nasal sinuses or the mastoid air cells is considered open, since it is in communication with air.

Depressed skull fractures often require surgical treatment to elevate the depressed bone fragments. If there are no untoward neurologic signs and the fracture is closed, repair may be done electively. Intraoperatively, the dura should be inspected and repaired.

Open skull fractures also require surgical intervention. Linear or stellate, nondepressed open fractures can be treated by simple closure of the scalp after thorough cleansing. Open fractures with severe comminution of underlying bone should be treated in the operating room, where thorough debridement can be carried out. The dura should be inspected to verify that a laceration has not been overlooked. Dural tears should be closed either primarily or with a fascial patch graft to reduce the risk of infection and prevent CSF leakage. Depressed, open skull fractures should be debrided, elevated, and closed in the operating room after preparations have been made for craniotomy, in case broader exposure of the underlying dura and/or brain is necessary.

Basal skull fractures involve the floor of the skull. Bruising may occur about the eye (raccoon sign) or behind the ear (Battle's sign), suggesting a fracture involving either the anterior or middle fossa, respectively. Isolated cranial nerve deficits can be seen in association with this fracture type because the nerves' exit foramina lie at the skull base. The facial nerve is frequently affected, with injury due to laceration or swelling. Most facial nerve deficits resolve spontaneously, so no specific therapy is warranted. On the other hand, complete transections of the facial nerve are usually explored, although the timing of exploration remains a matter of debate.

Any associated CSF rhinorrhea or otorrhea should be treated expectantly. Traumatic CSF leaks typically stop within the first 7 to 10 days. If a leak persists, lumbar CSF drainage can be implemented to seal the leak by lowering CSF volume and ICP. If drainage therapy fails, surgical exploration is indicated and accomplished by oversewing the defect with an intradural fascial patch graft. Fewer than 5 percent of patients require surgical repair. Prophylactic antibiotics are not used routinely because recent prospective studies have failed to demonstrate any benefit from their use.

Brain Injury

Injury to the brain is caused by rapid deceleration, acceleration, rotation, or a combination of these, associated with a blow to the head. The initial impact can produce neuronal and axonal disruption, which constitutes the *primary injury*. Any subsequent complication such as an intracranial hematoma, cerebral edema, hypoxia, hypotension, hydrocephalus, or endocrine disturbance characterizes *secondary injury*, which compounds the initial insult.

Mild head injury is usually not associated with significant primary brain injury, and neurologic deficits are limited to temporary loss of consciousness (concussion). Moderate to severe head injury, on the other hand, is typically associated with deficits that may or may not be reversible. Moreover, this degree of injury is usually accompanied by secondary injury.

Distortional forces causing the primary injury may be great enough to tear intraparenchymal capillaries, superficial subdural bridging veins, or epidural arteries and veins, allowing extravasation of blood and hematoma formation. Cerebral edema occurs in response to vasodilatation and blood- brain barrier disruption. Ischemia from hypotension or hypoxia can produce cell death and consequent cytotoxic edema. Disruption of CSF absorption by contamination of CSF with blood may lead to hydrocephalus. Inappropriate secretion of antidiuretic hormone or the development of diabetes insipidus can aggravate cerebral edema by altering fluid and electrolyte balance. These changes, either separately or in combination, can result in elevation of ICP.

Elevated ICP contributes to secondary brain injury by reducing *cerebral perfusion pressure (CPP)*, which is defined as the difference between mean arterial blood pressure (MABP) and cerebral venous pressure or ICP. Thus when ICP increases and MABP remains stable, CPP decreases. When CPP falls below 70 mmHg, cerebral perfusion is compromised, producing cerebral ischemia and compounding the primary brain injury with a secondary insult.

Intracranial hypertension is one of the most important factors affecting outcome in head injury. For this reason, aggressive management to circumvent cerebral blood flow reduction and secondary injury is imperative. Early resuscitative therapy should be initiated at the scene of the accident if possible with airway control and hyperventilation. In the absence of hypotension, osmotherapy can also be used early to reduce brain bulk.

Rapid clinical assessment is essential. Although extensive neurologic testing is limited in uncooperative or unresponsive patients, certain features of the exam are critical. The Glasgow Coma Scale (Table [40-1](#)), established in 1974, uses a numerical score to evaluate eye opening and verbal and motor behavior, both spontaneously and in response to stimulation. This scale is used to follow the patient's neurologic status and provides information about the ultimate outcome of the patient.

The initial neurologic exam determines whether diagnostic testing is indicated. It is unlikely that patients without headache, lethargy, or a focal neurologic deficit will suffer a secondary complication from their injury. Imaging studies are generally not indicated in the asymptomatic patient. Conversely, symptomatic patients with or without a focal deficit should undergo CT scanning of the head. If CT fails to disclose a lesion despite high clinical suspicion, carotid and/or cerebral angiography may be helpful to identify a vascular abnormality that cannot be appreciated by CT scan.

Treatment

After removal of any traumatic intracranial mass, the goals of management are normalization of *CPP* and prevention of secondary injury to damaged brain. Intracranial pressure

monitoring may be indicated, especially in those patients with marked depression or deterioration in neurologic function. Comatose patients who require emergency surgery (abdominal, thoracic, orthopaedic, etc.) should also have ICP monitoring, since frequent neurologic examinations are not possible during general anesthesia. Ventriculostomy to measure ICP also allows drainage of CSF, which can lower the pressure significantly. When the ventricular system is collapsed and ventriculostomy not feasible, intraparenchymal monitoring should be established.

Head elevation in the neutral position facilitates venous drainage; sedation reduces posturing and combative activity, both of which elevate ICP; hyperventilation to keep Pa_{CO2} around 35 mmHg lowers cerebral blood volume and ICP; prophylactic use of anticonvulsants prevents cerebral injury from seizures; mild dehydration with judicious sodium replacement and prompt treatment of SIADH (syndrome of inappropriate antidiuretic hormone secretion) protects the brain from insult secondary to fluid overload; prevention of hypotension reduces the extension of ischemic injury, and aggressive treatment of hypertensive episodes reduces cerebral blood volume and further disruption of the blood-brain barrier; and treatment of hyperthermia avoids an increase in the brain's metabolic demands. All of these management principles may be required simultaneously to achieve optimal homeostasis of the brain.

If ICP remains elevated despite these measures, mannitol (0.5 to 1.0 g/kg I.V.) and furosemide (1.0 mg/kg) can be used to reduce cerebral edema. Deep sedation with narcotics and the use of paralytic agents such as pancuronium may be helpful. Barbiturate coma, once popular in otherwise medically refractory ICP elevation, does not significantly change ultimate outcome and thus is seldom used. Corticosteroids, while occasionally used, have no proved benefit for severe brain injury.

Outcome after head injury depends on many factors. Increasing age and preexisting illness contribute to a poor prognosis. Penetrating injuries, particularly gunshot wounds, are associated with poorer outcome compared with blunt trauma. The presence of an intracranial hemorrhage also implies a suboptimal result. Subdural hematoma has a poorer prognosis than epidural hematoma. Combined subdural and intracerebral hemorrhage has the worst prognosis of all severe head injury subtypes. Other important factors that influence outcome include delay in treatment, multiple trauma, and systemic insults such as acidosis, hypoxia, and hypotension. Predictors of poor prognosis include evidence of brainstem dysfunction on the initial examination and refractory intracranial hypertension within the first few days of injury.

Diffuse Axonal Injury

Diffuse white matter injury is associated with anatomic disruption of axons throughout both cerebral hemispheres. This primary shearing injury is associated with high mortality and substantial neurologic morbidity. It results from rotational forces that are perpendicular to the axis of the white matter, causing axonal transection. Despite severe shearing injury, the brain may appear grossly normal, but numerous axonal transections are seen microscopically in white matter within 2 weeks of injury. Classically, small hemorrhages are present in the corpus callosum and cerebral peduncles on CT scan. Fiber tract demyelination subsequently develops, extending throughout the cerebral hemispheres and into the brainstem.

Epidural Hematoma

Hemorrhage between the inner table of the skull and the dura mater most commonly arises from a tear of the middle meningeal artery or one of its branches. Arterial bleeding strips the dura from the undersurface of the bone and produces a hematoma, which may increase in size and compress the underlying brain.

An epidural hematoma can also arise from torn venous channels in the bone at a point of fracture or from lacerated major dural venous sinuses. Since venous pressure is low, venous epidural hematomas usually form only when a depressed skull fracture has stripped the dura from the bone and left a space where the hematoma can develop.

Epidural hematoma classically follows a blow to the head that fractures the skull and causes a brief period of unconsciousness. After the patient regains consciousness, there may be a “lucid interval” during which there are minimal symptoms or signs. When the hematoma enlarges, hemispheric compression occurs. With time, the medial portion of the temporal lobe is forced over the edge of the tentorium, causing compression of the oculomotor nerve and dilation of the ipsilateral pupil. Similarly, compression of the ipsilateral cerebral peduncle causes contralateral hemiparesis, which may progress to decerebrate posturing. Coma, fixed and dilated pupil(s), and decerebration are a classic triad indicating *transtentorial herniation*.

Even though epidural hematomas are curable lesions, the mortality rate remains high because the severity of injury is often not recognized early. A patient may be seen during the “lucid interval” and discharged. Later, the patient may become unconscious because of progressive brain compression by the expanding hematoma.

Because of the danger of misdiagnosis, any patient with a history of a blow to the head leading to a period of unconsciousness should have a CT scan (Fig. 40-1). If an epidural hematoma is found, urgent craniotomy is indicated. If the CT scan is normal and the patient's examination shows no neurologic deficit, the patient may be discharged. An accompanying person should be instructed to awaken the patient frequently over the next 24 h to be certain that he or she remains arousable. Any deterioration in consciousness should prompt reevaluation.

Subdural Hematoma

Subdural hematomas may develop when veins bridging the cortex and the dura or venous sinuses are torn or when an intracerebral hematoma extends into the subdural space. They can be large even though the bleeding is of venous (low-pressure) origin.

Acute subdural hematomas are associated with severe head injury and arise from a combination of torn bridging veins, disruption of cortical vessels, and laceration of the cortex. The hematoma is best seen with CT scanning. Evacuation of the clot may lead to significant improvement, but often a major neurologic deficit remains because of the accompanying widespread parenchymal injury.

Subacute subdural hematomas become apparent several days after injury and are associated with progressive lethargy, confusion, hemiparesis, or other hemispheric deficits. Removal of the hematoma usually produces striking improvement.

Chronic subdural hematomas arise from tears in bridging veins, often after a minor head injury. Initially, the hematoma is small. Later, it becomes encased in a fibrous membrane, liquefies, then gradually enlarges. These lesions are more common in infants and the elderly. Typical presentation includes progressive mental status changes, with or without focal signs (hemiparesis, aphasia, etc.). Papilledema may be present. The diagnosis is confirmed by CT scanning. Treatment consists of hematoma drainage by trephination. Craniotomy may be necessary if the fluid reaccumulates.

Spinal Cord Injury

Traumatic injury of the spinal cord may result from vertebral fracture, fracture/subluxation, hyperextension of the cervical spine in the presence of a narrow spinal canal, herniation of intervertebral disc material into the canal, and penetrating injuries such as gunshots or stabbings. Neurologic involvement ranges from mild and transient to severe and permanent. Spinal fracture and cord injury should be suspected in head-injured patients, with or without coma, and in those patients with multiple injuries. It is best to assume that the spine is unstable initially and immobilize the patient on a backboard with a hard cervical collar until careful examination and diagnostic testing are done.

Clinical findings of spinal or spinal cord injury include spinal tenderness, extremity weakness, numbness or paresthesia, respiratory embarrassment, and hypotension. Spinal root involvement accounts for radiculopathy, characterized by motor and sensory impairment in the corresponding myotome and dermatome (Fig. 40-2). Spinal cord involvement produces myelopathy with variable manifestations.

A complete lesion, clinically defined as total loss of motor and sensory function below the level of injury, is associated with anatomic or physiologic transection of the cord. Acute transections are characterized by areflexia, flaccidity, anesthesia, and autonomic paralysis below the level of the lesion. Arterial hypotension is invariably present when the transection is above T5 because of the loss of sympathetic vascular tone. Common spinal cord syndromes are shown in Fig. 40-3.

Incomplete lesions of the cord may result in the Brown-Séquard syndrome, manifested by ipsilateral loss of motor function and position/vibratory sensation with contralateral loss of pain and temperature sensation below the level of injury. Anatomically, this presentation is explained by hemisection of the cord. The central cord syndrome is characterized by bilateral loss of motor function and pain and temperature sensation in the upper extremities, with relative preservation of these functions in the lower extremities. Typically, the distal upper extremities are more severely affected because the most medial portions of the corticospinal and spinothalamic tracts subservise these areas. The central cord syndrome is often seen following a hyperextension injury of the cervical spine, with or without fracture. The anterior spinal artery syndrome involves bilateral loss of motor function and pain and temperature sensation below the level of the lesion, with sparing of position, vibratory, and light touch sensation. This incomplete lesion develops when the anterior spinal artery is injured, rendering the cord ischemic within the distribution of the anterior spinal artery, affecting the anterior and lateral columns bilaterally. A common cause of anterior spinal artery syndrome is an acutely ruptured cervical disc.

Trauma to the lumbar spine may produce signs and symptoms of cauda equina compression. Presentation consists of multiple lumbosacral radiculopathies of variable severity. Lower extremity motor, sensory, and reflex functions may be affected, producing variable degrees of weakness, sensory loss (all modalities in the specific distribution of the roots involved), and diminution or absence of reflexes. Bladder distention from detrusor muscle paralysis, flaccidity of the anal sphincter, and loss of perineal sensation are common in severe injuries.

In addition to the neurologic deficit, acute spinal cord injury is accompanied by a variety of systemic responses. If the spinal cord is damaged above C3, respiratory efforts cease, accounting for this injury's high mortality at the scene of the accident. Although spontaneous ventilatory efforts can be initiated with injuries involving C4–C6, tidal volumes are often insufficient, accounting for progressive hypoxia and carbon dioxide retention. Airway obstruction, atelectasis, and pneumonia are common complications. Assisted ventilation is often required early after injury.

Ileus with gastric distention is common, necessitating nasogastric drainage. Similarly, bladder distention occurs because the bladder and pelvic floor muscles are flaccid. Bladder drainage will prevent overdistention, which may be severe enough to cause compression of the inferior vena cava and pelvic veins, impairing venous return to the heart and contributing to systemic hypotension.

Blood pressure is usually low if the cord injury is above the T5 level. This effectively denervates the sympathetic nervous system, which leads to increased venous capacitance and decreased venous return. The resulting hypotension is controlled by the administration of intravenous fluids. Colloid is preferred to reduce the threat of vascular overload and iatrogenic pulmonary edema. Postural changes that will precipitously drop the blood pressure, such as upper body elevation, should also be avoided.

Tachycardia is a common compensatory response to hypotension, but bradycardia is the rule when the cervical cord is damaged and the sympathetic input to the heart is lost. This type of bradycardia does not require treatment unless the patient is symptomatic or at risk for myocardial infarction or stroke because of age or other debilitating illness. If necessary, treatment with atropine and fluids is effective.

Once the patient is hemodynamically stable, spinal radiographs are essential, but only while the patient remains immobilized on a backboard in a hard cervical collar. Standard views are obtained, ensuring good visualization of the cervicothoracic junction (Fig. 40-4). Comatose and/or severely injured patients with multiple trauma should have good plain-film imaging of the complete spine. Fractured areas can be studied further with CT, using both axial and sagittal views (Fig. 40-5). If no abnormality is found by plain films or CT and a neurologic deficit exists indicating a spinal cord level, MRI or myelography followed by CT should be utilized to identify other causes of cord compromise, such as traumatic intervertebral disc rupture or spinal epidural hematoma.

Treatment

The objectives of treatment are to correct spinal alignment, protect undamaged neural tissue, restore function to reversibly damaged neural tissue, and achieve permanent spinal stability. Reduction and immobilization of any fracture/dislocation must receive top priority to meet these objectives.

Cervical spine malalignment can almost always be reduced by skeletal traction in the neutral position. Traction may be applied using skull tongs or a halo apparatus. Both are seated percutaneously through the outer table of the skull while the patient is kept supine and immobilized. The patient is then transferred to a special bed and traction begun. Frequent lateral view radiographs are obtained to document reduction and prevent overdistraction, which can lead to further cord injury. Once the spinal injury is reduced, traction should be maintained. Frequent follow-up films are then taken to confirm correct alignment.

Sometimes a cervical fracture cannot be reduced by traction alone without jeopardizing spinal cord function. Open reduction, usually through a posterior approach, combined with a fusion procedure may be necessary in those instances. This especially pertains to unilaterally or bilaterally locked facets.

Patients with thoracic and lumbar spine fractures are also treated with immobilization initially. Immobilization is less strict compared with that for cervical fractures, but the principles are the same. Patients are kept flat in bed without traction while flexion, extension, lateral bending, and rotational movements are avoided. They typically have fewer systemic complications associated with their neurologic injury but nevertheless require vigilance to prevent neurologic deterioration and provide the best chance for neurologic recovery.

Indications for early operation in patients with spinal cord injury include inability to reduce the fracture/dislocation satisfactorily by closed methods, neurologic deterioration in a patient with an incomplete cord lesion initially, severe compression of the spinal cord by an intraspinal mass shown by myelography or MRI, and a penetrating injury with or without a CSF leak. Open wounds, such as those inflicted by stabbings or gunshots, should be debrided and closed whether the cord injury is complete or incomplete. Early operation to stabilize the spine is warranted because this translates into early mobilization and rehabilitation. Either the anterior or the posterior approach may be used, depending on the nature of the spinal injury and the degree of instability.

If closed reduction is successful and the fracture is stable, external immobilization is necessary for a minimum of 3 months to ensure proper healing. If surgical reduction and/or fixation is necessary, external immobilization is still indicated. For the cervical spine, this involves a halo vest. Certain exceptions include anterior and posterior metal plating procedures in which a hard cervical collar may suffice (Fig. 40-6). The thoracic and lumbar spine usually require a plastic body jacket or plaster cast for a minimum, once again, of 3 months. Plain films are used to follow spinal alignment and the extent of fusion during the recovery period.

If any cord function is preserved immediately after injury, additional function usually returns, provided the cord and spine are protected from secondary injury. Patients with complete injuries rarely recover function below the level of the lesion. Rehabilitation for them is directed toward self-care and vocational readjustment. Most persons with these handicaps can eventually achieve independence. Life expectancy is shortened slightly in paraplegics and significantly in quadriplegics. Long-term problems associated with skin care and recurrent urinary tract infections account for the early mortality rate.

Peripheral Nerve Injury

Peripheral nerve injuries may be categorized functionally. *Neurapraxia* is a temporary loss of function without axonal injury. Structural damage does not occur. The foot that “goes to sleep” after crossing the legs is an example of functional loss without pathologic change. *Axonotmesis* is disruption of the axon with preservation of the axon sheath. Wallerian degeneration of the distal axon fragment occurs. Stretch or prolonged compression causes this functional and structural loss. Regeneration of the proximal axon occurs, but functional recovery depends on associated injuries, the amount of healthy proximal axon remaining after injury, and the age of the patient. *Neurotmesis* is disruption of both the axon and axon sheath with corresponding loss of function. Transection of a nerve causes this phenomenon. Regeneration occurs, but function rarely returns to normal.

Clinically, sensory and motor changes correspond to the peripheral nerve involved. A detailed history and a precise neurologic examination can localize the site of injury with great accuracy. Sensory findings are usually apparent early and remain so until regeneration is nearly complete. Compensatory motor function, often seen in the hand months after injury, is rarely seen acutely. A crude but clinically helpful sign of sensory regeneration is Tinel's sign. Percussion of the skin overlying the length of the injured nerve elicits paresthesias at the site where regeneration is occurring.

Radiographs of the injury site are helpful to look for fracture or foreign body. Electromyography (EMG) is not useful within the first 3 weeks of injury, but this diagnostic aid becomes highly effective to follow the state of the degeneration/regeneration process occurring later. Management decisions are often made depending on EMG findings weeks to months after trauma.

Treatment

Treatment of a lacerated nerve consists of primary repair when the wound is clean and uncomplicated, as in stab wounds, lacerations from glass, and surgical incisions. Secondary or delayed repair is indicated when the wound is dirty or complicated, as in gunshot wounds and avulsions, which disrupt tissue severely, making primary repair less successful. Secondary repair is best accomplished a few weeks after injury, when tissue viability is obvious, the likelihood of infection is reduced, and dissection planes are distinct. If end-to-end anastomosis of nerves is not possible because of tissue loss, nerve grafting using autologous sural nerve may be done. Intraoperative factors such as axial orientation of fascicles, proper coaptation, suture material, hemostasis, and suture line tension determine the outcome.

Nerve injuries in continuity (i.e., resulting from contusion or compression without laceration) are often explored if they do not improve within 6 weeks of injury, whether loss of function is complete or incomplete. Intraneural and extraneural scar tissue at the site of the lesion may prevent axonal regrowth by its constricting effect. Neurolysis releases the regenerating nerve fibers from the impinging scar and may improve functional recovery.

Prompt institution of physical therapy is also indicated for improvement of muscle function and maintenance of joint motion. It is the best means of minimizing the complications of denervation. The denervated portion of the limb is subject to muscle atrophy and fibrosis, joint stiffness, motor endplate atrophy, and trophic skin changes. The longer the denervation persists, the less likely will good function result.

Regeneration in a peripheral nerve occurs at 1 mm/day (roughly 1 inch per month), so improvement may not be obvious for many months. Factors that adversely affect the return of function include advanced age of the patient, proximal nerve injury, extensive nerve tissue loss, associated soft tissue injury, and presence of mixed sensorimotor function deficits. Unfortunately, incomplete neurologic recovery is often the rule. The use of tendon transfers should be considered to improve functional outcome if neurologic function is inadequate after recovery has ceased.

NEOPLASMS

Nervous system tumors represent almost 10 percent of all neoplasia. Of these, 15 to 20 percent occur in children. Nearly 70 percent of adult tumors are found above the tentorium (supratentorial), whereas 70 percent of childhood tumors are found below (infratentorial). CNS tumors are the most common solid tumors in children. Of all pediatric cancers, they are second in incidence only to leukemia.

The incidence of nervous system neoplasia decreases in the late teen years and begins to peak again by middle age. By then, only 25 percent of intracranial tumors are benign. This percentage, however, rises to 50 percent in the older patient because of the increasing incidence of meningiomas and schwannomas. Overall, there is a slightly greater incidence of tumors in men (55 percent), but schwannomas and meningiomas are more common in women. Of all CNS neoplasms, spinal tumors constitute approximately 15 to 20 percent.

Intracranial Tumors

Intracranial tumors exert both local and generalized effects by their presence within a closed bony structure, arising either from within or on the surface of the brain. Local effects of the tumor are either irritative or destructive. Focal seizures occur because of irritation of adjacent cortex, while a focal neurologic deficit develops because of compressive forces on nearby brain tissue. More generalized effects come from raised intracranial pressure due to the presence of the mass. These may be in the form of obstructive hydrocephalus, hemorrhage, cerebral edema, or simply the bulk of the neoplasm within the closed skull. The symptoms from generalized effects may be headache, occasional nausea and vomiting, decreased level of consciousness, and/or slowed cognitive function. Common tumors are listed in Table 40-2.

Astrocytoma

Astrocytes are glial (stromal or supporting) cells of the brain. Tumors arising from these cells make up over 60 percent of all intracranial tumors.

Low-grade astrocytomas (grades 1 and 2) constitute 5 to 10 percent of all intracranial tumors. When they involve the cerebral hemispheres, they typically arise during the fourth decade. They present with a 1- to 2-year history of illness, producing signs and symptoms such as headache, seizures, vomiting, mental status changes, papilledema, and focal neurologic deficits relevant to the hemisphere involved. On CT scan they appear as low-attenuation lesions with minimal contrast enhancement. These tumors are infiltrative and can rarely be totally excised. Surgery, however, is generally performed for diagnosis and debulking, usually followed by radiation therapy. Five-year survival rate with surgery and subsequent radiation is 35 to 50 percent. Some low-grade astrocytomas may become anaplastic and convert into higher grades of malignancy.

Low-grade astrocytomas of the cerebellum typically develop in children. They are often cystic with hamartomatous features, characteristics that carry a favorable prognosis. These tumors are often totally resectable. Radiation therapy is usually reserved for patients with incomplete resections, and follow-up has demonstrated continued tumor growth. The 10-year survival rate is over 80 percent for this common tumor of childhood.

Brainstem gliomas are also posterior fossa tumors that occur most often in children. The majority are benign astrocytomas. Patients may present with cranial nerve palsies, hemiparesis, and headache, often attributable to hydrocephalus. These tumors usually cannot be removed because of the risk of neurologic injury to the brainstem. They are

treated with radiation therapy once a diagnosis is established through either an open or closed biopsy procedure. Prognosis is a function of their location, with 5-year survival rates of 15 to 30 percent.

Optic nerve gliomas are astrocytomas involving the optic pathways anterior to the optic tracts. They occur in the chiasm in 75 percent of cases. Two-thirds of chiasmal tumors also invade the hypothalamus. The peak age for occurrence is 3 to 7 years, but their growth pattern is extremely variable. Some remain stable for years while others are relentlessly progressive. Those occurring strictly within one optic nerve can be excised, but more extensive tumors are generally irradiated. The value of radiation therapy, however, remains controversial.

High-grade astrocytomas (grades 3 and 4) are the most common primary intracranial tumor, constituting 20 to 30 percent of all intracranial tumors and half of all gliomas. Age at discovery ranges from 45 to 65 years. The frontal and temporal regions are most commonly involved. Due to their infiltrative nature, many of these high-grade neoplasms involve both cerebral hemispheres by invading through the corpus callosum (Fig. 40-7). Histologically these malignant lesions are composed of sheets of anaplastic cells with bizarre nuclei, numerous mitoses, endothelial proliferation, and abundant necrosis.

Patients with malignant gliomas present with a relatively short history of headache, focal neurologic deficit, mental status changes, or seizures. These tumors are readily imaged by CT and MRI, appearing as low-attenuation lesions with marked peritumoral edema and mass effect (Fig. 40-8). Ninety percent show enhancement with intravenous contrast. In spite of aggressive surgery, radiation therapy, and chemotherapy, these tumors are uniformly fatal. Although younger patients tend to do better, the average survival with surgical plus adjuvant therapy is about 12 months.

Chemotherapy is often used when tumor recurs, when the brain can no longer be irradiated, or when the tumor is not radiosensitive. It can be delivered intravenously, intraarterially, and/or intrathecally. Most drugs are directed at the tumor cells while others indirectly attack the cells by sensitizing them to subsequent radiation. Still in the early stages of development, immunotoxins toward tumor cells employ antibodies to tumor antigens, thus utilizing the host immune system as a means of destroying the cancerous tissue.

Meningioma

Meningiomas are relatively benign tumors that arise from the arachnoid layer of the meninges, usually occurring in the fourth through sixth decades of life. They affect women more often (65 percent) and arise in a variety of locations. The relative frequency of these locations include parasagittal 29 percent, sphenoid ridge 18 percent, convexity 13 percent, posterior fossa 10 percent, tuberculum sellae 9 percent, olfactory groove 5 percent, middle fossa 4 percent, foramen magnum 2 percent, and 1 percent or less occur in the orbit, gasserian ganglion, tentorium, or ventricular system. Altogether, meningiomas account for about 15 percent of all intracranial tumors.

Parasagittal (falcine) and convexity (hemispheric surface) meningiomas tend to present with seizures, focal neurologic deficits, and signs of intracranial hypertension. Sphenoid wing tumors often present with proptosis, decreased visual acuity, cranial nerve (III, IV, V, VI) palsies, seizures, or the more generalized effects of increased intracranial pressure (Fig. 40-9). Posterior fossa tumors present with cerebellar signs, hydrocephalus, lower cranial nerve palsies, or long tract signs (Fig. 40-10 A, B). Tuberculum sellae tumors are often accompanied by decreasing vision. Olfactory groove meningiomas may present with the classic Foster Kennedy syndrome of ipsilateral optic nerve atrophy and central scotoma with contralateral papilledema and bilateral loss of smell. Foramen magnum tumors may be difficult to recognize early and are often diagnosed as multiple sclerosis, syringomyelia, or cervical spondylosis. Patients complain of neck pain, clumsiness, and sensory disturbances in the upper extremities as well as gait difficulties.

Either CT or MRI is the principal means of diagnosis. Homogeneous pathologic contrast enhancement is characteristic of these neoplasms. Plain-film skull x-rays show abnormalities in 50 to 75 percent of patients with meningiomas, showing either tumor calcification, hyperostosis, increased vascular channels, or bone erosion. Angiography classically shows a “blush” in the late arterial phase.

The treatment for meningiomas is surgical. The techniques and goals are individualized for each specific tumor location, taking into account the patient's age and symptoms. In general, the goal is total excision, including the dura at the site of the tumor's attachment. If the meningeal origin cannot be excised, it is usually cauterized generously. Some meningiomas, such as those along the sphenoid ridge, may be “en plaque” tumors, which are flat and tend to spread along the inner table of the skull. This unique feature often precludes total resection.

The recurrence rate of meningiomas after surgery depends on the completeness of excision initially and the presence of malignant histologic features. If the tumor is totally removed, including its dural attachment, the recurrence rate is about 10 percent. If the dural attachment can only be cauterized, the recurrence rate is 20 percent. With subtotal resection, the recurrence rate is approximately 40 percent. Nonmalignant and malignant meningiomas have 15-year survival rates of 68 and 34 percent, respectively. Radiation therapy is generally reserved for malignant meningiomas, although it has also been used for incompletely removed nonmalignant meningiomas. Highly focused radiation therapy coupled with stereotactic guidance to the tumor site, referred to as radiosurgery, is now available. The single treatment, delivered in a few minutes, is an outpatient procedure. The efficacy of this form of therapy remains debatable, but it is believed to provide prolonged survival.

Medulloblastoma

Medulloblastomas are part of the primitive neuroectodermal class of brain tumors. They are thought to arise from primitive cells of the cerebellum, most likely the external granular layer. They constitute 5 percent of all gliomas and one-third of all fourth ventricular region tumors. Two-thirds of medulloblastomas occur in children, with an average age of onset of fourteen years. In children, these are more likely to occur in the midline, usually within the fourth ventricle, whereas adult tumors are frequently positioned more laterally. They commonly metastasize throughout the subarachnoid space via the CSF and can be found outside the central nervous system.

Children with medulloblastomas commonly present with elevated ICP secondary to hydrocephalus. Cerebellar signs may be prominent, with truncal ataxia and nystagmus. Some patients may present initially with symptoms related to spinal metastases. The diagnosis is generally made by identifying an enhancing mass on CT or MRI. MRI provides better images of this common posterior fossa tumor. Cytologic examination of the CSF is often positive for neoplastic cells.

Treatment involves aggressive surgical removal of tumor, followed by irradiation of the brain. If CSF cytology or spinal myelography or MRI is positive, the spinal axis is included in the radiated field. Chemotherapy is commonly used as well. The 5-year survival rate depends on the extent of tumor resection and the efficacy of radiation therapy. Some centers now report 5-year tumor-free rates of 75 percent in patients who have had gross total tumor resections, radiation therapy, and chemotherapy.

Schwannoma

This benign tumor arises from the Schwann cells that surround axons as they leave the central nervous system via cranial nerves. Schwannomas constitute 8 percent of all intracranial tumors and are almost twice as common in females as males. They occur usually in midlife. If associated with von Recklinghausen's neurofibromatosis, they may be multiple.

Schwannomas usually occur on sensory cranial nerves. The vestibular portion of the acoustic nerve (i.e., acoustic neuroma) is by far the most common site (Fig. 40-11). Depending on their size, these tumors usually produce hearing loss. As they enlarge, they may create facial numbness by compression of the trigeminal nerve and loss of coordination by compression of the adjacent cerebellar hemisphere. Only very late in their course do they cause facial weakness. Symptoms of increased intracranial pressure due to obstructive hydrocephalus may also present late.

Much less commonly, schwannomas arise from the trigeminal nerve, presenting as a mass in the middle fossa, associated with facial numbness. Some patients complain of lancinating, burning, episodic pain similar to trigeminal neuralgia. Only rarely do other cranial nerves serve as the primary site of origin for schwannomas. These less common sites would be more apt to be involved in cases of neurofibromatosis. Schwannomas occurring in patients with neurofibromatosis are often associated with other intracranial tumors such as meningiomas, astrocytomas, and ependymomas.

The treatment of schwannomas is surgical, with total resection being curative. Microsurgical techniques have reduced the risks of surgery and allowed preservation of cranial nerve function. Acoustic neuromas may be resected via the suboccipital or translabyrinthine approach, with a 95 percent probability of preservation of facial nerve function with small tumors. Even with the largest tumors, preservation is predictable in 50 to 70 percent of cases. The translabyrinthine approach is reserved for those patients who have complete hearing loss due to the presence of the mass. If this loss is incomplete, the suboccipital approach is preferred, allowing for hearing preservation by leaving the middle and inner ear structures intact. Preservation of hearing is directly related to tumor size. Mortality rate for acoustic tumor surgery is a function of tumor size as well, with most deaths occurring in patients with very large tumors.

Ependymoma

These tumors originate from the ependyma, which is formed by cells that line the ventricular system. They constitute 3 percent of all gliomas. Their incidence peaks at the age of 5 years and again at 34 years. They arise below the tentorium, usually within the fourth ventricle, in two-thirds of cases. When they occur above the tentorium, 50 percent are intraventricular. Infratentorial ependymomas are far more common in children.

The presenting symptoms are related to the tumor location. Typically, children harboring fourth ventricular tumors present with headache and vomiting related to associated obstructive hydrocephalus. They also tend to have ataxia secondary to cerebellar compression. Patients with supratentorial tumors present with signs of raised intracranial pressure as well, but this is usually due to brain edema. Focal neurologic deficits are common. These tumors are usually well circumscribed on CT scans as enhancing masses. MRI is the study of choice when the tumor involves the posterior fossa, since it far surpasses the imaging capabilities of CT in this region.

Treatment involves aggressive surgical excision first. Fourth ventricular tumors, however, can rarely be totally removed because they invade the fourth ventricular floor. They are quite radiosensitive, so radiotherapy can prolong survival. The 5-year survival rate with surgery and irradiation for both children and adults is about 50 percent. Chemotherapy is reserved for malignant ependymomas. All ependymomas have a propensity to recur despite aggressive management.

Oligodendroglioma

Tumors arising from oligodendroglial cells (the supporting cells that make myelin in the CNS) make up 4 percent of gliomas and usually occur in the cerebral hemispheres. Half of all hemispheric lesions occur in the frontal lobes. Up to one-third are mixed tumors containing populations of astrocytes in addition to oligodendrocytes. They are considered slow-growing tumors with benign histologic features, but as many as 25 percent may show some degree of anaplasia.

Patients present with a history measured in years; 80 percent have a history of seizures. Patients commonly have focal neurologic deficits and papilledema. These tumors are unique in their tendency to calcify, attesting to their indolent nature. On CT scan, a high percentage show calcification and most have contrast enhancement.

Treatment is maximum surgical resection. Radiation therapy is reserved for those tumors displaying malignant histologic features or for cases of subtotal excision. The 5-year survival rate is 50 to 80 percent. There are well-documented long-term survivors, however, with a reported 20-year survival rate of 10 percent.

Germinoma

These tumors occur most commonly in the region of the pineal gland and have thus been misnamed “pinealomas” in the past. The suprasellar region involving the hypothalamus is the second most common location. They are much more common in males and tend to occur in the second and third decades. They are composed of two cell populations, tumor cells and lymphocytes, and are histologically indistinguishable from testicular seminomas and ovarian dysgerminomas.

Germinomas in the pineal region tend to produce obstructive hydrocephalus and Parinaud's syndrome, characterized by upward gaze paresis, loss of convergence, and small, unreactive pupils. Those in the hypothalamic region may result in diabetes insipidus and emaciation (diencephalic syndrome). Germinomas are extremely radiosensitive and do not require aggressive surgical resection, but biopsy analysis is recommended prior to irradiation. The 5-year survival rate approaches 75 percent after radiation therapy.

Other, related germ cell tumors include teratomas, embryonal cell tumors, endodermal sinus tumors, and choriocarcinomas. These tend to arise in the same locations as germinomas. The germ cell tumors often produce compounds that serve as tumor markers. These markers are identified with immunoassays of the CSF or with immunoperoxidase stains of the tissue under light microscopy. All choriocarcinomas and 10 percent of germinomas secrete beta-human chorionic gonadotropin (beta-HCG). Endodermal sinus tumors produce alpha-fetoprotein (AFP). Embryonal cell tumors typically synthesize both beta-HCG and AFP. Teratomas secrete neither. These markers are useful for confirming the diagnoses and for measuring responses to treatment.

Epidermoid, Dermoid, and Teratoma

Epidermoids and dermoids arise from benign inclusions of epithelial elements within the central nervous system during closure of the embryonic neural groove. Epidermoids contain only stratified squamous epithelium. Dermoids contain skin appendages such as hair follicles, sebaceous glands, and sweat glands in addition to the squamous epithelium. Epidermoids tend to arise off the midline in such locations as the cerebellopontine angle, parasellar region, and in the diploë of the skull. Dermoids, in contrast, usually arise in midline structures such as the cerebellar vermis or the fourth ventricle. They are often accompanied by overlying bone and skin defects. Both tumor types are easily identified on CT and MRI scans. Surgical removal is the preferred treatment. Recurrences are usually managed with further surgery.

Teratomas are more common in children. By definition, they are composed of tumor cells representing all three embryonic germ cell layers. The more differentiated teratomas often contain cartilage and bone. If these elements are primitive, they are considered malignant. Like dermoids, they tend to occur in the midline in such locations as the pineal region, third ventricle, and posterior fossa. Treatment is surgical excision, along with irradiation in certain malignant types.

Hemangioblastoma

These benign tumors are quite vascular and usually occur in the cerebellum. They are uncommon and constitute only 1 percent of intracranial tumors. Typical presentation is in the fourth decade. Fifteen percent of patients have von Hippel-Lindau disease, an autosomal dominant disorder consisting of central nervous system hemangioblastomas, retinal angiomas, renal and pancreatic cysts, and renal cell carcinoma. Irrespective of this syndrome, many patients with hemangioblastoma have polycythemia. Sizable neoplastic cysts are present in 60 percent of cases, often with only a small mural nodule of tumor. Total surgical removal is curative. Radiation therapy is used when resection is not possible. Reoperation is recommended for recurrences. The long-term survival is excellent, with up to 80 percent of patients alive at 10 years.

Metastatic Tumor

Twenty-five percent of all intracranial tumors are metastatic. Malignant cells invade the central nervous system hematogenously and tend to lodge at the gray and white matter junction. Metastatic tumors occur singly or multiply and may involve virtually any portion of the brain or, less commonly, spinal cord. Although any malignancy has the potential to metastasize, the most common primary sites are lung, breast, kidney, testis, colon, and skin. The presenting symptoms are determined by the site(s) of the metastases. In general, these commonly include headache, mental status changes, seizures, and hemiparesis.

Metastatic lesions are best imaged with high-resolution MRI, but they may be mimicked by other lesions such as meningiomas, abscesses, primary brain tumors, and even aneurysms. If a metastasis is suspected, a workup to find the primary source is recommended. If the primary site is not identified, an excisional biopsy is indicated to establish the diagnosis.

In general, a symptomatic, solitary lesion that is surgically accessible should be removed. Surgery, however, should not be undertaken for multiple lesions or in patients who are severely afflicted by their primary disease. Treatment should also include preoperative dexamethasone, as in any brain or spinal cord tumor, to reduce adjacent brain edema. Whole brain irradiation is almost always indicated. Prognosis depends on tumor type, with the median survival ranging from 1 to 2 years. Long-term survivors have been reported with surgical removal of solitary brain metastases. Quality of life is almost always improved. There is little evidence that chemotherapy plays a significant treatment role.

Tumor metastasis to the leptomeninges (meningeal carcinomatosis) is also quite common, particularly in the childhood leukemias and in adults with lymphoma, breast and lung cancers, and melanoma. Patients may present with cranial nerve palsies, radiculopathies, or obstructive hydrocephalus. They often have signs and symptoms suggestive of meningitis. Analysis of the CSF is usually critical, often revealing an increased opening pressure, an elevated white cell count and protein level, and a decreased glucose level. There may or may not be identifiable malignant cells, but cytologic examination should always be done.

Treatment of meningeal carcinomatosis usually involves radiation therapy and intraventricular chemotherapy. Methotrexate is a common chemotherapeutic agent. The outlook for patients with leptomeningeal tumor spread is generally poor, but again a few long-term survivors emerge.

Spinal Tumors

Spinal tumors constitute approximately 20 percent of all CNS tumors. They are classified as intradural or extradural. Of the intradural variety, 84 percent are outside the spinal cord (extramedullary) and 16 percent are within it (intramedullary). Intradural tumors are almost always primary CNS tumors, whereas the majority of extradural tumors are either metastatic or primary bone tumors. The majority of intradural spinal neoplasms are benign and can often be excised surgically. Tumors occurring within the cord (intradural, intramedullary) tend to produce weakness, spasticity, and sensory loss. Extramedullary lesions present with radicular pain from nerve root (lower motor neuron) compression as well as with long tract (upper motor neuron) signs from cord compression. Patients with lesions involving the conus medullaris region may have early loss of bladder and bowel function; those with lesions in the cauda equina present primarily with leg pain and only later develop sphincter disturbances.

The definitive study for spinal tumors is MRI, although abnormalities on plain films and myelograms may be diagnostic. Plain films may show widening of the interpeduncular distance, bony erosion, enlargement of neural foramina, or a paraspinous mass. Myelography helps to determine the tumor's relationship to the spinal cord and dura. Postmyelogram CT can further define that relationship.

Neurilemoma and Neurofibroma

Typically benign, these are the most common spinal cord tumors, comprising almost 30 percent of the total. They are usually intradural, extramedullary in location. Of these, 13 percent have extradural extension through an adjacent foramen, producing the classic “dumbbell” shape of the tumor. Fourteen percent are totally extradural. The extradural component tends to enlarge the involved foramen. Treatment is surgical removal. Multiple neurofibromas are associated with von Recklinghausen's neurofibromatosis. In these instances, only symptomatic tumors should be removed.

Meningioma

Meningiomas constitute 26 percent of spinal cord tumors, are benign, and are usually intradural, extramedullary (Fig. 40-12). Fifteen percent occur extradurally. Two-thirds arise in the thoracic spine, affecting women in their fourth through sixth decades in 80 percent of cases. Surgical excision is the treatment of choice.

Ependymoma

Arising from the ependymal cells of the central canal of the cord, these intramedullary tumors constitute 13 percent of all spinal cord tumors. They occur more frequently in males. Nearly 60 percent are found in the conus medullaris region. Ependymomas should be surgically excised. Their distinct borders often allow complete resection; when total removal is not possible, radiation therapy is usually employed.

Astrocytoma

These glial tumors are derived from astrocytes and are often intramedullary. Their incidence is about the same as that of spinal ependymomas. Total excision is rarely possible due to their infiltrative nature. Low-grade astrocytomas, if recurrent, are usually reoperated. Radiation therapy is reserved for the malignant astrocytomas, but this is usually only palliative. While the growth rate of spinal cord astrocytomas is slow, prognosis is generally poor.

Lipoma

Lipomas constitute 10 percent of spinal tumors and are often associated with spina bifida and a subcutaneous lipoma. Although benign, they tend to be intertwined with cord tissue and are usually only partially excised. These do not require radiation, and the mortality is low.

Dermoid

Dermoids are congenital lesions usually found in the lumbosacral area. They often have an associated sinus tract to the skin surface and may present with infection. The treatment is surgical resection, including the sinus tract. The resection of the portion entering the spinal cord is usually incomplete. The long-term prognosis is good.

Metastatic Tumor

Up to 25 percent of all spinal neoplasms are metastatic in origin and most appear in an extradural location (Fig. 40-13). Common primary sites include breast, lung, prostate, and kidney. If the primary site is not known or if the neurologic decline is rapid, treatment is surgical decompression with biopsy. Otherwise, local radiation therapy is the treatment of choice. Other extradural malignant tumors include lymphoma, myeloma, plasmacytoma, chordoma, and osteogenic sarcoma. When significant bone destruction or the surgical decompression renders the spine unstable, surgical stabilization through an anterior or posterior route is often necessary.

Peripheral Nerve Tumors

The peripheral nervous system includes the peripheral and cranial nerves, spinal roots, and autonomic nervous system. Tumors can arise from any of these elements. The more common tumors are discussed below. More unusual tumors include gangliogliomas, neuroblastomas, paragangliomas, chemodectomas, and pheochromocytomas.

Schwannoma

This tumor arises from the peripheral nerve Schwann cells that provide the myelin sheaths for axons. Schwannomas tend to displace the nerve of origin and thus usually present as a painless mass. With continued growth, they can create pain in the distribution of the nerve. As they enlarge, nerve function deteriorates. They tend to arise from sensory nerves but may also be found on motor nerves. The treatment is surgical excision. The nerve of origin can usually be preserved. At times, however, total excision may mean division of the parent nerve. If the nerve serves a significant function, it is preferable to leave a portion of the tumor in order to spare the nerve. This is justified because malignant transformation is rare.

Neurofibroma

Neurofibromas differ from schwannomas in that they actually engulf the nerve of origin, since they arise from the nerve itself. They are often cutaneous, making it difficult to identify one specific nerve of origin. When associated with von Recklinghausen's neurofibromatosis, they are usually multiple. When found singly, treatment is resection. When multiple tumors are present, only the symptomatic ones should be resected. Removal requires sacrifice of the nerve if that nerve is expendable. If the function of the nerve is critical, a portion of tumor should be left attached to the nerve. Unlike in the case of schwannomas, patients with neurofibromas should be followed closely, since these tumors have a higher incidence of malignant transformation.

Malignant Nerve Sheath Tumor

These tumors typically occur after the age of thirty. The treatment of choice is radical wide resection. If there is evidence of muscle or soft tissue invasion, amputation of the involved extremity is recommended. These tumors are generally resistant to radiation therapy.

CEREBROVASCULAR DISEASE

Cerebrovascular disease is the third most common cause of death in the United States and is a significant cause of disability. Death and disability are due either to ischemia causing focal or diffuse infarction or to hemorrhage causing compressive mass lesions. Although discussed separately, cerebrovascular problems producing infarction may become hemorrhagic, and hemorrhagic lesions may lead to infarction.

Ischemic Vascular Disease (Stroke)

Ischemia and subsequent infarction of the brain can occur in the distribution of any of the cerebral vessels; thus any portion of the cerebrum, brainstem, or cerebellum may be affected. Because the carotid circulation provides the greatest blood supply to the brain, ischemia and infarction within its distribution are most common. Ischemia may be the result of diminished flow secondary to stenosis or occlusion of major arteries or due to transient or permanent occlusion of smaller arterioles from intravascular emboli.

The most common cause of stenosis or occlusion of large vessels is atherosclerosis. This disease often develops extracranially at the origin of the internal carotid artery in the neck but may occur in the carotid siphon (that portion of the artery within the cavernous sinus), the distal internal carotid, or even the proximal middle cerebral artery (Fig. 40-14).

Arterial emboli usually originate either from atherosclerotic ulceration in the region of the carotid bifurcation or from sources within the heart. The heart is a common source of emboli when a mural thrombus forms after a myocardial infarction or as a result of atrial fibrillation. Other risk factors for cerebral ischemia include hypertension, diabetes, hypercholesterolemia, obesity, smoking, and family history of stroke.

Since there is no effective medical or surgical therapy for a completed stroke, the goal of neurosurgical intervention is to identify stroke-prone patients and reduce their risk of cerebral ischemia. These high-risk patients are best identified by a history of transient ischemic attacks (TIAs), which take the form of either transient cerebral ischemia or amaurosis fugax. Transient cerebral ischemia in the carotid circulation usually consists of temporary hemianesthesia, hemiparesis, or aphasia. Amaurosis fugax is transient loss of vision in one eye. Ischemia in the vertebrobasilar system may cause transient diplopia, dizziness, dysarthria, dysphagia, weakness, numbness, loss of vision, or even loss of memory.

Most ischemic episodes last seconds to minutes and rarely longer than 30 min. As long as the neurologic deficit resolves within 24 h, the episode is, by definition, a TIA. A reversible ischemic neurologic deficit (RIND) is one that lasts 24 h to 3 weeks. Ischemic deficits lasting longer are considered completed strokes. Careful questioning of individuals with completed strokes reveals that 60 percent had a prior history of TIAs, 20 percent presented in a slow, stepwise fashion, and only 20 percent were sudden in onset.

Patients with TIAs or slow-onset strokes are potential candidates for preventive surgical intervention. Surgical procedures to prevent stroke are directed toward either removal of the source of emboli or augmentation of blood flow to the brain. Operations for these conditions include carotid endarterectomy and microvascular bypass. Potential candidates generally undergo a CT or MRI scan of the brain to evaluate any degree of cerebral infarction and to rule out other diagnoses such as tumor, subdural hematoma, or subarachnoid hemorrhage. Patients then undergo angiography, including the aortic arch and the carotid, vertebral, and cerebral arteries. Noninvasive studies of the carotid circulation are less accurate, although they are useful as screening procedures because of their low risk.

Carotid endarterectomy is indicated when ipsilateral symptoms of cerebral ischemia or amaurosis fugax exist and angiography demonstrates either significant stenosis (usually more than 75 percent) or ulceration in the accessible portion of the common and/or proximal internal carotid arteries (Fig. 40-15). The procedure consists of opening the affected portion of the carotid artery under systemic heparinization and removing the atherosclerotic plaque. The mortality from carotid endarterectomy is about 1 percent and the neurologic morbidity 5 percent in experienced hands.

A number of patients present with cerebral ischemia ipsilateral to an occluded internal carotid artery or with stenosis of the internal carotid or middle cerebral artery that is not surgically accessible. For these patients with inadequate collateral cerebral circulation, a microvascular bypass procedure is sometimes indicated. The most common of these is the superficial temporal artery to middle cerebral artery anastomosis (STA- MCA).

Intracranial Aneurysm

Intracranial aneurysms are diseased dilatations of the cerebral arteries, their walls consisting of ballooned-out tunicae intima, media, and adventitia with a variable degree of intraluminal or mural thrombus. Most are congenital in origin, evolving and developing during life. They may become atherosclerotic. Aneurysms are typically found at the bifurcation of the major vessels of the circle of Willis. Up to 20 percent of patients with aneurysms have multiple aneurysms, and 1 percent demonstrate an associated arteriovenous malformation (AVM). If aneurysms are found more peripherally in the cerebral vasculature, secondary causation such as trauma or infection should be considered.

Over 85 percent of cerebral aneurysms occur in the carotid or "anterior" circulation. Approximately 30 percent arise from the intracranial portion of the internal carotid artery, usually at or near the origin of the posterior communicating artery. Another 30 percent occur in the region of the anterior communicating artery. About 25 percent arise from the middle cerebral artery, usually at its first major branch point, which is commonly a trifurcation. Aneurysms of the vertebrobasilar or "posterior" circulation are most frequently found at the tip of the basilar artery but may occur more proximally along its trunk. The origin of the posterior inferior cerebellar artery is the next most common location.

Patients with intracranial aneurysms most commonly present with signs and symptoms of subarachnoid hemorrhage (SAH). Eighty percent of nontraumatic subarachnoid hemorrhages are caused by aneurysm rupture. The patient notes a sudden severe headache commonly followed by neck stiffness and photophobia due to associated meningeal irritation caused by the subarachnoid blood. Transient loss of consciousness may occur. Some patients may develop a focal neurologic deficit or become comatose as a result of the acute rise in intracranial pressure. The severity of the SAH can be graded, as shown in Table 40-3. In general, the lower the grade, the better the outcome.

Not all patients with aneurysms present with symptoms related to rupture. Through mass effect, an internal carotid artery (ICA) aneurysm may compress the optic (II) nerve (Fig. 40-16 A, B), causing monocular blindness, or the oculomotor (III) nerve, producing a palsy characterized by diplopia, ptosis, and dilated pupil. An ICA aneurysm within the cavernous sinus may compress the abducens (VI) nerve and create diplopia. A giant aneurysm (larger than 25 mm in diameter) of the basilar tip may block the cerebral aqueduct and create hydrocephalus. Rarely, an aneurysm may be large enough to be mistaken for a tumor.

The diagnosis of SAH is usually made clinically and confirmed either by noting blood within the subarachnoid spaces on CT scan or by finding bloody CSF with xanthochromia on a lumbar puncture (LP). The CT scan should be obtained first, since it usually spares the patient an LP and also eliminates the potential risk of brainstem compression from herniation if an unsuspected mass lesion is present. Complete cerebral angiography is then employed to identify and delineate the aneurysm and, at the same time, rule out multiple aneurysms or an associated AVM.

Once the diagnosis of aneurysmal rupture is confirmed, the patient is placed on a medical regimen to reduce the risk of rebleeding. This includes strict bed rest with the head elevated. Stimulation is kept to a minimum. Blood pressure is tightly controlled to keep it below 150 mmHg systolic. Careful observation is necessary to watch for signs of raised intracranial pressure which may be attributable to delayed hydrocephalus. Anticonvulsants are started for seizure prophylaxis. Calcium channel blockers are used to reduce the risk of vasospasm.

The ultimate treatment of aneurysms is microsurgical dissection and obliteration, usually by placing a metallic clip across the aneurysms's neck via a craniotomy. Timing of surgery depends on the clinical grade of the patient (see Table 40-3). Good-grade (I and II) patients should usually undergo operation within 72 hours of rupture. Poor-grade (III and IV) patients should continue intensive medical management until they improve to a lower grade, if possible, because mortality rises with grade. Surgically accessible unruptured aneurysms should be operated on electively to prevent rupture. Some inaccessible aneurysms can be effectively obliterated by embolization using interventional neuroradiologic techniques.

Complications of aneurysmal rupture include a 30 percent rebleeding rate within the first 8 weeks if the lesion remains unrepaired, hydrocephalus from obstruction of the arachnoid villi by subarachnoid clot, vasospasm, intracerebral hematomas, raised intracranial pressure, and seizures. The most significant and least understood of these is cerebral vasospasm. This phenomenon occurs most frequently within 4 to 7 days after the hemorrhage and results in narrowing of adjacent cerebral arteries. Vasospasm may be noted on angiography without any untoward clinical effects, or it may produce profound and life-threatening cerebral ischemia in the distribution of the involved vessels.

Angioplasty done by interventional neuroradiologic techniques can, in selected cases, increase cerebral blood flow in order to overcome the spasm. Improved perfusion through dilated spastic arterial segments combined with induced hypertension can reverse ischemic neurologic deficits. The use of inotropic support and intravascular volume expansion, usually with colloid and red cell transfusion, is also beneficial. Cardiovascular status is monitored continuously by means of a Swan-Ganz catheter.

Patients who undergo elective clipping of unruptured aneurysms have better outcomes than those with ruptured aneurysms because the brain has not been injured by the subarachnoid hemorrhage. In addition, aneurysms of the internal carotid artery carry less risk than those of the vertebrobasilar system, with the exception of complex anterior communicating artery aneurysms. In general, if the aneurysm can be clipped and vasospasm avoided or effectively overcome, most patients do well.

Arteriovenous Malformation

Arteriovenous malformations (AVMs) occur within the central nervous system as congenital abnormalities that allow blood to be shunted directly from arteries to veins, bypassing the normally interconnecting capillary bed. These malformations may be quite small with only a single feeding artery, or they may encompass several lobes of the brain and arterial feeders from multiple sources. They may occur in virtually any portion of the brain, including the cerebellum and brainstem. In the cerebral parenchyma, where they are most commonly located, the lesion takes on a conical shape, with the apex deep, often reaching the lateral ventricle. Rarely, AVMs occur within the spinal cord, and they may exclusively involve the dura either intracranially or within the spinal canal.

Patients with AVMs tend to develop symptoms before age thirty. The most common initial presentation is hemorrhage (50 percent of cases and 10 percent of all intracerebral hemorrhages, second only to aneurysms). Bleeding usually occurs within the brain substance but may occur within the ventricular system or the subarachnoid space. The patient experiences a sudden headache, often associated with loss of consciousness and/or a neurologic deficit. The next most common presenting symptom is a seizure. In a few cases, seizures may be frequent and refractory to medical therapy. AVMs may also present with the insidious onset of a focal neurologic deficit due to mass effect, increased venous pressure, or vascular steal phenomenon. Occasionally, young patients with severe unrelenting headaches are found to harbor an AVM.

From the time of discovery of an unruptured AVM, the risk of hemorrhage is about 1 percent per year cumulatively. Once an AVM has bled, the risk of rebleeding increases to 5 percent per year. With each hemorrhage, the risk of dying is approximately 10 percent and morbidity at least 15 percent. Smaller AVMs are more likely to bleed than larger ones. Most AVMs remain stable in size, but some enlarge with time. Up to 10 percent have an associated aneurysm on a feeding artery. In these cases, hemorrhage is usually due to rupture of the aneurysm.

An AVM can be identified on contrast-enhanced CT scanning as a hyperdense mass, part of which has a serpentine configuration related to the presence of large draining veins. Its configuration and extent are more easily delineated with MRI. After hemorrhage, an unenhanced CT scan usually demonstrates intracerebral or subarachnoid blood. AVMs may be too small to be seen on CT, so careful angiography may be necessary to identify the source of hemorrhage. Lumbar puncture may be necessary if subarachnoid hemorrhage is suspected clinically but not verified by the CT scan.

In all cases of suspected or proven AVM, complete cerebral angiography must be undertaken to define carefully the extent of the malformation (Fig. 40-17 A, B). All feeding arteries, including any from the external carotid system as well as the draining veins, must be evaluated. It is only from high-resolution angiography that a treatment decision can be reached. If angiography fails to delineate a lesion despite suspicion, MRI may identify the “angiographically occult” abnormality.

The treatment of AVMs is dependent on the size and location of the lesion, the presenting symptoms, and the age and condition of the patient. Because of the risk of rebleeding, an AVM that has bled should be surgically excised if possible. The treatment decision is more difficult in the patient who presents with seizures. If the patient is young and the malformation is readily accessible, surgical resection is usually recommended, especially when the seizures are medically refractory. Operation involves the microsurgical dissection and resection of the entire malformation, rather than simple ligation of feeding arteries. The results of operation are related to the size and location of the malformation. Overall, the operative mortality is less than 5 percent and the morbidity less than 10 percent.

Alternative or adjunctive methods of treatment include intraarterial embolization and radiation therapy. Using interventional neuroradiologic techniques, particulate matter or glues may be introduced into AVMs via feeding vessels to occlude the vascular shunt nidus. It is rarely possible to obliterate these lesions completely with this method, however. This technique nevertheless can reduce flow through the AVM prior to direct surgical intervention.

Ionizing radiation, on the other hand, has the capability of completely obliterating selected small- to medium-sized AVMs. Focused gamma or proton beam irradiation has demonstrated efficacy, and occasional success with conventional irradiation has been reported. Ionizing radiation causes endothelial proliferation and may take 6 months to 2 years to obliterate the lesion. Focused irradiation is recommended for deep, surgically inaccessible AVMs.

Intracerebral Hemorrhage

Spontaneous hemorrhage is most commonly associated with systemic hypertension and occurs in predictable locations, including the putamen, thalamus, cerebellum, and pons. Hemorrhage can also occur within the various lobes of the brain. Nonthypertensive causes of brain hemorrhage have been discussed above, such as rupture of AVMs and aneurysms and hemorrhage into areas of ischemia. Additional causes include induced or endogenous coagulopathies, primary or metastatic brain tumors, and rare conditions such as amyloid angiopathy.

Chronic hypertension results in lipohyalinosis of the vessel wall, which sets the stage for either vascular occlusion or rupture. Occlusion results in infarction, and rupture produces an intracerebral hemorrhage. The shorter penetrating arteries of the brain appear to be the most vulnerable. The lenticulostriate and thalamoperforating vessels are involved in putaminal and thalamic hemorrhages, and affected basilar branches contribute to pontine hemorrhage.

Although brain hemorrhage is often devastating, it may be surprisingly well tolerated. The hematoma tends to dissect along axonal planes, separating rather than destroying vital structures. If the resultant mass is tolerated by the patient, the blood is slowly resorbed by macrophages along the periphery, leaving only a hemosiderin-stained slit in the brain. Patients may worsen clinically anytime after the initial hemorrhage as a result of associated edema formation.

Hemorrhage into the putamen accounts for the majority of hypertensive hematomas. Presentation is characterized by the lack of headache with gradual development of hemiparesis progressing to hemiplegia. This may be associated with a hemisensory loss, aphasia, hemianopia, and/or ipsilateral deviation of the eyes, depending on the size of the hematoma and its direction of dissection. The patient may, of course, progress into coma if the lesion is large. Similarly, thalamic hemorrhage presents initially with a hemisensory loss and hemiparesis. Localizing features include downward eye deviation with limitation of vertical gaze and small, sluggish pupils due to involvement of the nearby mesencephalon. Headache is uncommon.

Cerebellar hemorrhage is sudden in onset and presents with headache. Vomiting, ataxia, and dizziness are accompanying features. This hemorrhage is extremely dangerous in that it may cause coma and ultimately death due to brainstem compression and acute hydrocephalus. Brainstem hemorrhage (usually pontine) is the most devastating and often presents with quadriplegia, decerebrate posturing, pinpoint pupils, and coma. Most patients do not survive if the hematoma is larger than 1 cm in size. Moreover, those who do survive have a high degree of morbidity.

Lobar hemorrhage is less likely to be associated with hypertension and, in general, is better tolerated by the patient. The symptoms depend on the area of brain involved.

CT scanning has become an invaluable tool in diagnosing and defining brain hemorrhage. CT not only delineates the hemorrhage but also permits assessment of ventricular size, the presence of edema, and often the cause of the hemorrhage (e.g., AVM, tumor, aneurysm). The hematoma appears hyperdense in the acute phase (Fig. 40-18). With time, as the blood breaks down, the clot progresses to a hypodense lesion. At any time, from days to weeks, the hematoma may demonstrate an enhancing ring. If a vascular lesion is suspected, careful angiography is indicated. In all cases, appropriate coagulation studies should be obtained.

The treatment of brain hemorrhage may be medical or surgical depending on the size of the lesion, its location, and the condition of the patient. Surgical resection is recommended if the patient is deteriorating neurologically, no matter what the size of the hematoma. Cerebellar hematomas are particularly important to remove, since a small change in surrounding reactive edema may result in life-threatening brainstem compression and/or hydrocephalus.

Because hematomas are mass lesions, medical management is directed toward keeping the intracranial pressure under control. If the hemorrhage renders the patient unconscious, hyperventilation and hyperosmolar agents may be required to control ICP. ICP monitoring may be a helpful adjunct to direct treatment. Steroids are useful in controlling brain edema if the patient has bled into a tumor, and coagulopathies should be corrected. Despite medical and/or surgical therapy, mortality and morbidity remain high from all types of brain hemorrhage.

DEGENERATIVE SPINE DISEASE

Anatomy and Pathophysiology

The spinal column is composed of 33 vertebrae making up those divisions labeled cervical (7), thoracic (12), lumbar (5), sacral (5 fused), and coccygeal (4 fused). Each vertebra consists of a body, which bears weight, and the posterior elements (pedicles, laminae, spinous and transverse processes), which provide the flexibility and stability to the vertebral column. The spinal canal has an ovoid shape in the transverse plane through the cervical and thoracic regions and assumes a more triangular shape in the lumbar region. Most spine movement occurs in the cervical and lumbar regions. Flexion and extension are greatest in the lower cervical and lumbar segments, and maximum rotation occurs predominantly in the upper cervical and lumbar segments.

The intervertebral disc consists of two parts. The circumferential anulus fibrosus, making up the outer portion, is composed of dense, fibrous tissue. The central nucleus pulposus consists of fibrocartilage, which has little tensile strength but substantial elasticity. The fibrocartilage may fragment acutely or degenerate gradually. It heals poorly because of limited blood supply. The anulus heals well and is buttressed by heavy anterior and posterior longitudinal ligaments. Intervertebral disc disease may occur at any level from C1 to S1. The lower segments of the cervical and lumbar areas are affected most often. Thoracic disc disease is rare.

The spinal cord extends from the cervicomedullary junction at the base of the skull to the conus medullaris at the L1–L2 level. The cord is centrally placed within the spinal canal and moves rostrally and caudally a few millimeters during spinal flexion and extension. Lateral motion of the cord is restricted by intradural dentate ligaments. The blood supply is provided by radicular arteries, which arise from the vertebral arteries and the thyrocervical trunks in the neck, from the intercostal arteries in the thorax, and from the lumbar arteries in the low back. An arterial confluens, the artery of Adamkiewicz, is typically found in the T10–L2 region, usually on the left side. It supplies the lower thoracic cord and conus medullaris.

Three fiber tracts of the spinal cord are important clinically (see Fig. 40-3). The laterally positioned *corticospinal tracts* carry motor fibers from the cortical upper motor neurons to the

spinal lower motor neurons located in the ventral horns of the spinal cord. These tracts cross the midline at the pyramidal decussation in the lower medulla. The *spinothalamic tract*, also positioned laterally, transmits pain and temperature sensation from the contralateral side of the body. Its axons cross through the anterior commissure of the cord within two or three segments of each dorsal root entry zone and ascend to the ipsilateral thalamus. The *dorsal columns* carry sensory fibers conveying position, vibratory, and light touch sensation from the dorsal roots rostrally on the same side, then cross to the opposite cerebral cortex through a decussation in the brainstem.

Dorsal and ventral nerve roots emerge from the spinal cord separately and pass to their respective intervertebral foramina, where they exit from the spinal canal (see Fig. [40-2](#)). The roots join to form a spinal nerve within the neural foramen. In the cervical spine, the roots exit above the corresponding vertebrae; for instance, the C5 root exits above the C5 pedicle. Since there are eight cervical roots, C7 exits above the C7 pedicle and C8 exits below it. Consequently, all roots below C8 exit below the pedicle of their corresponding vertebra.

Lumbar and sacral roots form the cauda equina below the conus medullaris. The sacral roots are more centrally located adjacent to the filum terminale. Because a lumbar root (e.g., L4) passes laterally toward the neural foramen as it descends within the spinal canal, it crosses the adjacent intervertebral disc (e.g., L4–L5) at its extreme lateral edge, hugging the pedicle of the L4 vertebra laterally. The nerve root that descends to the next lowest foramen (e.g., L5) passes across the disc space (e.g., L4–L5) more medially, making that root more vulnerable to disease involving that disc (Fig. [40-19](#)).

Intervertebral Disc Disease

If the nucleus of an intervertebral disc extrudes (herniates) through the anulus, adjacent neural structures may be compressed. In the cervical and thoracic spine, compression of the spinal cord may result in paraparesis or quadriparesis, depending on the spinal segment involved. At all levels, compression of a spinal root may cause weakness and sensory loss in structures innervated by that root. The severity of the clinical syndrome depends on the site and severity of compression by the displaced disc fragment. In some cases the anulus and adjacent ligament hold, preventing complete extrusion of the fragmented disc. The anulus may only stretch sufficiently to allow the disc to bulge into the spinal canal or foramina, accounting for back and leg pain, but often without neurologic deficit.

Often the nucleus does not extrude, but simply fragments in response to the forces exerted on the spinal column. This is intensified by the concomitant dehydration and loss of elasticity of the disc as it ages. The disc space gradually narrows, the joint becomes loose, and the cartilaginous endplates of the adjacent vertebral bodies abut and wear more quickly. Bony spurs (osteophytes) develop at the joint in reaction to the increased mobility and decreased elasticity.

Formation of osteophytes around the joints of vertebrae, termed spondylosis, is a common disorder that represents the normal process of aging. If an osteophyte forms in a neural foramen, the nerve root passing through may be chronically irritated and compressed. If the osteophyte develops within the cervical or lumbar canal, the cord or cauda equina may be compromised.

The onset of symptoms and signs of an extruded disc fragment may be acute or chronic. Acute symptoms may or may not be related to trauma. In disc disease of the cervical spine, neck and radicular discomfort occur simultaneously. Spinal cord symptoms are rare. There is usually limitation of neck motion, with loss of normal cervical lordosis. With foraminal osteophytes, episodes of cervical discomfort recur over many months or years before radicular symptoms appear. Interscapular aching, suboccipital headaches, and even chest pain are common complaints.

Nerve root compression produces radiculopathy, often characterized by pain and hypoesthesia in the distribution of the involved root. Associated loss of deep tendon reflex with or without weakness may be seen on examination. Cervical cord compression causes myelopathy characterized by progressive spastic quadriparesis or paraparesis, mild to moderate sensory changes in the lower extremities and trunk with cervical dermatomal sensory loss, weak upper extremities, hyperreflexia, and extensor plantar response.

Cervical Disc Disease

Cervical disc disease must be differentiated from other ailments. These include inflammatory disease of the soft tissues and joints of the arm and shoulder, nerve entrapment syndromes, and neoplasms. The pain must be distinguished from that which accompanies cardiac disease. Spinal infections, congenital lesions, and posttraumatic disorders are other important considerations.

Plain radiographs typically demonstrate loss of the lordotic curve of the cervical spine, with narrowing of one or more disc spaces. Osteophyte formation may be seen. In cervical spondylosis, there is usually radiologic evidence of osteophytes and disc space narrowing at multiple levels. In most cases, the anterior-posterior diameter of the cervical spinal canal is narrowed. Myelography with CT is very useful in the diagnostic workup of nerve root compression. The use of intrathecal contrast medium enhances the power of CT to delineate the lesion. MRI is suitable for investigating myelopathies. In addition to defining the compressive lesion, MRI often shows intrinsic cord abnormalities related to compression. Electromyography may confirm the diagnosis and localize the lesion more specifically, particularly when myelographic defects are multiple.

Treatment

Painful cervical disc disease may be treated medically as long as there is no evidence of a progressive neurologic deficit (motor loss and bowel and bladder dysfunction being most important). Adequate medical therapy includes immobilization of the neck with a soft or hard cervical collar, analgesics, muscle relaxants, and local heat. These methods, in association with a good physical therapy program, provide relief under most circumstances.

Up to 75 percent of patients with cervical disc disease improve following an adequate trial (10 to 14 days) of medical therapy. Some have recurrence of radicular symptoms on return to full activity. In many cases, these patients can be managed for years with intermittent cervical traction and a cervical collar, but some require surgical therapy. For the 25 percent who do not respond to conservative means, operation is often helpful.

There are two approaches for the surgical treatment of cervical disc disease. Anteriorly, nerve roots, spinal cord, or both may be decompressed through discectomy with or without bone graft fusion. The other approach is posteriorly through a laminectomy and/or foraminotomy. The choice of operative direction is based on consideration of the patient's anatomic lesion. Improvement follows operative treatment of symptomatic cervical disc disease by either approach in approximately 80 percent of patients who fail to respond to medical treatment. Surgical treatment of cervical spondylotic myelopathy results in improvement in most cases. Arrest of the progressive myelopathic deficit usually occurs.

Lumbar Disc Disease

Herniated lumbar discs often produce some degree of nerve root compression. The severity of the syndrome depends on the degree of root compression. Occasionally, the entire cauda equina may be involved, resulting in loss of motor and sensory function, including bowel and bladder sphincter control. Sometimes disc rupture may occur in the midline, compressing centrally positioned sacral roots preferentially, without involvement of laterally placed lumbar roots.

Fragmentation of a lumbar disc may occur without extrusion of the nucleus pulposus as described above for cervical disc disease. Because of loss of elasticity within the disc, mobility of the intervertebral joint is increased. The anulus fibrosus may simply bulge without tearing. With time, osteophytes may form around the degenerated disc and encroach on the spinal canal and neural foramina. This degenerative hypertrophy involves the ligamentous structures as well. Stenosis of the lumbar spinal canal is the eventual result, a spondylotic condition common in the elderly.

In the lumbar spine, over 90 percent of clinical problems arise from the L4–L5 and L5–S1 intervertebral discs. Pain is usually chronic, but its onset may be acute when associated with frank herniation. There may be back pain, leg pain, or both. Radiation of low back pain into the buttock, posterior thigh, and calf is usually the same with disease at the L4–L5 and L5–S1 levels. This radiating pain may be exacerbated by coughing, sneezing, or straining. Bending and sitting accentuate the discomfort, while lying down characteristically relieves it. Pain is typically described as aching but frequently has a sharp or shooting element and is limited to one lower extremity. With lumbar stenosis, patients are unable to extend their spine without developing pain, numbness, and/or weakness, usually in both lower extremities. In the upright posture, either standing or walking, the cauda equina becomes relatively ischemic, producing neurogenic claudication. Relief is obtained by sitting or flexing forward (Fig. [40-20](#)).

Palpation usually reveals tenderness over the sciatic notch, the popliteal fossa, or both. Paravertebral muscles may be in spasm. With true nerve root compression, straight-leg raising produces leg pain that is accentuated by dorsiflexion of the foot. Ipsilateral leg pain produced by contralateral straight-leg raising is highly suggestive of lumbar disc herniation. Sensory loss, weakness, and loss of tendon jerks may occur in a variety of combinations and to variable degrees.

Back pain with radiation to the leg has many causes besides lumbar disc disease. The differential diagnosis includes bony abnormalities such as subluxation, degenerative facet fracture, and osteophyte formation; primary and metastatic tumors of the cauda equina, spine, and pelvis; inflammatory disorders, including abscess, arachnoiditis, ankylosing spondylitis, and rheumatoid arthritis; degenerative lesions of the spinal cord; peripheral neuropathies; peripheral vascular occlusive disease including abdominal aortic aneurysm; and gynecologic problems such as endometriosis.

Plain films of the lumbosacral spine can identify congenital or acquired bony changes. Disc space narrowing is an unreliable sign of symptomatic disease, since narrowing of the disc space may occur without clinical symptoms. Flexion and extension lateral views reveal concomitant instability. Myelography can be diagnostic in symptomatic lumbar disc disease, but CT alone can delineate the lesion in most cases (Fig. [40-21](#)). MRI has replaced myelography and CT at some centers in the workup of lumbar radiculopathy. Electromyography may confirm the diagnosis, especially when physical examination is unable to localize the involved nerve root.

Treatment

Initially medical treatment is indicated in all patients who do not have neurologic deterioration. Bed rest, local heat, analgesics, and skeletal muscle relaxants are usually effective within a few days. Physical therapy and limited exercise often help when the acute episode passes. A back brace partially immobilizes the patient and can minimize muscle spasm. With aggressive conservative management, most patients improve sufficiently to return to full activity. Recurrent symptoms may be treated in a similar fashion, often successfully. Surgical treatment is reserved for the patient with an acute or progressive neurologic deficit, chronic disabling pain, or both. The acute onset of weakness or sphincter disturbance constitutes an emergency, demanding prompt diagnosis and early operation.

Operation usually entails a unilateral laminotomy with removal of the offending disc fragment. Foraminotomy may be necessary in the presence of osteophyte formation. With lumbar stenosis, multilevel laminectomy is curative. Should plain films demonstrate any instability preoperatively, combining the laminectomy with posterior fusion, either with or without instrumentation, is generally indicated. If the imaging studies demonstrate an extruded disc fragment that accounts for the clinical signs and symptoms, 85 to 90 percent of patients recover with surgical treatment. If the syndrome is atypical, the myelogram equivocal, and the patient poorly motivated, operation is less effective. Emotional factors, psychologic disturbances, litigation, and industrial injury play an important role in the eventual outcome, whether the treatment is medical or surgical. Alternatives to the laminotomy approach include percutaneous discectomy done by endoscopic excision of the problematic nucleus pulposus. Clinical results for the alternative approaches are less predictable.

INFECTIONS

The central nervous system may be infected by viruses, bacteria, fungi, and parasites. Development of infection depends on the host's resistance (i.e., immune defenses) and on the infecting agent's virulence. Bone, brain, spinal cord, meninges, and cerebrospinal fluid may be involved separately or in combination. The routes of infection include hematogenous dissemination, local extension from a neighboring source, and direct contamination through an open wound. The infection may be diffuse, as in meningitis, or focal, as in brain abscess.

The clinical spectrum of signs and symptoms of CNS infection varies from nonspecific (such as fever, confusion, and lethargy) to highly specific (such as jacksonian epilepsy and focal neurologic deficits). Consequently, CNS infections present difficult diagnostic and therapeutic problems. Early diagnosis and treatment are critical to achieve a successful outcome.

Bacterial

Subgaleal Abscess

Localized infection between the galea of the scalp and the pericranium constitutes subgaleal abscess. Usually, the process is initiated by contamination of an open scalp wound by staphylococci, streptococci, or anaerobic cocci. Localized scalp tenderness, warmth, and swelling are signs of abscess formation. Osteomyelitis of the skull may occur secondarily. Subgaleal infections rarely extend intracranially, unless the skull has been penetrated. Treatment includes open drainage, debridement, and systemic antibiotics.

Osteomyelitis

Osteomyelitis may develop from extension of a localized infection, such as sinusitis or mastoiditis, from direct contamination at operation or after trauma, or, rarely, by hematogenous spread from a distant source such as the respiratory or urinary tract. An established skull or spine infection may extend to the epidural space, producing a localized abscess (Fig. 40-22). The usual osteomyelitis pathogens are staphylococci and anaerobic streptococci. Occasionally, gram-negative organisms and fungi are responsible. Treatment consists of drainage, debridement of infected bone, and appropriate antibiotics for a prolonged period, usually 6 weeks.

Epidural Abscess

Spinal epidural abscess is much more common than intracranial epidural abscess. It is characterized by fever, local spinal tenderness, and rapid progression of neurologic deficits, often constituting a medical and surgical emergency. Radicular pain and impairment of cord function, with early motor and sensory deficits including sphincter disturbances, occur within a few days. Most epidural abscesses are caused by local extension of osteomyelitis or by hematogenous spread from a distant suppurative focus. The diagnosis is suggested by the clinical presentation. The CSF often has a markedly elevated protein level with mild pleocytosis. MRI defines the extent of the epidural mass. If the dura is intact, infection rarely extends across it.

The most common causative organisms are *Staphylococcus aureus* and the streptococci. Treatment should be immediate, beginning with broad antibiotic coverage until the offending agent is identified. Specific antibiotic therapy should be continued for a prolonged period, often up to 6 weeks. Surgical drainage is necessary when neurologic deficits progress despite aggressive medical therapy. Corticosteroids in the perioperative period are beneficial in reducing localized edema, although prolonged use can reduce the host immune response to the infection. Recovery of neurologic function is directly related to the duration and severity of impairment before treatment.

Subdural Empyema

Subdural empyema is a purulent infection of the subdural space. It accounts for approximately 25 percent of all intracranial infections and is usually a complication of sinusitis, meningitis, or open contamination of the subdural space at operation or after trauma.

With sinusitis, infection can spread intracranially by transcranial emissary vein thrombophlebitis. Staphylococci, streptococci, and anaerobic cocci are commonly responsible. Once the subdural space is violated, infection can spread over the convexity of the brain. The accumulation of purulent material may be sufficient to produce an intracranial mass, provoking adjacent brain swelling. The clinical result is rapid neurologic deterioration, often with lateralizing signs, coma, and death. Treatment includes craniotomy with debridement, drainage, and intravenous antibiotics. The source of the infection must be treated aggressively. A sinus or mastoid drainage procedure is often required if this is the source. Mortality from acute fulminant subdural empyema from a paranasal source remains about 25 percent.

The diagnosis of intracranial subdural empyema is made readily by CT or MRI scan, but it may be difficult to distinguish from subacute or chronic subdural hematoma. The mass itself may be isodense, necessitating the administration of intravenous contrast. Including the sinuses on the scan may demonstrate the source of infection. Lumbar puncture to obtain CSF for analysis risks transtentorial herniation; thus LP should be avoided if the scan shows significant mass effect.

Spinal subdural empyema is rare. It usually develops from local extension transdurally or through the arachnoid in the presence of meningitis. Spinal cord compression and transverse myelitis may develop. Treatment is emergent, consisting of surgical drainage and prolonged antibiotic administration.

Meningitis

Bacterial meningitis is an acute, purulent infection of the leptomeninges. It is manifested by fever, lethargy, headache, nausea, vomiting, and nuchal rigidity. Seizures occur in approximately 20 percent of patients and cranial nerve palsies in about 5 percent. Coma may develop in up to 10 percent of patients with missed diagnoses, heralding a poor prognosis. Untreated bacterial meningitis is almost always fatal.

A CSF Gram stain may demonstrate the offending organism in 75 percent of cases. Cultures provide a diagnosis 90 percent of the time. When CSF cultures are negative despite high clinical suspicion, as in a mild case or in an incompletely treated case of meningitis, latex agglutination studies are helpful. These immunologic studies are specifically directed at *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis* and are highly sensitive when an organism is present, even with negative cultures. Blood cultures may be positive and thereby helpful in the diagnosis, particularly with infections caused by *S. pneumoniae* and *N. meningitidis*. CSF pleocytosis with a preponderance of polymorphonuclear cells is typical of untreated bacterial meningitis. CSF glucose level is almost always reduced, and the protein content is typically increased.

Meningitis that develops after a penetrating wound or a neurosurgical procedure is usually caused by staphylococcal, streptococcal, or gram-negative organisms. Meningitis occurring after closed head trauma with either a skull fracture or CSF rhinorrhea is most often caused by *S. pneumoniae*. Ventricular shunt and reservoir infections leading to meningitis are more likely due to *Staphylococcus epidermidis* or *aureus*.

The treatment for acute bacterial meningitis depends on the causative organism, its antibiotic sensitivity, and the primary source of infection from which the meninges were contaminated. The presumed diagnosis is made clinically, a sample of CSF is obtained via lumbar puncture, and broad-spectrum intravenous antibiotics are immediately started. Once culture results are available, the choice of antibiotics is changed to an appropriate single agent. Bacterial endocarditis, pneumonia, sinusitis, concurrent subdural empyema, and brain abscess are sometimes associated with meningitis. Treatment should be directed at both the meningitis and the primary source.

The extent of antibiotic penetration into the CNS varies, depending on the degree of meningeal inflammation. Intrathecal administration of those antibiotics that do not readily cross the blood-CSF barrier may be necessary (the commonly used intrathecal preparations are gentamicin and vancomycin). This is especially true when a foreign body, such as a ventricular shunt, is present. Shunt removal is often necessary despite intravenous and intrathecal antibiotic administration. Rarely, therapy may also include the use of steroids and/or osmotic diuretics if intracranial pressure is elevated as a result of cerebral edema or localized brain abscess.

Complications of bacterial meningitis include communicating hydrocephalus, brain abscess, subdural empyema, and subdural effusions, particularly after *H. influenzae* meningitis in infants. The risk of complications is significantly reduced by prompt, early treatment.

Brain Abscess

Brain abscess is a purulent lesion of brain tissue, beginning as a focal infection, usually in the white matter, surrounded by a typical inflammatory response. The blood-brain barrier becomes disrupted. Necrosis and liquefaction follow the acute inflammatory stage. Eventually, either the process is encapsulated by fibrous granulation tissue or the infection spreads through the parenchyma to the subarachnoid spaces and the ventricular system.

Brain abscess is usually secondary to focal infection elsewhere. Abscesses that develop by direct intracranial extension are usually solitary and are typically found in the frontal and temporal lobes near adjacent nasal sinuses or mastoid processes where the infection began (Table 40-4). Multiple brain abscesses that develop in the septic patient are often related to bacterial endocarditis, pneumonia, and diverticulitis. Cyanotic congenital heart disease with concurrent infection is a frequent source. Direct contamination of the brain through a penetrating wound, especially when accompanied by in-driven bone fragments, is another cause of abscess. Abscess formation is frequent among patients with compromised immunity either from an underlying illness such as HIV infection or during pharmacologic immunosuppression such as in organ transplantation.

Signs and symptoms of brain abscess are related to its mass effect. Headache, focal, neurologic deficits, and impaired mentation are often noted. There may be little or no evidence of systemic infection, and the patient may be afebrile. Conversely, the patient may be moribund from bacteremia with fever, hypotension, and a markedly elevated white blood cell count. Seizures may occur. Progressive mass effect leads to brain shifts followed by coma.

Contrast CT and MRI are highly accurate in detecting brain abscess and should be done before CSF is sampled (Fig. 40-23). The CSF of patients with brain abscess may be entirely normal, but usually some pleocytosis is noted. The causative organism can be identified and cultured from the abscess itself in 60 to 80 percent of cases, provided cultures are processed carefully for both aerobic and anaerobic organisms. Blood cultures are also helpful, particularly if the abscess is secondary to systemic infection.

In certain cases of early abscess formation or high surgical risk, medical therapy alone with appropriate parenteral antibiotics may be sufficient. The most effective therapy, however, is drainage of purulent material with simultaneous administration of appropriate intravenous antibiotics. Although needle aspiration may be successful, craniotomy with evacuation and removal of the abscess wall may be necessary. Surgical drainage reduces the mass effect, thereby reducing the most critical and dangerous aspect of the infection, and allows accurate bacteriologic analysis.

Results of treatment for brain abscess depend on the patient's neurologic status initially, the efficacy of the antibiotic used, the extent to which the intracranial mass is controlled by surgery, and the effective treatment of the primary source of the abscess. Despite aggressive surgical and medical management, mortality rates associated with brain abscess approach 40 percent, especially in the malnourished, chronically debilitated, or immunosuppressed patient.

Postoperative Infection

Any or all of the pyogenic infections described earlier may develop after operation. Once identified, characterized, and treated with appropriate antibiotics, the infection will almost always subside. Commonly isolated organisms include *Staphylococcus aureus* and *epidermidis*. If a foreign body such as prosthetic material or a ventricular shunt is involved, eradicating the infection becomes more difficult, often requiring a combination of intravenous and intrathecal antibiotics and removal of the foreign material. Occasionally, infections can be treated satisfactorily in the presence of retained foreign bodies, such as a shunt, provided the infection is indolent.

Fungal

Fungi may become pathogenic as a result of depression of the host immune system, prolonged systemic antibiotic therapy, or severe systemic illness. When the CNS becomes infected, it is usually associated with pulmonary fungal infection and depressed host resistance. The CNS involvement may be a diffuse meningitis or a focal abscess. Multiple abscesses may be present. Treatment requires long-term systemic antifungal chemotherapy. Surgical intervention is reserved for drainage of abscesses and resection of symptomatic mass lesions. Hydrocephalus, a potential late complication, is treated with a ventricular shunt.

Parasitic

While relatively uncommon in North America and western Europe, parasitic diseases of the CNS are a major cause of neurologic disability and death worldwide. Control of these diseases remains a public health problem. A major emphasis is placed on their prevention, for once the CNS is infested, therapeutic options are limited. Treatment, both medical and surgical, is usually ineffective or palliative at best.

Cysticercosis

Taenia solium, the pork tapeworm, infests the human CNS by transmission of its larvae through the blood following ingestion. It is most prevalent in eastern Europe, Latin America, China, Pakistan, and India. Its presence may take one or all of four forms. Meningeal cysticercosis is characterized by parasitic vesicles throughout the basal cisterns and CSF pathways, usually with resultant hydrocephalus. Parenchymal cysticercosis diffusely involves the brain, sometimes forming large cysts. Seizures and focal deficits are common. The ventricular variety resembles the meningeal form. Obstructive hydrocephalus is commonplace. Spinal cysticercosis may be intramedullary or extramedullary, producing either a transverse myelitis or a compressive myelopathy.

The diagnosis rests on serologic and radiologic testing. The presence of intracranial cysts and calcifications within skeletal muscle is often presumptive of the diagnosis. Praziquantel, an anthelmintic agent, is effective medical therapy for systemic infestation. Anticonvulsants, CSF shunting, and occasional removal of symptomatic cysts are additional treatment options.

Echinococcosis

Hydatid disease is caused by *Echinococcus granulosus*, the dog tapeworm. It is prevalent in southern South America, northern and eastern Europe, Australia, Africa, China, and the Middle East. Humans may serve as intermediate hosts by ingesting the larvae. The liver and lungs are preferentially involved through hematogenous dissemination with subsequent formation of hydatid cysts. When the CNS is involved, cysts are usually solitary, large, and confined to white matter. There is a negligible inflammatory response.

Most cysts produce signs and symptoms related to their mass effect. Diagnosis of the infection is made serologically. CT and MRI of the brain and ultrasonography of the liver and spleen may be definitive. Chest x-ray often shows calcified pulmonary cysts. Treatment consists of patient isolation from the source and surgical removal of symptomatic cysts. Care must be taken to remove the intact cyst to avoid seeding with viable larvae. Hydatid disease of the CNS is disabling, but rarely fatal, provided cysts are removed when they become symptomatic.

CONGENITAL AND DEVELOPMENTAL ABNORMALITIES

Approximately 2 percent of newborns possess some type of congenital abnormality. Sixty percent of these involve the central nervous system, and over half of those are related to defective development or closure of the dorsal midline structures. Many have associated hydrocephalus. The commonly encountered neurologic malformations are listed in Table 40-5.

Spinal Dysraphism

Between 18 and 28 days of embryonic development, the neural groove closes posteriorly in the midline to form the neural tube. This tube is encircled by bone derived from adjacent somites and is covered superficially by skin derived from ectoderm. Abnormal closure of the neural groove, failure of fusion of the adjacent bone, and/or maldevelopment of the overlying ectoderm can lead to a variety of spinal dysraphic states. Thus, dysraphism implies an abnormal fusion of normally united parts.

Failure of the bony structures to close with normal closure of the neural groove is called spina bifida occulta. Patients with this anomaly have a normal spinal cord and normal cord function. The abnormality usually is unnoticed unless seen on plain radiographs. If the meninges fail to close, a meningocele develops, producing a cutaneous abnormality. The underlying neural structures, however, develop normally, so there is no compromise of neurologic function.

Failure of the underlying neural tissue to fuse has been called spina bifida cystica, or more recently, spina bifida aperta. Myelomeningocele, the more common form, involves incomplete closure of the neural groove, usually in the lumbar region, with the abnormal, unfused neural tissue on the dorsal surface, exposed through an associated defect in the spinal column. This may be partially or totally covered with epithelium. The accompanying neurologic deficit usually consists of complete absence of motor and sensory function below the level of spinal cord involvement.

The most severe form of spinal dysraphism is myeloschisis, which is much less common than myelomeningocele. The spinal cord is unfused and presents directly on the surface of the back without overlying meninges or epithelium. It usually occurs at the thoracolumbar region and is virtually always associated with paraplegia and absence of bladder function.

Both myelomeningocele and myeloschisis are associated with hydrocephalus. Hydrocephalus is caused by a developmental abnormality of the hindbrain called Arnold-Chiari malformation, which is associated with the more severe forms of spinal dysraphism. This malformation is composed of caudal displacement of the cerebellar tonsils, vermis, inferior fourth ventricle, and medulla. There is a dorsal kink in the cervicomedullary junction and breaking of the quadrigeminal plate. Associated anomalies include agenesis of the corpus callosum and obstructive hydrocephalus.

The treatment of spinal dysraphism is surgical. Meningoceles are excised and the skin is closed primarily after watertight closure of the posterior meningeal defect. Myelomeningoceles and myeloschises are closed as early as possible to reduce the risk of superficial infection and subsequent meningitis. The goal is to preserve as much neural tissue as possible, untether the spinal cord from surrounding soft tissue, and fashion a dural closure to prevent CSF leakage. Accompanying hydrocephalus is treated by shunting.

Survival of infants with these dysraphic states continues to improve. Those newborns with lower-level lesions have better outcomes than those with higher lesions overall. The more severe the dysraphic state, the higher the morbidity and mortality. Risk of sepsis from bladder infection is reduced with intermittent catheterization when indicated. Timely revision of failed shunts placed for hydrocephalus preserves potential for intellectual development.

Cranial Dysraphism

Cranial dysraphic states are one-tenth as common as their spinal counterparts. Encephaloceles, although rare, are the most common manifestation. They consist of a midline skull defect through which a small portion of brain protrudes. Most encephaloceles are covered with skin, and only some 35 percent have associated hydrocephalus. Once believed to arise from defects in the closure of the primitive neural tube, they probably develop because of an overlying mesodermal abnormality, with subsequent perturbation of underlying cerebral tissue. In North America and Europe, 70 percent of encephaloceles occur in the posterior cranial vault; the remainder are found in the anterior cranial vault. In southeast Asia, this distribution is reversed for unknown reasons.

Surgical repair involves early resection of malformed and devitalized brain and dural closure. The mortality of patients with encephaloceles is extremely variable. Prognostic factors include the size and location of the anomaly, the extent of brain protrusion, and the presence of associated hydrocephalus, seizure disorder, or cerebral dysgenesis. The smaller and more anterior defects generally have a better outcome. Of those who survive, only 35 percent attain normal intelligence.

Hydrocephalus

The term hydrocephalus implies an increase in the amount of CSF within the ventricular system (Fig. 40-24). This is almost always due to a decrease in the absorption of fluid, although there are rare cases of choroid plexus papillomas causing hydrocephalus by an increase in CSF production. Hydrocephalus is traditionally classified as communicating and noncommunicating. In the former, the ventricular system continues to communicate with the subarachnoid spaces outside the brain through the fourth ventricular foramina of Luschka and Magendie. In the noncommunicating variety (often termed obstructive), it does not. The common causes of hydrocephalus vary with age and are listed in Table 40-6.

Infantile Hydrocephalus

Hydrocephalus occurs most frequently between birth and 2 years of age and is most commonly due to congenital abnormalities of the brain. These abnormalities typically produce noncommunicating hydrocephalus. Stenosis of the cerebral aqueduct is one such common congenital anomaly. Another is the Arnold-Chiari malformation, described earlier. The Dandy-Walker malformation produces a markedly enlarged fourth ventricle as a result of congenital obstruction of CSF outflow from the fourth ventricle, with resultant hydrocephalus. Other, less common congenital lesions include arachnoid cysts, vascular anomalies, and congenital tumors.

Acquired hydrocephalus in the infant is often the result of meningitis or intracranial hemorrhage, both potentially causing obstruction of either the CSF absorptive mechanism or the intraventricular pathways. Aqueductal stenosis may develop well after birth because of infection or hemorrhage and thus be considered "acquired." Tumors may also obstruct the outflow of CSF, resulting in noncommunicating hydrocephalus.

Infants with hydrocephalus usually, but not invariably, present with an enlarging head circumference. They often have a tense, bulging anterior fontanelle with distended scalp veins and split cranial sutures. They may appear to have "sun setting" of the eyes, with only the tops of the irises visible (Parinaud's syndrome). The head may transilluminate because of a lack of cerebral substance. Hydrocephalus usually does not impair the infant initially because the open cranial sutures allow for cranial vault expansion. In the more chronic forms, or in older infants with closed sutures, papilledema, optic atrophy, and VIth nerve palsies may be seen.

Childhood Hydrocephalus

Hydrocephalus in children over 2 years of age may have a more acute presentation because of the decreased ability of the more mature brain and skull to accommodate the increase in CSF. Consequently, raised intracranial pressure may cause headache, nausea, vomiting, lethargy, coma, and even death. Slower onset may result in decreased mentation, behavioral changes, diminished performance in school, VIth nerve palsies, optic atrophy, paralysis of upward gaze, spastic leg weakness, and endocrine (hypothalamic) disorders. Causes of hydrocephalus in this age group include tumors, meningitis, intracranial hemorrhage (both spontaneous and traumatic), and aqueductal stenosis. Ventricular shunt malfunction can cause acute hydrocephalus in the shunt-dependent patient, regardless of the patient's age or the underlying cause of the hydrocephalus.

Adult Hydrocephalus

Hydrocephalus in adults may also result from obstructive tumors, meningitis, and intracranial hemorrhage, but it may also be more insidious in onset. An entity called normal pressure hydrocephalus occurs in the older population, involving a communicating hydrocephalus with relatively normal intraventricular pressure. The cause remains unknown but is thought to be due to subclinical hemorrhage or infection in the patient's remote past and malabsorption of CSF. The classic symptom triad of ataxia, urinary incontinence, and failing mentation suggests the diagnosis. Treatment is by ventricular fluid shunting.

Regardless of the cause, the treatment of hydrocephalus is essentially the same. Either the cause must be removed (e.g., tumor) or a shunting procedure must be performed to divert accumulated CSF. Sometimes both measures are necessary. The most commonly used procedure is a lateral ventricle to peritoneal cavity shunt with a one-way pressure-regulating valve in the system. If the peritoneal cavity is not suitable for shunting, the distal catheter may be placed in the right atrium of the heart or in the pleural cavity. In selected cases of communicating hydrocephalus, a lumbar subarachnoid to peritoneal shunt may be used. Common complications of indwelling shunts include shunt obstruction and infection.

Craniosynostosis

Craniosynostosis is the premature closure of one or more cranial sutures, typically manifested within the first 6 months of life. Since the brain doubles in size during the first 6 months of life and grows another 50 percent by age two, the cranial sutures must remain open to allow for skull expansion to accommodate this growth. Usually, when one suture fuses prematurely, the brain is not compressed significantly, but the skull will then develop in a distinctly abnormal shape. If more than one suture is fused, brain growth may be restricted.

The sagittal suture fuses prematurely most often. The skull then develops an elongated shape with a narrow biparietal diameter, often referred to as scaphocephaly. The supraorbital ridge may be square-shaped due to overexpansion of the open metopic suture. Associated congenital anomalies are rare.

The next most common suture to be involved is the coronal, which may close prematurely on one or both sides. Unilateral involvement produces an asymmetrically shaped forehead with flattening on the affected side and compensatory enlargement on the opposite side. This is called plagiocephaly and is not usually associated with other abnormalities. Bilateral coronal synostosis produces a more severe foreshortening of the entire anterior fossa and is often manifested by shallow orbits with exophthalmos and hypertelorism. This entity is often associated with inherited congenital disorders such as Crouzon's disease and Apert's and Carpenter's syndromes.

Less common forms of craniosynostosis include premature closure of the lambdoid suture or the metopic suture. With unilateral synostosis of the lambdoid suture, the skull appears flattened in the affected occipital area, which may be confused with birth molding. With premature closure of the metopic suture, the forehead assumes a triangular shape (trigonocephaly). Neither of these forms is associated with other congenital anomalies.

The treatment of craniosynostosis is surgical and generally involves the opening of the affected suture along its entire length. This should be carried out as soon as possible after the diagnosis is made, since early surgical intervention provides the best cosmetic result. In cases of multiple suture involvement, prompt treatment provides early skull expansion to accommodate brain growth.

NEUROSURGICAL MANAGEMENT OF PAIN

The majority of neurosurgical patients have pain, either as their primary complaint or as a secondary manifestation of their disease process. Painful conditions can be categorized as acute processes, such as arm pain from a herniated cervical disc, or chronic processes, such as extremity pain from an invasive neoplasm. For most acute pain states, the cause can be identified and treated, but for chronic pain there is often no ready solution. In this section, the more common neurosurgical procedures available to manage chronic pain will be described.

At one time, the perception of pain was thought to involve a relatively simple system of pathways extending from the peripheral receptors to the brain. It has since been shown that this system is an extremely complex network of pathways, with a considerable amount of modification at multiple synaptic levels. Impulses from pain receptors reach the spinal cord via the dorsal root ganglion and may be significantly modified in the various laminae of the dorsal horn. This information is then relayed to the thalamus, but again may undergo considerable modification in the area of the brainstem reticular formation. This sensory input is subsequently relayed to the cortex for conscious interpretation. Modifiers in this complex system include the endogenous substances endorphins, enkephalins, and substance P. In addition, the psychologic state of the patient influences the perception of painful stimuli.

Traditionally, neurosurgical procedures for chronic pain have been ablative or destructive, but a variety of neuromodulating or stimulating procedures have been developed in recent years. These procedures are generally reserved for those chronic pain conditions that have failed medical therapy.

Cerebrum

In general, few painful states warrant procedures involving the cerebral hemispheres or deep brain nuclei. Bilateral rostral cingulotomies have been performed for treatment of intractable pain. This procedure disrupts the cingulum, a large fasciculus running deep to the cingulate gyrus. It has proved useful for affective disorders involving pain when performed bilaterally.

As the primary relay station for pain impulses, the thalamus has been the target for stereotaxic ablative procedures. Thalamotomies are performed with considerable accuracy and have been used for thalamic pain disorders, phantom limb pain, and pain from invasive tumors. More recently, deep brain electrodes have been placed for stimulation rather than ablation of these structures. In general, thalamic stimulation has proved useful in patients suffering from deafferentation pain, pain that typically does not respond to opiates. Conversely, brainstem stimulation is successful in chronic painful states that are responsive to opiates. Long-term success (control of pain for a minimum of 2 years) is reported to be about 60 and 80 percent for the respective groups.

Cranial Nerves

Trigeminal neuralgia (tic douloureux) is one of the more commonly occurring neuropathic painful conditions. It presents as an intermittent, shocklike pain in one or more divisions of one trigeminal nerve. It most commonly involves the second (maxillary) and/or third (mandibular) division of the nerve and is rarely bilateral. The pain usually lasts for seconds but is extremely severe and may be incapacitating. It is often triggered by touching the face, talking, or chewing. The pain may be present for weeks or months, then spontaneously disappear, only to return with increased severity. For most patients the pain can be controlled initially with phenytoin or carbamazepine, but eventually many will require surgical intervention. A small percentage of patients may have a posterior fossa tumor causing the pain, so evaluation should include a CT or MRI scanning before therapy.

In the past, surgical treatment involved ablation of the involved branch(es) of the trigeminal nerve. This could be accomplished peripherally by surgical section or alcohol ablation of the supraorbital, infraorbital, or inferior alveolar nerves. Pain control through these neurectomies was usually short-lived, however. Experience showed that preganglionic lesions must be made for more permanent relief. Retrogasserian rhizotomy is now carried out by open surgical approaches subtemporally or through the posterior fossa, or percutaneously by placing a radiofrequency electrode through the foramen ovale into the ganglion.

A nonablative approach involves microvascular decompression of the trigeminal nerve in the posterior fossa. The theory behind this approach is that trigeminal neuralgia is caused by external pressure on the nerve by vascular structures (an artery or vein) near its entry into the brainstem. With the use of the operating microscope, the offending artery can be moved or the vein ablated, thus decompressing the nerve. This procedure has a high success rate but carries more risk than the percutaneous method. An advantage is that the nerve's function remains intact.

Spinal Cord

Just like other surgical procedures for chronic pain, those involving the spinal cord have traditionally been ablative. Cordotomy, designed to obliterate the spinothalamic tract (see Fig. 40-3), can be performed by open operation or percutaneously. Anterolateral cordotomy may provide excellent temporary relief of pain for patients with terminal malignancies, but it is rarely effective for chronic benign conditions such as low back, postherpetic, or phantom limb pain.

For selected cases of severe pain of peripheral nerve origin, such as brachial plexus injury, postherpetic neuralgia, traumatic limb amputation, and root avulsion, ablative lesions can be made at the dorsal root entry zones (DREZ) of the spinal cord. These lesions are made with a radiofrequency lesion generator or laser through an open exposure of the cord via a laminectomy. Several levels are usually included. About 50 percent of patients obtain good relief from pain.

Chronic pain that develops in chest, flank, or abdominal incisions may warrant an ablative procedure. Since these regions do not contain critical sensory areas, unilateral dorsal nerve roots may be sectioned to deprive the involved area of sensation. It is necessary to ablate at least three adjacent levels in order to denervate one dermatome adequately because of overlapping sensory distribution. Dorsal rhizotomy may be carried out openly through a laminectomy or percutaneously through radiofrequency thermocoagulation. Pain relief from rhizotomy is good initially, then less so as time passes.

Intrathecal morphine can be given temporarily or permanently by infusion of small but effective doses. This newer method of pain control is particularly effective in debilitated patients with terminal illnesses. The procedure involves the subcutaneous implantation of a constant infusion pump that can be recharged periodically.

In chronic painful states of nonmalignant spinal origin, such as low back and/or leg pain, a nonablative neuromodulation technique may play a role in therapy. This involves transcutaneous excitatory nerve stimulation (TENS), which "blocks" nerve conduction of pain impulses. It is simple, safe, and relatively inexpensive. Electrodes are taped to the skin, usually over the region of the pain or directly over the affected major nerve, and then connected to a small, portable stimulating device. The device has a variable pulse width, frequency, and amplitude, which the patient can adjust in order to achieve maximum pain relief. Use of this device for a brief time often provides long-lasting relief with no untoward side effects. Direct stimulation of the spinal cord by surgically implanted electrodes combined with percutaneous telemetry can also provide relief from chronic pain.

Peripheral Nerve

Pain from a partial or complete nerve injury usually involves the nerve's sensory distribution but may include the whole extremity. Chronic pain developing after an amputation may be present in the remaining portion of the limb at the site of the amputation ("stump pain") or in the nonexistent amputated portion ("phantom pain"). The causation of the pain may be related to the sensory component of the nerve or to its associated sympathetic nerve supply.

With partial or complete peripheral nerve transection, a painful neuroma may form. The usual treatment is excision of the neuroma, with prevention of recurrent formation by burying the nerve end in bone or muscle or wrapping it in tantalum or Silastic. Neuromodulation techniques may also be applied in cases of painful neuromas.

Chronic pain resulting from peripheral nerve injury may be significantly altered by interruption of the sympathetic nerve supply to the affected extremity. The classic example of this dysautonomic state is major causalgia. This term implies a partial injury to a major nerve in an extremity. Minor causalgia is reserved for an injury to a more distal minor sensory nerve, which may also become a source of significant pain. It is also well recognized that a dysautonomic state may be created by major or minor trauma to an extremity that does not involve a peripheral nerve. This has been termed major or minor traumatic dystrophy. The entire collection of causalgias and traumatic dystrophies makes up a syndrome called reflex sympathetic dystrophy.

Major causalgia is most commonly related to partial injury of the sciatic or the median nerve. Typically, symptoms begin in the affected nerve's distribution but may progress to involve the whole extremity. The extremity first becomes swollen, warm, erythematous, and quite sensitive to touch. With time, it becomes cool and pale. Hyperhidrosis (excessive sweating) may follow. Due to lack of joint motion, the normal flexion and extension creases disappear and the skin becomes smooth and flat. Plain radiographs may demonstrate osteoporosis. Eventually the extremity may become completely useless. A constant, burning pain develops and persists throughout these various stages. It may be exacerbated by touching or moving the extremity. Even temperature changes and emotional stress can trigger worsening of pain. Minor causalgia and the traumatic dystrophies may be accompanied by similar, but less severe symptoms.

The treatment of these dysautonomic states is complex. They may be helped by disruption of the sympathetic nerve supply to the extremity. Sympathetic denervation may be extremely rewarding in major causalgia but is less so in minor causalgia and the traumatic dystrophies.

Less severe and more easily treated pain may arise from chronic compression of selected peripheral nerves. The most common are compression of the median nerve at the wrist (carpal tunnel syndrome) and compression of the ulnar nerve at the elbow. Chronic compression may result in pain, paresthesias, numbness, and eventually weakness and atrophy of muscles in the distribution of the affected nerve. These compression syndromes are diagnosed clinically and confirmed by finding denervation and slowed nerve conduction on electromyography and nerve conduction velocity testing. Treatment is surgical decompression of the involved nerve, with prompt and long-lasting relief in most cases.

EPILEPSY AND MOVEMENT DISORDERS

Epilepsy

That removal of specific areas of the brain can cure epilepsy has been known for many years. Inadequate diagnostic tools and poorly defined indications for surgery, however, precluded widespread use of surgery to treat the disease. About 2 million people in the United States have epilepsy, and most undergo treatment for it. Most varieties of epilepsy can be categorized into either generalized (grand mal) seizures characterized by loss of consciousness with tonic/clonic movements, or partial seizures manifested by involuntary movements but no loss of consciousness. Epilepsy can be controlled fairly well by available medications in about 80 percent of patients. Of the remaining 20 percent, nearly 100,000 may benefit from available surgical procedures. Interest in surgical management of epilepsy has increased significantly in recent years because of the availability of accurate physiologic localization of seizure foci and high- quality imaging of the brain by CT, MRI, and *PET* scanning.

Epilepsy has many etiologies. Congenital anomalies of the brain are relatively common causes in the pediatric age group. Birth injury is also a well-known precursor of seizure disorders in children. Neoplasms and vascular anomalies, including arteriovenous malformations, are common seizure sources in adults. A large number of adults have epilepsy because of mesial temporal sclerosis, a condition that probably results from hypoxia and/or hypermetabolic states during early development of the brain. Some seizures have no known cause.

Patients can be tested effectively in order to select those who can clearly benefit from surgery. Electroencephalography and MRI are initial monitoring tools. More invasive monitoring involves recordings from depth electrodes placed stereotactically and subdural electrodes placed by open craniotomy. Specific foci can be identified with prolonged monitoring using video recording coupled with continuous EEG monitoring. Areas of brain that serve specific functions such as speech and motor/sensory functions can also be identified by specific tests, including PET scanning and functional MRI.

Operation is planned once a focus has been identified that is amenable to surgical resection without undue risk. Often seizure surgery is performed with the patient under local anesthesia. Electrocorticography, which is similar to electroencephalography except that it is performed directly from brain cortex, can further identify the focus of seizures at the time of operation. Stimulation of brain is also used to map the cortex and to identify regions that subserve speech, vision, and sensory/motor functions.

Outcomes of seizure surgery today are often gratifying. Many patients with uncontrolled epilepsy are not able to function in society because of their seizures. About half of those who have uncontrolled seizures can be made seizure-free by surgical techniques. Another one-quarter can have significant reduction in both seizure frequency and seizure intensity by surgical therapy.

Movement disorders

Movement disorders result from a variety of central nervous system diseases and are manifested by abnormal volitional movement. Typical examples are the resting (or "pill-rolling") tremor of Parkinson's disease and the intention tremor that accompanies volitional movement as a consequence of cerebellar disorders. Many treatments have been used since these conditions were first recognized, but most have been ineffective. Because of advances in neurophysiology, neuropharmacology, neuropathology, and anatomic imaging during this century, the balance between excitatory and inhibitory effectors of movement has come to be better understood, and the imbalance responsible for abnormal volitional movements better explained. Improvements in both medical and surgical treatments have resulted.

The symptoms of Parkinson's disease, characterized by resting tremor, cogwheel rigidity, and bradykinesia, were not effectively controlled until it was known that there is a dopamine deficiency in the substantia nigra resulting in inhibition of the voluntary initiation of movement. Dopamine supplements were then developed, and symptoms were often improved.

Surgical treatment began in the 1950s with ablative lesions of the basal ganglia designed to interrupt imbalanced inhibitory/excitatory circuitry and its effect on movement coordination. An ablative lesion in the globus pallidus, called pallidotomy, was not very effective at first, though it did seem to help rigidity and bradykinesia. Similarly, thalamotomy proved to have a modestly beneficial effect on tremor. Neither operation was used extensively because medical treatment, particularly with dopaminergic agents, was generally effective.

Stereotactic methods, particularly pallidotomy, have become popular because medical therapy has not remained effective in many patients over the long term. Escalating doses of medication have often been necessary to control parkinsonism, resulting in widely fluctuating symptoms and signs. Patients who have resting tremor without bradykinesia are considered for thalamotomy, and those with rigidity, bradykinesia, and gaze and balance difficulties are candidates for pallidotomy.

Other movement disorders that respond to ablative lesions in the basal ganglia include essential tremor, intention tremor of multiple sclerosis, and the choreiform movements of Huntington's disease. Hemiballismus and hereditary dystonia can also be treated by stereotactic lesions in the basal ganglia.

RADIOSURGERY

Radiosurgery is a highly focused ionizing radiation derived from an external source delivered to a stereotactically defined intracranial target. Treatment of lesions in the depths of the brain by focused radiosurgery began about 20 years ago. The technique, using the so-called "gamma knife," allows delivery of radiation to a specific target from a variety of sources, ablating the target without significant adverse effects on tissue surrounding it. In some instances, the technique is delivered by fixed cobalt sources delivering photon radiation to the target, aided by a stereotactic frame attached to the patient's head and a computer program coupling the radiation source to the frame on the patient. Identification of the intracranial target is made by CT or MRI scanning, using a computer to relate the target to the stereotactic frame and the patient within it. Other systems involve a similar principle, but the radiation source, delivered by linear accelerator, moves around a fixed target.

Currently radiosurgery is given in a single dose requiring little or no time in the hospital. Fractionated treatments using radiosurgery are being developed, reducing the risk of treatment and increasing its effectiveness. Radiosurgery is used for a variety of conditions, including arteriovenous malformations and neoplasms, and is particularly adaptable to deep intracerebral lesions. It causes sclerosis of vascular structures, resulting in progressive occlusion of arteries, such as those within arteriovenous malformations, and cell necrosis in tumors. Radiosurgery can gradually ablate small to medium-sized arteriovenous malformations with 85 percent efficacy over a 2- to 3-year follow-up period.

Radiosurgery is also used to treat neoplasms, including intracranial schwannomas and meningiomas. In both instances the target is usually well defined and adjacent brain is not involved. A tumoricidal dose can be delivered in a single shot without significant adverse affects on surrounding tissue. Collimation of the beam allows it to be shaped to fit the anatomic dimensions of the target lesion.

The role of radiosurgery in other tumors, such as metastatic tumors and gliomas, is uncertain at this time. It is probable that metastatic intracerebral lesions will be treated more by this technique in the future because it is relatively noninvasive, and it is probably as effective as surgery for small lesions. Larger lesions are less well suited to radiosurgery because they would require radiation doses larger than can be delivered safely.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 41 - Orthopaedics

Randy N. Rosier

GENERAL CONSIDERATIONS

Pain

Anatomy and Physiology

Definition

The International Association for the Study of Pain defines pain as “an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.” Afferent nociceptive impulses produced by injurious stimulation are transmitted to the central nervous system, where they are given meaning by the perceptual state of the individual. The perceptions of pain are modified by many factors, including past and present experience, state of awareness, concomitant sensory stimuli, and emotional status. Stimulation of peripheral receptors by noxious agents produces a spatiotemporal pattern of nervous impulses that is interpreted as pain within the higher cerebral centers. Teleologically, pain serves the useful function of prevention of tissue damage and is a homeostatic mechanism. It is only in the postoperative setting or when pain becomes chronic that it serves no useful function. Acute pain implies the presence of actual tissue damage or its potential unless the noxious stimulus is removed; it may be associated with autonomic hyperactivity such as hypertension, tachycardia, sweating, and vasoconstriction. Chronic pain implies the absence of a threat of tissue damage yet is described in terms of such damage (discomfort, suffering). Generally pain is considered to be chronic when its duration exceeds 3 to 6 months.

Transduction

Pain is initiated by a stimulus that is detected by nociceptors (*transduction*) and is the most common manifestation of disorders of the musculoskeletal system. Pain can be produced by a wide variety of physical phenomena, including pressure, puncturing, squeezing, tension, and extremes of temperature, and by chemical effects such as change of pH or release of polypeptide mediators, including histamine-like substances, serotonin, bradykinin, and other polypeptides. Prostaglandins may lower the pain threshold for some stimuli, and inflammation accompanied by local acidosis can enhance perception of pain. In addition, a number of local peptide mediators such as substance P are released at sites of injurious stimuli and trigger patterns of nervous activity interpreted as pain.

Transmission

Most cutaneous sensory nerve endings consist of unmyelinated fibers, which produce sensations of pressure, touch, or pain, depending on the impulse pattern invoked. These nerve endings are found also in periosteum, arteries, joint capsules, and synovium, with similar myelinated fibers in muscle. The signal is then transmitted via sensory afferent nerve fibers and dorsal root ganglia to the spinal cord. The smaller-diameter myelinated A-delta and unmyelinated C afferent fibers are those primarily responsible for pain transmission (Table 41-1). Synapses in the dorsal horn interact with other afferents as well as the spinoreticular pathway and transmit information via the spinothalamic tract to the thalamus and somatosensory cortex.

Modulation

Modulation of the nociceptive information occurs in the dorsal horn through a number of neurotransmitters (Fig. 41-1). Descending modulating systems are generally inhibitory and are influenced by multiple factors, including emotional state. Endogenous opioid systems (enkephalins and endorphins) regulate descending modulatory activity, and multiple opioid receptors for these substances have been identified in the central and peripheral nervous system. These same receptors are responsible for the inhibitory effects of opioid analgesics on pain level.

Additional modulation may occur within the dorsal horn, according to the gate theory of pain proposed by Melzack and Wall (Fig. 41-2). This hypothesis involves interactive influences of myelinated and unmyelinated fibers within the dorsal horn, determining the net level of nociceptive output of the transmission cells. Both fiber types are thought to directly excite the transmission cell, while they have opposing effects on the interneurons in the substantia gelatinosa, which also regulate the level of transmission-cell activity. Transcutaneous electrical nerve stimulation as a pain control method is based on this hypothesis, and its efficacy supports the gate theory of pain. Selective stimulation of large-diameter myelinated afferents blocks pain, since the unmyelinated A-delta and C pain fibers have a high electrical threshold and remain unstimulated. This technique is innocuous, noninvasive, and of great benefit to some patients with postoperative or chronic pain.

Perception

The endpoint in the pain pathway is its *perception* by the patient. Pain occurs with various qualities, such as aching, burning, spasmodic, radiating, lancinating, dull, or sharp. Local pain is felt at the site of injury, while diffuse pain appears to be more characteristic of deep structures, and radicular pain radiates along peripheral nerve pathways, often in association with neurologic deficits such as sensory or motor loss. Referred pain occurs in a location remote from the site of tissue pathology, and represents a misplaced cortical pain perception. Common examples include knee pain as a manifestation of hip joint pathology, and gluteal or posterior thigh and leg pain as a manifestation of spinal pathology. Referred pain tends to follow spinal segmental innervation and must be differentiated from radicular pain. For instance, injection of saline into interspinous ligaments has been demonstrated to cause referred gluteal and lower extremity pain (Fig. 41-3). Radicular pain, however, will follow specific dermatomal distributions.

Tissue Patterns

Bone

Bone is not thought to have any sensory endings within it, although the periosteum is richly innervated with both myelinated and unmyelinated nerve endings. Small unmyelinated fibers have been identified in association with blood vessels within bone and probably are sympathetic fibers responsible for bone blood flow regulation. Bony lesions such as tumors or infections cause a deep, boring type of pain that may result from pressure sensation mediated by blood vessel-associated fibers; pain from fractures, in contrast, has a sharper quality and is characteristically relieved by rest.

Muscle/Tendon

Muscle pain may result from direct injury, overuse or chemical irritation from metabolites such as lactic acid resulting from tissue anoxia. Types of local injury causing pain include contusion, partial or complete tendon or muscle rupture, excessive stretch or load under tension (muscle strain), and inflammatory disorders (e.g., myositis). Muscle injury usually is characterized by tenderness of the muscle to palpation and soreness aggravated by movement or voluntary contraction of the muscle. *Muscle spasm* refers to a sustained involuntary muscle contraction, which can cause severe, paroxysmal cramplike pain and is a common response to muscle injury. Spasm also can result from injury to the innervation of a muscle, as in sciatica, or from metabolic abnormalities such as hypocalcemia, alkalosis, or the presence of toxins such as tetanus. Myalgia, or aching pain in muscle, can occur as a symptom of systemic viral infections, or in association with chronic idiopathic disorders such as fibrositis or fibromyalgia.

Ischemia of muscle causes regional lactic acidosis, which can produce aching pain. More severe degrees of ischemia, usually associated with significant trauma such as fractures, but occasionally seen with excessive exertion, can result in a *compartment syndrome*. Given that muscle is bounded by inelastic fascial compartment coverings, ischemia that causes muscle swelling can result in elevated tissue pressure within muscle compartments, which in turn impedes vascular inflow and starts a vicious cycle of worsening ischemia. The result is severe pain out of proportion to the severity of the initial injury as well as pain with passive stretch of the muscle(s) involved. Emergent surgical release of involved compartments is required to prevent permanent muscle and nerve damage in this condition.

Tendon-related pain is most commonly seen with tendinitis, an inflammation of the tendon and associated sheath (paratenon). Tendinitis usually is a result of overuse, and as such it may reflect the result of mechanical disruption of some of the collagen fibers in a tendon. Common sites include the rotator cuff tendons of the shoulder, the tendons about the wrist, the patellar tendon, and the Achilles tendon. Local tenderness and swelling result, and the pain is generally invoked by contraction of the associated muscle. Tendinitis may presage a rupture of the involved tendon, occasionally seen with the Achilles, patellar, and rotator cuff tendons, and with the tenosynovitis of rheumatoid arthritis in the hand and wrist. Treatment usually involves rest of the affected area and anti-inflammatory medication. Ruptures require surgical repair, and persistent tendinitis may require surgical removal of mechanical causes or tenosynovectomy.

Joint

Joint pain may result from hyperemia or inflammation of synovium, joint effusion producing capsular distention, instability causing traction on capsular or ligamentous structures, or degeneration of articular cartilage. Cartilage is avascular and lacks nerve endings, indicating that pain resulting from cartilage injury or degeneration originates in the underlying bone or adjacent capsule and synovium via secondary mediators.

Neurogenic Pain

Peripheral nerves may cause pain in response to pressure, ischemia, stretching, herpes zoster infection, toxins (lead, arsenic), or metabolic disturbances (vitamin deficiencies, diabetic or alcoholic neuropathies). A characteristic of neurogenic pain is its radicular nature and association with neurologic symptoms such as paresthesias, sensory or motor loss, and secondary muscle atrophy. Hyperesthetic pain may be encountered with herpes zoster infection or Guillain-Barré syndrome, or during axonal recovery from mechanical injury. HIV infection also has been identified as a cause of chronic peripheral neuropathic pain.

Treatment

The mechanism of action of many of the commonly used nonnarcotic analgesics is unknown. Multiple receptors that respond to narcotic medications as well as endogenous analgesics (endorphins) have been identified. For chronic pain from conditions that are not life threatening, efforts are made to avoid the use of narcotics because of the potential for progressive dependence and development of tolerance to the drugs. Nonsteroidal anti-inflammatory and analgesic drugs are useful, and for chronic neurogenic pain antidepressants (amitriptyline) and anticonvulsants (carbamazepine) may be useful.

Postoperatively, patient-controlled analgesia techniques, wherein the patient can self-administer intravenous narcotic medications at low doses as needed, are widely used. This method not only gives better and more even relief of postoperative pain but also has been shown to decrease the total amount of medication needed. The use of epidural anesthesia for surgical procedures also allows the use of the epidural catheter postoperatively for pain control. Local anesthetics or narcotics can be used in conjunction with a pump to deliver minute doses of the agents directly to the epidural space around the spinal cord, usually with total relief of pain.

Implantable subcutaneous or epidural pumps also are useful for continuous delivery of opioids in patients with severe chronic pain such as those with metastatic cancer or AIDS. In selected cases of chronic severe pain involving an extremity, neural ablations with injections of alcohol or phenol have been used successfully to improve pain control.

The topical analgesic capsaicin, which is thought to deplete local tissue stores of substance P, has been used successfully for pain from arthritis and from herpes zoster infection.

Upper Extremity Wrist and Hand

Radiocarpal, carpometacarpal, metacarpophalangeal, and interphalangeal arthritides frequently cause pain in the region of the wrist and hand, usually with swelling and stiffness of the affected joints. Additional causes include tendinitis, de Quervain's disease, compression of the median nerve at the wrist beneath the transverse carpal ligament (carpal tunnel syndrome), compression of the ulnar nerve beneath the volar carpal ligament (ulnar tunnel syndrome), or radicular pain from compression of these nerves at a more proximal level. Such neurogenic pain can be accompanied by sensory or motor deficits, and when associated with vasomotor trophic changes can indicate the presence of a reflex sympathetic dystrophy (RSD). RSD, which can occur in the upper or the lower extremity after an injury or surgery, is a sustained abnormal reflex caused by efferent activity arising from sympathetic nerves (Fig. 41-4). When the inciting injury is known to involve a major nerve trunk, it is referred to as *causalgia*. The pain typically is described as burning in character, and there is associated hyperesthesia and dysesthesia (painful response to normally nonpainful stimuli), swelling, joint stiffness, vasomotor instability, and marked osteopenia in the region (Sudeck's atrophy). In the upper extremity, adhesive capsulitis of the shoulder may occur with RSD of the hand (shoulder-hand syndrome). Regional sympathetic blockade aids in diagnosis as well as treatment by breaking the reflex arc, and physical therapy and corticosteroids may be helpful.

Elbow

The most common causes of elbow pain are medial or lateral epicondylitis (“golfer's elbow” and “tennis elbow,” respectively), arthritis of the elbow joint, and compression neuropathy of the ulnar nerve at the elbow (cubital tunnel syndrome). Treatment is rest, anti-inflammatory medication, and, in refractory cases, surgical release of the affected muscle origin or nerve decompression.

Shoulder

Common causes of shoulder pain include arthritis of the glenohumeral or acromioclavicular joints and impingement syndrome, in which tendinitis and bursitis of the rotator cuff result from impingement of these structures on the coracoacromial ligament. Forward elevation and internal rotation (impingement test) bring the supraspinatus tendon in contact with the coracoacromial ligament, reproducing the pain. Injection of lidocaine and cortisone in the subacromial bursa is helpful diagnostically and therapeutically. In refractory cases, anterior excision of the acromion and coracoacromial ligament (Neer acromioplasty) may be helpful. Impingement chronically can lead to rotator cuff tears, the majority of which can be managed conservatively. Persistent symptoms or complete rupture in a young person are indications for rotator cuff repair. Diagnosis of these lesions is aided by arthrography and magnetic resonance imaging (MRI) scans. Bicipital tendinitis also can present with anterior activity-related shoulder pain, with local tenderness and pain with resisted biceps muscle contraction. Conservative treatment usually suffices. Occasionally rupture of the tendon of the long head of the biceps can occur, particularly in association with impingement syndrome and rotator cuff pathology.

Cervical Origin

Cervical arthritis or nerve root irritation from osteophytes or disc herniations can cause both referred and radicular pain in the shoulder area. In addition, visceral pathology in the heart, lungs, or pleura can be referred to the shoulder or arm. Brachialgia (brachial neuralgia) is characterized by upper extremity pain associated with paresthesias, altered sensation, weakness, and reflex changes in a radicular distribution of the brachial plexus. Occasionally sympathetic plexus disturbances are seen, with vertigo, tinnitus, or visual disturbances. Causes of brachialgia include tumors of the spinal cord or nerve roots, infections, disc herniation, brachial plexus trauma, Pancoast's tumor of the apex of the lung, congenital anomalies of the cord such as syringomyelia, and cervical vertebral subluxations or dislocations. Another possible cause of brachialgia is the thoracic outlet syndrome, in which compression of the brachial plexus and vascular outflow obstruction can result from cervical ribs, anomalous fibrous bands within the scalene muscles, or changes caused by trauma to the clavicle or scalenes. Scalene muscle or rib resection can improve the symptoms if conservative measures fail. Differential diagnosis of the causes of cervical and shoulder pain is aided by radiographs to delineate bony abnormalities, and by computed tomography (CT) or MRI, which can demonstrate both bony and soft-tissue pathology. Cervical strain syndrome (whiplash injury) can result from rapid acceleration or deceleration of the head and is common in motor vehicle accidents. Pain typically is present in the cervical area and often is also referred to the shoulder and arm. Treatment is conservative, with a cervical collar for rest and appropriate medications followed by a cervical isometric exercise program, and, in refractory cases, intermittent cervical traction.

Cervical Disc Disease

Cervical disc herniations most commonly occur at the C4–C6 levels, the region responsible for the majority of flexion/extension motion in the cervical spine. Degenerative changes in the discs or trauma can cause herniation of the gelatinous nucleus pulposus through the annulus fibrosus, allowing impingement on nerve roots or the spinal cord. Neck pain and radicular pain in the distribution of the involved root level(s) can result. Symptoms may be unilateral or bilateral, depending on whether the herniation is central or lateral; in some cases central herniations can cause cord symptoms such as Brown-Séquard syndrome, hyperreflexia below the lesion, incontinence, and gait disturbances. Lateral herniations are the most common, with corresponding local and radicular symptoms. Local paraspinous muscle spasm is common, with rigidity and pain with motion. Vertical compression, particularly with head tilt to the affected side, exacerbates the symptoms, and vertical traction tends to diminish them. Depending on the affected level, the brachioradialis, biceps, or triceps reflexes may be depressed (C5, C6, or C7, respectively), and corresponding muscle weakness or dermatomal sensory changes may be present. Definitive diagnosis can be made by cervical myelography or MRI, along with electromyographic studies to assess the degree and level of neurologic involvement. Treatment consists of cervical traction followed by anti-inflammatory medication and a cervical collar. If neurologic deficit does not respond promptly to traction, surgical discectomy and fusion of the involved vertebrae may be necessary.

Cervical Spondylosis

Degenerative changes in the cervical discs can lead to narrowing of the intervertebral foramina and osteophyte formation in the adjacent facet joints. This can cause impingement on nerve roots, with neck pain and neurologic symptoms as described above for disc herniations. In severe cases the facet joint hypertrophy can lead to stenosis of the cervical spine, resulting in cervical myelopathy as well as radicular symptoms. Sphincter disturbances occur in about one-third of patients, but incontinence is unusual. The causes of the myelopathy are multifactorial but include ischemia to the anterior spinal cord from vascular compression, ligamentous instability that places strain on the cord, and pressure by osteophytes. When conservative measures such as traction, cervical collar, and postural exercises fail to prevent neurologic progression, surgery may be indicated. Anterior interbody fusion (Cloward procedure) and posterior laminectomies or laminaplasty provide relief of symptoms and neurologic improvement in a high proportion of patients. Because of the instability caused by laminectomies, laminaplasty is preferable.

Lower Extremity Foot and Ankle

The most common causes of foot pain are metatarsalgia and plantar fasciitis, which result from repetitive loading of the metatarsal heads or of the attachment of the plantar ligament to the calcaneus. Appropriate shoe inserts to relieve pressure (heel cups, metatarsal pads or bars, insoles), and anti-inflammatory medication or local cortisone injections generally alleviate the symptoms. In refractory cases surgical release of the plantar fascia or metatarsal osteotomies or head resections are occasionally indicated. Arthritis of any of the joints in the midfoot or forefoot can cause pain, and the first *metatarsophalangeal (MTP)* joint is particularly susceptible to acute gouty arthritis as well as osteoarthritis. Surgical arthrodesis (fusion) or arthroplasty is occasionally necessary when conservative management is unsuccessful. Additional causes of foot and ankle pain include peroneal tendon subluxation, stress fractures of the navicular or metatarsals, and compression of the posterior tibial nerve at the ankle or distally (tarsal tunnel syndrome). Surgical decompression of the nerve in tarsal tunnel syndrome is not associated with as successful an outcome as it is in treating other compression neuropathies.

Lower Leg

Repetitive loading of the tibia can lead to “shin splints,” or activity-related pain and tenderness over the tibia. Several underlying causes have been identified, including an exertionally induced compartment syndrome and a periostitis involving inflammation of the attachments of the posterior tibialis fascia to the tibia. Exercise-induced compartment syndrome can be diagnosed by measuring compartment pressures in the leg before and after strenuous exercise with a slit catheter (see Fig. 41-22 B). Release of the fascia may be indicated in persistent cases. Stress fractures of the tibia also can occur and present similarly, although radiographs often indicate periosteal reaction, and a nuclear bone scan will be abnormal. Protected weight bearing, alone or with cast immobilization, allows healing. Female runners with amenorrhea have been identified to be at risk for osteopenic tibial stress fractures secondary to decreased estrogen levels.

Pain in the posterior calf can result from partial tears of the medial gastrocnemius muscle; pain was formerly ascribed to rupture of the plantaris muscle. Muscle cramps in the calf, commonly occurring at night, can result from overuse. Deep venous thrombosis in the calf must always be considered in the differential diagnosis of leg pain and usually is associated with diffuse swelling, warmth, tenderness, and pain with passive stretch of the gastrocnemius muscle or toe flexors (Homans' sign). Ultrasonography and venography are important diagnostic aids. Calf pain that progresses with ambulation but is relieved at rest may indicate ischemic claudication. The neurogenic claudication of spinal stenosis continues when the patient stands at rest, which differentiates it from ischemic claudication.

Knee

Traumatic intraarticular derangements of the knee, including tears of the menisci, ligamentous sprains, and osteochondral fractures, are a common cause of knee pain. A history of locking can be found with intraarticular loose bodies, osteochondral defects, or meniscal tears. Arthritis of the knee also can cause pain and local swelling, often with joint space narrowing or osteophyte formation visible radiographically. Because of the common innervation of the medial aspect of the knee and the hip joint by the obturator nerve, knee pain can be a manifestation of hip joint pathology. In these cases local tenderness and swelling of the knee are absent, and forced passive rotation of the hip usually elicits the knee pain.

Anterior knee pain is frequently caused by the patellofemoral pain syndrome, which is often associated with degeneration and fibrillation of the articular cartilage of the patellar facets (chondromalacia). Causes include blunt direct trauma to the patella, malalignment of the extensor mechanism, or recurrent subluxation or dislocation of the patella. Patellar tendinitis (jumper's knee) also can present with anterior knee pain and focal tenderness over the patellar tendon. Diagnosis of chondromalacia is aided by demonstration of patellofemoral crepitus, pain with patellar compression, and tenderness of a patellar facet. Treatment generally is conservative, with isometric quadriceps exercises. If malalignment exists, lateral retinacular release or realignment of the extensor mechanism can be considered, and arthroscopic debridement of chondral fibrillations can be helpful. MRI is often helpful as a noninvasive method for evaluation of intraarticular pathology, including ligament tears, osteochondral fracture, meniscal tears, and chondromalacia and is used increasingly in place of diagnostic arthroscopy.

Hip

An important clinical diagnostic consideration in the evaluation of hip pain is the localization of the site of pathology. Intraarticular hip disorders usually present with anteriorly localized (inguinal) pain, aggravated by weight bearing or by passive rotation of the hip. Referred pain or radicular pain such as sciatica more commonly presents in the gluteal area, as does pain derived from sacroiliac joint pathology. Hip pain can be secondary to osteoarthritis, avascular necrosis of the femoral head, synovitis, septic arthritis, stress fractures of the femoral neck, or avulsion fractures of the anterior inferior iliac spine or the lesser trochanter. Lateral hip pain and tenderness can result from greater trochanteric bursitis, and anteromedial hip pain can be caused by adductor tendinitis, iliopsoas bursitis, or an iliopsoas abscess. With psoas abscess or septic arthritis, the hip tends to be held in flexion and external rotation. Piriformis syndrome, or entrapment of the sciatic nerve under the piriformis muscle, causes posterior hip and thigh pain (sciatica) and is aggravated by internal rotation of the hip. Hip flexion contractures frequently result from chronic hip joint pathology.

Low Back

The *low back syndrome* refers to a disease or injury of the lumbosacral spine, of an acute or a chronic nature. There are a wide variety of causes of low back pain; a summary is presented in Table [41-2](#). Acute low back pain, which is most common in the third to fifth decades, can be activity related, associated with paraspinous muscle spasm, and aggravated by sneezing or coughing. Pain usually is not radicular in nature, but it can be referred to the buttocks or legs. When nerve root irritation is present, paresthesias, neurologic deficits, and radicular radiation of the pain can occur. In the majority of patients with low back pain, no specific pathoanatomic cause can be identified, and treatment is symptomatic. For all idiopathic cases of low back pain, 90 percent of patients recover within 3 months, and 50 to 60 percent of patients with acute back pain recover within 1 week.

In disease processes such as tumors or infections involving the spine, the pain tends to be severe, unremitting, and not relieved by rest. Mechanically caused back pain is activity dependent, while morning pain and stiffness can be associated with ankylosing spondylitis. Examination of the patient should include assessment of the spinal range of motion, straight-leg- raising test for sciatic irritability, and a complete neurologic examination. Low back syndrome is most effectively treated by a short period of bed rest in the semi-Fowler's position (knees and hips flexed), anti- inflammatory medication, local heat, and occasionally muscle relaxants for spasm. Patients mobilized after 2 days of rest have been shown to recover more quickly than those rested 7 days. With subsequent mobilization, isometric back and abdominal exercises are helpful, as is the occasional use of a corset or back brace. Cardiovascular fitness training appears to be important both as a preventive measure against recurrences and for improved functional capacity. Such regimens include regular walking, swimming, or cycling.

Occupational low back pain (OLBP) is one of the most common and costly problems for workers in industry, and it is the second most frequent cause of worker absenteeism. As many as 50 percent of workers in the United States are affected at some time, and low back pain is the most common work-related complaint treated by primary care physicians. This disorder is the leading cause of disability in persons in the 19- to 45-year-old range. Risk factors include repetitive lifting, twisting, whole body vibration, and chronicity, and disability may also be influenced by psychological factors, including work stress and low job satisfaction. The natural history of OLBP is favorable, with 60 percent of patients improved by 4 weeks and 80 percent by 6 weeks. Initial treatment is conservative, as outlined in the algorithm in Fig. [41-5 A](#). When recovery seems very slow, further evaluation may be indicated as outlined in Fig. [41-5 B](#).

Spondylolisthesis

Spondylolisthesis is a forward subluxation of one vertebral body on another. It can be caused by (1) spondylolysis (a defect in the pars interarticularis), (2) fracture of the posterior elements, (3) congenital facet deficiency, (4) facet deficiency caused by degenerative disc disease, and (5) isthmic elongation of the pars interarticularis. Illustrations of the types of spondylolisthesis are shown in Fig. [41-6](#). Although back pain may occur along with hamstring tightness, sciatica, and, rarely, neurologic symptoms and signs, spondylolisthesis often is asymptomatic. The L5–S1 articulation is most commonly affected, and the deformity is best visualized on a lateral radiograph, as seen in Fig. [41-7](#). Pars interarticularis defects (spondylolysis) are best visualized with oblique radiographs and are generally believed to result from incomplete healing of traumatic stress fractures. From 5 to 25 percent of patients presenting in childhood will demonstrate progression of the displacement and can require posterolateral fusion, along with excision of the posterior elements (Gill procedure) if neurologic deficit is present. Otherwise treatment is conservative, with rest and abdominal exercises. There has been renewed interest in reduction of severely displaced vertebral bodies before fusion, as well as in direct repair of pars interarticularis defects.

Sciatica

Sciatica is a symptom rather than a disease, and the term is used to describe radicular-type pain in the lower extremity. Sciatica can be caused by nerve root compression by a herniated disc, tumor, abscess, or osteophyte, or by peripheral nerve compression within the pelvis or gluteal area by tumor, hematoma, or abscess. Disc herniations or degenerative arthritis with facet hypertrophy and foraminal stenosis are by far the most common causes. Inflammatory disorders of nerves and nerve roots also can cause sciatica. Alcoholic or diabetic neuropathy, arsenic or lead poisoning, syphilitic or herpes zoster infections, and vasculitis associated with collagen vascular diseases all have been associated with sciatica.

Disc herniations occur most commonly at the L5–S1 and L4–L5 levels, when a tear or degeneration in the annulus fibrosus allows herniation of the soft, gelatinous nucleus pulposus posteriorly into the spinal canal. Impingement on nerve roots then causes back pain and sciatica, sometimes with radicular neurologic symptoms. Because the posterior longitudinal ligament provides support in the midline, most disc herniations are posterolateral, and hence symptoms often are unilateral. Disc herniations are uncommon in children and in older adults. The low incidence of disc herniations in older individuals probably is related to the age- dependent loss of water content and disc volume. Pain from disc herniations usually is aggravated by sitting, coughing, sneezing, and forward flexion, all of which increase disc pressures. Radicular symptoms also are elicited by straight-leg raising, particularly with additional dorsiflexion of the foot (Lasègue's sign). Depressed reflexes in the affected distribution (knee jerk—L4; posterior tibialis reflex—L5; ankle jerk—S1) as well as associated muscle weakness and dermatomal sensory deficits can be found.

Radiographs may be normal or may show narrowing of the affected disc space. The lesion can be demonstrated radiographically by CT scan, myelography, or MRI (Figs. [41-8](#) and [41-9](#)). Treatment usually is conservative initially, with 80 to 90 percent of patients improving spontaneously and not requiring surgery. Helpful therapeutic measures include bed rest, analgesics, abdominal isometric exercises, and sitting in a reclining position, with mobilization as tolerated as the symptoms subside. Avoidance of heavy lifting and bending helps to prevent recrudescence. Surgical excision of the extruded portion of the disc or digestion by percutaneous injection with chymopapain or collagenase into the disc can be effective when symptoms progress despite conservative measures. Fusion at the time of discectomy has not been shown to be of benefit in the absence of spinal instability. Microdiscectomy and percutaneous suction discectomy have recently been used as less invasive surgical alternatives but require further validation of efficacy and delineation of indications.

Spinal Stenosis

Spinal stenosis, a narrowing of the spinal canal or neuroforamina, can be acquired, as in the case degenerative disc disease, or congenital, as in achondroplasia. Patients present with back or leg pain, generally exacerbated by standing and walking and, unlike in discogenic back pain, relieved by sitting. Leg pain secondary to spinal stenosis can mimic vascular claudication but is present while standing still, unlike claudication. Neurologic signs, including hyporeflexia and muscle weakness, may be present; sciatic irritability may be evident. Facet arthropathy associated with lumbar spondylosis is readily visualized on lateral and oblique radiographs, but the degree of spinal canal, lateral recess, or foraminal narrowing is better assessed on axial CT or MRI of the lumbar spine (Fig. [41-10](#)). Treatment consists of abdominal isometrics, flexion bracing, and anti-inflammatory medication. Epidural steroid injections can be of some benefit. In refractory cases, wide posterior surgical decompression (laminectomies, facetectomies, foraminotomies) with or without fusion can be undertaken.

Vertebral Tumors

Malignant and benign lesions can occur in the spine as primary tumors. Benign lesions include osteoid osteoma, osteoblastoma, aneurysmal cyst, giant cell tumor, and eosinophilic granuloma. Malignant primary lesions are rare, with the exception of multiple myeloma, and include chondrosarcoma, Ewing's sarcoma, and lymphoma. Chordoma, a rare tumor that evolves from remnants of the notochord, occurs in the sacrococcygeal and occipitocervical areas. The most common malignant tumor involving the spine is metastatic carcinoma, usually from a primary tumor in the breast, prostate, lung, kidney, or thyroid gland. Back pain is the usual presenting symptom, but neurologic signs and symptoms, including paralysis, can occur. Lower-extremity hyperreflexia, spasticity, and positive Babinski signs indicate a danger of progression to paralysis and require emergent treatment. Primary benign tumors such as aneurysmal cyst or osteoblastoma are treated by surgical excision, with fusion when necessary. Primary malignant tumors such as chondrosarcoma or chordoma require wide surgical excision, while more radiosensitive lesions such as plasmacytoma or Ewing's sarcoma are treated with radiotherapy and chemotherapy, as are most metastatic lesions. In the presence of spinal instability (defined as a lesion causing a progressive neurologic deficit, or mechanical disruption with the potential to cause such a deficit), surgical decompression through an anterior or a posterior approach, along with fusion and instrumentation, is performed. Decompression without fusion and instrumentation can lead to instability, progressive deformity, and recurrent neurologic deficit. In cases requiring surgical intervention in which the patient has not had prior radiotherapy, postoperative treatment after wound healing is indicated to prevent local tumor recurrence.

Vertebral Osteomyelitis

Pyogenic osteomyelitis is usually caused by a *Staphylococcus aureus* infection that starts in the vertebral end plates adjacent to a disc space by hematogenous spread from another infectious focus. Back pain, radiographically visible destruction of the disc space, and sometimes neurologic deficit result. Patients may have systemic symptoms of infection; usually leukocytosis and elevation of the erythrocyte sedimentation rate are present. Marked collapse of the vertebrae with gibbus deformity is unusual with pyogenic infections and is seen more often with tuberculous infection. Acute, fulminating sepsis is more common in children, and chronic osteomyelitis is more common in adults. Radionuclide technetium diphosphonate bone scans and MRI are helpful in diagnosis when radiographic findings are subtle. Figure [41-11](#) demonstrates the radiographic differentiation between tumor and infection, and Fig. [41-12](#) shows the MRI findings in similar lesions. Treatment involves identification of the organism by blood cultures or percutaneous CT or fluoroscopically guided disc space aspiration, followed by immobilization and intravenous antibiotics, usually for 6 weeks. Failure to respond promptly to conservative therapy and the presence of neurologic deficits are indications for surgical debridement with anterior bone grafting to preserve stability.

Chest Wall

Causes of chest wall pain include infections such as herpes zoster (shingles), spinal tumors with intercostal nerve root compression, osteomyelitis, and referred pain from visceral disease of the underlying lung, pleura, mediastinum, or heart. Rib fractures and injuries of the costochondral junctions can cause local pain. Inflammation of the costochondral junctions (costochondritis or Tietze's syndrome) is characterized by painful enlargement and local tenderness, often affecting several ribs simultaneously.

Cumulative Trauma Disorders

Cumulative trauma disorders encompass a spectrum of musculoskeletal problems generally related to repetitive loading or injury to tissues, frequently in a workplace setting (Table [41-3](#)). These disorders may cause pain, swelling, restriction of movement, or neurologic symptoms, depending on the specific condition. Treatment generally involves rest of the affected part, anti-inflammatory medications, and, when possible, workplace modifications based on ergonomic considerations.

DISORDERS OF MUSCLE

Anatomy and Physiology

Muscle fibers consist of multiple bundles of individual muscle cells containing actin and myosin, contractile elements that are regulated by neurogenic stimulation of the motor end plates. The macroscopic and microscopic anatomy of muscle is shown in Fig. [41-13](#). Muscle fibers are divided into two major types having different functional properties. Type I slow-twitch fibers have low glycogen and glycolytic enzyme content, but high mitochondrial oxidative phosphorylation activity and myoglobin content. These fibers function in high-repetition and low-load endurance activities. Type II fast-twitch fibers have higher glycolytic capacity, with subtypes (IIA, IIB, IIC, and IIM) differing in the form of myosin present. Type II muscles are generally involved in activities requiring power and speed. The fiber types can be distinguished histochemically and are genetically determined, with different muscles having differing proportions of fiber types, depending on functional requirements. Type I fibers are fatigue resistant and respond to endurance training, and type II fibers are fatigable, respond to resistance training, and are responsible for muscle hypertrophy. Types of muscular contraction and associated terms are defined in Table [41-4](#).

Motor paralysis is defined as loss of voluntary control of muscle contraction. Normal muscle has some resting tone, or tension, which is absent with lower motor neuron lesions, causing flaccid paralysis. Tendon reflexes also are abolished with interruption of the lower motor neuron pathway. Spasticity refers to abnormal increases in muscle tone with passive stretch and is caused by loss of normal central inhibitory control resulting from upper motor neuron lesions. Loss of inhibitory control of tendon reflexes with upper motor neuron lesions also causes hyperreflexia and repetitive reflexive muscle contraction in response to stretch, or clonus. Lower motor neuron lesions, disuse, and immobilization result in muscle atrophy. The clinical grading of muscle strength is shown in Table [41-5](#).

Electrodiagnosis

Stimulation of a peripheral nerve by surface or needle electrodes results in conduction of the stimulus to the motor end plates, with depolarization and resultant muscle contraction. The time for conduction and depolarization, referred to as *latency*, is prolonged in conditions of nerve injury or compression. The nerve conduction velocity, which can be determined by stimulation at two points a known distance apart along the nerve, is useful in localizing peripheral nerve lesions. *Electromyography (EMG)* is the measurement of electrical potentials within muscles using needle electrodes, both at rest and with voluntary contraction. Spontaneous electrical impulses from individual resting muscle fibers known as fibrillation potentials occur in muscle that has been denervated.

Intrinsic Muscle Diseases

Differentiating myopathies from other causes of muscle weakness or paralysis requires consideration of family history, age of onset, and presence of muscle group involvement without common innervation. Biochemical tests for serum aldolase and creatine kinase and muscle biopsy can be helpful in diagnosis.

Muscular Dystrophies

Muscular dystrophies are hereditary disorders resulting in progressive muscular degeneration. Duchenne's muscular dystrophy is inherited as an X-linked recessive disorder manifesting in male children between the ages of 3 and 6 years. The genetic defect has been identified as a deficiency of a protein called dystrophin, which functions as a calcium transport protein. Duchenne's muscular dystrophy is uniformly fatal, usually by the age of 20 years. Late cardiac and respiratory muscle involvement are the causes of death. The disease presents with difficulty standing and walking, and the child uses the upper extremities to push to an upright posture (Gower's sign). Pseudohypertrophy of the gastrocnemius, deltoids, and quadriceps is common. Equinus contractures of the ankles occur early, as does rapidly progressive scoliosis. Other forms of muscular dystrophy include limb girdle and facioscapulohumeral dystrophies, which have an higher age of onset and are inherited as autosomal recessive and dominant traits, respectively. Facioscapulohumeral dystrophy, with upper-extremity and facial weakness and normal life span, is the most benign.

Pathology

Loss of the integrity of the muscle cell membrane results in progressive degeneration of muscle fibers. Fibers become variable in diameter and rounded rather than polygonal in cross-section and appear hyalinized. Fat and fibrous tissue replace the degenerating muscle fibers, with no evidence of muscle regeneration.

Diagnosis

In addition to the clinical features, serum aldolase and creatine kinase levels are elevated, particularly in the Duchenne form. Urinary creatine and amino acid levels are increased, and creatinine level is decreased. EMG shows lower potentials and a polyphasic pattern during voluntary contraction, and muscle biopsy examination confirms the diagnosis.

Treatment

Given the recent identification of dystrophin deficiency as the cause of muscular dystrophy, the disease is a good candidate for treatment by gene therapy in the future. At present, treatment is symptomatic, with exercise to maintain function as long as possible and bracing or surgery to control deformities. Achilles tendon lengthening is occasionally needed for equinus deformities, and early aggressive treatment of progressive scoliosis with curvature over 25 or 30 degrees is indicated, with posterior rodding and segmental wiring (Luque technique) to maintain sitting balance and prevent restrictive pulmonary compromise.

Myotonias

Myotonic dystrophy is an autosomal dominant inherited disorder of muscle, which usually presents in children or young adults with facial and distal extremity weakness. Stimulation of muscle contraction results in prolonged contraction. Myotonia congenita (Thomsen's disease) is a hereditary disorder characterized by difficulty initiating voluntary movement. Diagnosis of these disorders is made clinically and by EMG, the presence of elevated serum muscle enzyme levels, and muscle biopsy examination. Contractures do not develop, and surgical treatment is not necessary. Procainamide, prednisone, and quinine can help patients with myotonia congenita.

Myositis

Inflammatory diseases of muscle can cause pain and weakness and can be associated with viral, parasitic, or bacterial infections, or with collagen vascular diseases (e.g., dermatomyositis, systemic lupus erythematosus, scleroderma, rheumatoid arthritis). Polymyositis can be treated with corticosteroids, and exercise, physical therapy, and appropriate bracing can be helpful. Muscle biopsy examination reveals inflammatory cells and muscle necrosis, and EMG demonstrates fibrillation potentials, distinguishing polymyositis from muscular dystrophies. In elderly patients, an occult primary malignancy sometimes is associated with polymyositis.

Extrinsic Muscle Diseases (Neurologic Disorders)

Poliomyelitis

Poliomyelitis is an infectious viral disease characterized by central nervous system infection with destruction of anterior horn cells in the spinal cord causing flaccid paralysis. Poliomyelitis, formerly one of the most common diseases causing orthopaedic deformities, is now rarely seen in the United States, because of widespread vaccination. Deformities and leg length discrepancies can result. In about one-third of poliomyelitis patients an initial febrile illness occurs, with headache, malaise, and fever, lasting 48 h. The patient may recover or go on to a second acute phase after 4 or 5 days, which also may resolve without paralysis. Headache, fever, neck stiffness, and muscle spasms may occur during the acute phase, which lasts from several days to a week. Paralysis can develop during the third or fourth day, with loss of deep reflexes and muscle spasm. The lower limbs are more frequently involved than the upper extremities. Death can result from bulbar paralysis with respiratory insufficiency. The convalescent phase follows, and some motor improvement may occur for up to 2 years. Treatment is supportive, with physical therapy to maintain joint range of motion and ventilatory support when necessary. Bracing is occasionally needed to assist function of paralyzed extremities. Later, correction of joint contractures, stabilization of flail joints by arthrodesis, and correction of leg length inequalities can be undertaken during the residual phase. Tendon transfers can be useful to improve muscle balance, decrease contractures, or restore functions. Only muscles with a strength grade of 4 or better are suitable for transfer, since one grade of muscle function is lost with transfer (see Table [41-5](#)). A postpolio syndrome has been identified in older adults several decades after disease onset, consisting of increasing weakness in affected muscle groups and fatigue. The cause of this syndrome remains unknown.

Bony Stabilization

For flail joints, when muscles of sufficient strength for transfer are unavailable or bracing is unsuccessful, surgical arthrodesis can be performed. Joint contractures require stretching exercises, splints, or, for severe deformity, surgical release. Epiphysiodesis, the surgical ablation of the growth plate, or bone lengthening/shortening procedures can be used to correct significant limb length inequality.

Treatment of Deformity

Foot deformities are common and can be treated in some cases by orthoses. The most common deformities and functional losses and their corresponding surgical treatments are shown in Table [41-6](#). Many of the procedures listed are useful also for correction of deformities or functional deficits resulting from neurologic injury from other causes, such as traumatic nerve injuries. Surgical correction of foot deformities may be accomplished by extraarticular subtalar arthrodesis (Grice procedure) in skeletally immature patients, or triple arthrodesis (fusion of the subtalar, calcaneocuboid, and talonavicular joints) in adults (Fig. [41-14](#)). Operations must be individualized according to the deficit. Patients with quadriceps muscle paralysis often are able to stabilize the knee in extension for ambulation with the gluteus maximus and gastrocnemius muscles. Muscle transfers about the hip are variably successful, and if hip subluxation or painful degenerative change occurs, arthrodesis may be needed.

Cerebral Palsy

Cerebral palsy (CP), which occurs in about 3 births per 100,000, can be caused by a number of factors, including birth trauma, childhood head injury, anoxic brain damage, and viral diseases such as encephalitis, measles, and cytomegalovirus. Recent epidemiologic data suggest that a relatively small proportion of CP cases result from perinatal events; the majority are idiopathic and are related to defects in central nervous system development. CP is classified as spastic (50 percent), athetoid (25 percent), ataxic (5 percent), rigid (5 percent), or mixed (15 percent). Sixty percent of patients with CP have hemiplegia (ipsilateral upper and lower extremity involvement), with diplegia (both lower extremities) and quadriplegia (all four extremities) being less common. Athetoid types exhibit involuntary repetitive motions that are dysfunctional; ataxia (staggering, broad-based gait) reflects cerebellar dysfunction, and rigidity results from diffuse cerebral involvement usually associated with birth anoxia.

Treatment

Two-thirds of CP patients have an IQ below 70, which compromises treatments that require patient cooperation, such as muscle strengthening, exercise, and gait-training programs. Treatment is directed toward prevention of contractures and surgical correction of deformities that develop from muscle imbalance.

Orthotics

Orthoses are helpful to control varus or valgus deformities of the foot and equinus deformities of the ankle. Long leg braces can help to control knee flexion or valgus deformities and assist in ambulation. Upper-extremity bracing can help to prevent flexion deformities of the wrist and digits and adduction of the thumb. Leg adductor spasticity is common and can cause hip subluxation or dislocation, which becomes painful in about 50 percent of patients who develop it. Abduction bracing can help to prevent this problem.

Surgical Treatment—Lower Extremity

When contractures or deformities progress despite conservative treatment, an operation sometimes is necessary. Surgery also can improve function by improving muscle balance or removing deforming forces. An important adjunct in surgical planning is gait analysis by video recording or dynamic EMG. A major emphasis in the management of the crouched, scissoring gait of patients with CP has been on correction of knee flexion deformities, but evidence suggests that while stance phase of gait is improved, swing phase is not. The rectus femoris muscle has been shown by dynamic EMG to be more active in swing phase, causing inadequate knee flexion and poor toe clearance. Transfer of the rectus femoris to the medial or lateral hamstring muscles has been shown to improve this phenomenon. Excessive length of the patellar tendon also can result from the stretching effect on the rectus femoris when the hamstrings are chronically tight. This can be diagnosed by the presence of patella alta on lateral radiograph or greater passive extension than active extension. Shortening the patellar tendon in conjunction with hamstring lengthening or release can improve gait by correcting these problems.

Equinus deformities of the ankle constitute another common problem that can be corrected by Achilles tendon lengthening. Care must be taken not to overlengthen, as a calcaneus deformity can result, which is difficult to brace. Varus or valgus deformities of the hindfoot can be corrected by a Grice procedure, or by triple arthrodesis at skeletal maturity (see Fig. 41-14). Adduction deformities of the hips are treated by adductor tenotomies along with neurectomy of the anterior branch of the obturator nerve, and if hip subluxation is present varus and derotational osteotomies of the proximal femur with internal fixation may be indicated (Fig. 41-15).

Surgical Treatment—Upper Extremity

An operation on the shoulder is almost never necessary. The most common problems are related to wrist and finger flexor spasticity and thumb adduction deformities. Wrist/finger flexion deformities have been divided into three groups: Type I or mild deformities (digits extend with wrist in less than 20 degrees of flexion) rarely require surgical intervention. Type II or moderate deformities (digits extend with wrist in less than 50 degrees of flexion) benefit from proximal wrist/finger flexor release with dorsal flexor carpi ulnaris transfer. Type III or severe flexion deformities require distal wrist and finger flexor tenotomies, with transfer of the sublimis to the distal profundus tendons. Proximal row carpectomy or wrist arthrodesis occasionally is indicated. Release of the adductor and interphalangeal fusion or tendon transfers can correct thumb-in-palm deformities.

Spinal Deformity

Neuromuscular scoliosis with a C-shaped progressive curve is common in severe cases of CP. Bracing can be of some value for limited or nonprogressive curves. Progression can lead to impaired sitting balance, pelvic obliquity, and ischial decubitus ulcers, and therefore correction of scoliosis with anterior and/or posterior spinal instrumentation and fusion is commonly indicated. Generally, posterior instrumentation with rods and segmental fixation (Luque procedure) from the upper thoracic spine to the sacrum is the preferred approach. With severe rotational deformities in the lumbar spine, anterior instrumentation also may be required.

Myelodysplasia (Spinal Dysraphia)

Myelodysplasia refers to a developmental defect in the vertebral column associated with a neurologic deficit. *Spinal dysraphia* refers to defects involving failure of fusion of midline structures and may be present without cord involvement (spina bifida occulta), or with a myelomeningocele, a neural tube defect at the level of the lesion. Eighty percent of patients have associated hydrocephalus, and paralysis is generally present below the level of the defect. Antenatal diagnosis of neural tube defects is possible using amniocentesis with assay of alpha-fetoprotein. Supplementation with folate during the first trimester of pregnancy can dramatically reduce the incidence of neural tube defects.

Meningoceles (cystic enlargement of the lower meninges) or myelomeningoceles (cystic enlargement of the meninges and intradural contents) are treated by early closure and shunting for hydrocephalus. Prognosis depends on the degree of neurologic involvement and corresponding level of lower-extremity paralysis and disruption of bowel and bladder function. Patients with lesions below the L4 level will usually be ambulatory, although lower-extremity deformities such as talipes equinovarus and hip subluxation can occur, requiring surgical correction. Additionally, lack of sensory function makes pressure sores a common recurring problem.

Orthopaedic Management

Function is strongly dependent on the level of the neurologic lesion. Patients with lesions below L3 or L4 usually can ambulate but may require orthoses. Maintenance of a plantigrade foot is essential, and 50 percent of patients have foot deformities at birth. Lower-extremity deformities correspond to the level of spinal involvement. Appropriate bracing to minimize progressive contractures and allow ambulation when possible is indicated. Young children with flaccid paralysis of the lower extremities can ambulate and maintain an upright standing position in a parapodium. The energy costs of ambulation with this device are so high, however, that older children and adults are unable to use it and generally are confined to wheelchair. A reciprocating gait orthosis allows ambulation in older patients with meningomyelocele or paralysis secondary to spinal cord injury. Custom-molded orthoses lined with pressure-distributing materials such as Plastizote are important in the prevention of pressure sores from lack of sensation.

Management of paralytic deformities is similar to that with the flaccid paralysis of poliomyelitis. Talipes equinovarus can be quite severe and is treated by posteromedial release or talectomy in severe or recurrent cases. Contractures frequently require release and tendon transfers to remove deforming forces when appropriate muscles are available. The Grice procedure is corrective for valgus hindfoot deformities, and the triple arthrodesis is appropriate for correction of hindfoot deformities in older children (see Fig. 41-14). Maintenance of concentric reduction of the hips can necessitate proximal femoral osteotomies or acetabular osteotomies in severe cases. Spinal deformities include severe kyphosis at the level of the lesion, or scoliosis. Virtually all patients with a functional disturbance of the spinal cord at the level of L3 or higher will develop scoliosis requiring surgical intervention, as compared to approximately 60 percent at the L4 level. Lumbar kyphosis is seen in 8 to 15 percent of patients and is almost always progressive. Kyphectomy with instrumentation and fusion can be helpful, and scoliosis is corrected by posterior rodding with segmental fixation and fusion, occasionally in addition to anterior release and fusion. In addition, tethering of the spinal cord can occur with growth, usually presenting as a worsening of the neurologic deficit. This can necessitate surgical release of the filum terminale.

Degenerative Neurologic Diseases with Skeletal Deformity

Peroneal Muscle Atrophy (Charcot-Marie-Tooth Disease)

This is an inherited autosomal recessive or dominant trait leading to a degenerative neuropathy, manifested initially in the distribution of the peroneal nerve. Patients develop cavus and varus deformities of the feet and a dropfoot gait. Intrinsic atrophy in the hands also may be apparent. Correction of cavus deformities by midfoot osteotomy and clawtoe deformities by interphalangeal fusions and extensor tendon recession are helpful.

Friedreich's Ataxia

This familial disease begins in childhood, involving the spinocerebellar tracts, corticospinal tracts, and posterior columns. Patients have progressive gait and speech disturbances and scoliosis and foot deformities. Intervention for foot deformities or scoliosis may be indicated early in the course of the disease to maintain ambulation.

Syringomyelia

This degenerative condition of the spinal cord involves destruction of neurons in the central portion of the cord with formation of a cystic cavity, or syrinx. Onset of symptoms usually is in the second or third decade. The intrinsic muscles of the hand are involved initially, followed by progressive loss of motor and sensory function in the upper and lower extremities. Orthopaedic treatment entails bracing for prevention of contractures, or arthrodesis of neuropathic joints. Laminectomy occasionally is indicated to relieve expanding cavities of the spinal cord.

Multiple Sclerosis

This is a progressive demyelinating disease of the central nervous system of unknown pathogenesis. The disease has a variable course, with exacerbations that are sporadic and can be ameliorated by use of ACTH or corticosteroids. Orthopaedic problems include contractures of the ankle, knee, and hip and scoliosis. Treatment involves range-of-motion and stretching exercises, splinting, and occasionally surgical release of contractures. Spinal instrumentation and fusion rarely are indicated. Death can result from progressive muscle weakness causing respiratory failure.

Orthopaedic Management of Stroke

Rehabilitation of stroke patients should begin as early as possible, usually within a few days after the cerebrovascular event. Initially patients exhibit flaccid paralysis, which later can give way to spasticity and the development of contractures. Early physical therapy can maintain mobility and prevent deformity and decubitus ulcers. Persistent spasticity can be temporarily ameliorated by peripheral nerve blocks with phenol or alcohol, but more permanent control generally requires release or tenotomy of the affected muscle groups. The most common deformity is foot and ankle equinus, which can be corrected by Achilles tendon lengthening. Similarly, flexion contractures of the knee and adduction contractures of the hip are corrected by surgical release. Paralysis of the shoulder girdle musculature can lead to a painful “frozen shoulder,” which can impede rehabilitation. Active-assisted exercises may help to maintain range of motion, and flexion deformities of the fingers and wrist can be improved by flexor slide, a release of the flexor muscles in the proximal forearm.

POSTURE

Disturbances in Gait

Gait disturbance of mechanical origin must be distinguished from that of neurologic causation. Neurologic gait disturbances include ataxia, hemiplegia, and spasticity. Ataxia, a wide-based, uncoordinated gait, can result from cerebellar lesions, Guillain-Barré syndrome, Friedreich's ataxia, or defects in peripheral sensation and proprioception such as tabes dorsalis or vitamin B₁₂ deficiency. Spastic gait, common in cerebral palsy, is characterized by a crouched position with scissoring of the legs. With hemiplegic gait, the patient appears to drag the affected extremity.

Mechanical Disorders

Gait disturbances can be produced by numerous abnormalities of joints, including congenital hip dysplasia or dislocation, slipped capital femoral epiphysis, avascular necrosis of the hip, Legg-Calvé-Perthes disease, and knee joint abnormalities (e.g., arthritis, osteochondritis dissecans, genu valgum, genu varum, meniscal injury). Congenital foot and ankle deformities, such as talipes equinovarus (clubfoot), and limb length discrepancies also can cause a limp. Any condition causing pain on weight bearing in a lower extremity can cause an antalgic gait in which the patient “short steps” on the affected extremity, minimizing the stance phase of gait. Pain in the hip joint can cause an abductor limp, wherein the patient lurches or leans toward the affected side during stance phase to move the center of gravity of the body over the hip, decreasing the abductor force across the joint. With abductor paralysis, a Trendelenburg gait results, in which the pelvis tilts toward the opposite side during stance phase on the affected extremity because of the inability of the abductors to maintain a level pelvis. This also can result from ineffective abductor contraction in congenital hip dislocation or coxa vara. Use of a cane in the opposite hand can reduce the force across the hip joint by a factor of five, the number of times body weight the abductors must contract to maintain the trunk and pelvis level during one-legged stance. In patient with hip problems the use of a cane thus can dramatically improve gait and reduce hip pain when present. Generally, orthopaedic management of a limp requires appropriate diagnostic evaluation and treatment directed toward the cause.

Limb Length Discrepancy

Limb length discrepancies may be secondary to trauma with growth plate injury, shortening of a bone, or growth stimulation from injury. Additionally, congenital aberrations in growth or radiation treatment for tumors can lead to significant and progressive limb length discrepancies in children. Asymmetric growth plate disturbances can produce angular deformities. Orthopaedic treatment includes use of a shoe lift for discrepancies of up to 1 or 2 cm, epiphysiodesis to stop the growth of the longer limb for larger discrepancies, or limb-lengthening or limb-shortening procedures. Growth charts (Moseley chart) are used to plot the growth of the limbs as measured by calibrating radiographs called scanograms versus time. Comparison to normal data allows prediction of the appropriate timing for surgical disruption of the growth plates (usually distal femoral and proximal tibial plates) so that the remaining growth in the short limb will allow leg length equalization at skeletal maturity. Disadvantages include loss of stature, the need to operate on the well leg, and the limitations in the amount of length equalization that can be attained in this manner. For modest discrepancies, however, this is a well-accepted and useful method.

For severe limb length differences a new method of limb lengthening, called *distraction osteogenesis* or *callotasis*, has become popular. This method, first implemented by Ilizarov

in the Soviet Union, relies on intramedullary membranous osteogenesis after cortical osteotomy and gradual distraction with an external fixator using small Kirschner wires (K-wires) under tension and circumferential ring supports (Fig. 41-16). The bone can be lengthened by 1 mm daily, and the multiplanar fixation device allows simultaneous correction of angular and rotational deformities. Care is taken at the time of the corticotomy not to disrupt the medullary bone, and the procedure is done through small incisions without any periosteal stripping. Lengthenings of up to 10 cm are obtainable with this method, and more than one bone can be lengthened in an extremity. The method is also useful in the treatment of nonunions.

Spinal Deformities

The spine is characterized by a series of curves that aid its physiologic functions. The cervical and lumbar spine normally exhibit lordosis, while the thoracic spine has a normal kyphosis. The intervertebral discs serve to maintain flexibility of the spine and the function of shock absorption. The discs consist of a tough, fibrous peripheral component, the annulus fibrosus, and a gelatinous central portion, the nucleus pulposus. With age, loss of water content occurs with concomitant loss of disc height.

Kyphosis

Kyphosis is an increase in the normal posterior convexity of the thoracic spine involving a number of vertebral bodies. A gibbus deformity is an acute kyphotic angular deformity that may be congenital, posttraumatic, or secondary to tumor or infections such as tuberculosis.

Adolescent Kyphosis

Postural kyphosis, or “round shoulders,” occurs in children, occasionally in association with muscular coordination problems. With time, this habitual kyphotic posture can become fixed. Treatment is with extension exercises, and bracing is unnecessary. A discogenic form of adolescent kyphosis that is progressive, known as Scheuermann's disease, also occurs. This disorder is characterized by abnormalities in the growth plates of the vertebral bodies and herniations of disc material into the vertebrae (Schmorl's nodes). Scheuermann's kyphosis tends to be progressive and is generally treated by extension exercises and Milwaukee bracing for more severe deformities. The need for surgical treatment is unusual, but refractory cases with back pain can benefit from staged anterior and posterior fusion with posterior instrumentation.

Osteoporotic Kyphosis

In osteoporotic kyphosis, cumulative effects of compression fractures or anterior wedging of multiple vertebral levels as a result of mechanical failure of osteopenic bone lead to progressive kyphosis in the thoracic spine. Most commonly this is seen in women with postmenopausal osteoporosis. The discs often bulge into the weakened vertebral end plates. Compression fractures can occur as discrete, acute events following minor trauma, or with a more insidious onset and progression, probably the result of multiple microfractures. Back pain is the presenting complaint, and patients may experience loss of height and a roundback deformity (dowager's hump). Neurologic deficit beyond intercostal radicular pain almost never develops despite severe degrees of deformity. Multiple myeloma and metastatic carcinoma also can cause vertebral compression fractures and must be ruled out, as well as osteomalacia, renal osteopathy, and other metabolic derangements. Treatment of acute fractures is with rest, analgesics, and an extension orthosis, followed by extension exercises and therapy for the underlying cause of the osteopenia when possible (see Metabolic Diseases, below).

Scoliosis

Any lateral deviation or curvature of the spine is referred to as scoliosis. Because of rotation of the spine, which is almost always a component of the deformity, the abnormality is best demonstrated in physical examination in forward flexion, in which asymmetry of the paraspinous region or rib prominence will be most apparent. Scoliosis has been classified as shown below (modified from Ponseti and Freedman, and R.B. Winter):

- NonstructuralPostural scoliosis Nerve root irritation
- 1

Disc herniation
- 1

Inflammatory
- Muscle spasm secondary to injury Structural scoliosisIdiopathic
- 1

Cervicothoracic
- 1

Thoracic
- 1a

Infantile—age of onset under 3 years
- Resolving
- Progressive
- 1a

Juvenile—age of onset 4–9 years
- 1a

Adolescent—age of onset 10 years to skeletal maturity
- 1

Thoracolumbar
- 1

Lumbar
- 1

Combined thoracic and lumbar
- Osteopathic
- 1

Congenital vertebral anomalies
- 1

Thoracogenic after thoracoplasty or empyema
- 1

Osteochondrodystrophy
- 1a

Mucopolysaccharidosis
- 1a

Diastrophic dwarfism
- 1a

Spondyloepiphyseal dysplasia
- 1a

Multiple epiphyseal dysplasia
- 1

Degenerative disc disease
- 1

Postirradiation
- Neuropathic
- 1

Congenital
- 1

Postpoliomyelitis
- 1

Neurofibromatosis
- 1

Syringomyelia
- 1

Charcot-Marie-Tooth disease
- 1

Friedreich's ataxia
- 1

Cerebral palsy
- 1

Spinal muscular atrophy
- 1

Meningomyelocele
- 1

Dysautonomia (Riley-Day syndrome)
- Myopathic

1 Arthrogryposis multiplex congenita

1 Muscular dystrophies

1 Myotonic dystrophy

Metabolic

1 Osteogenesis imperfecta

1 Ehlers-Danlos syndrome

1 Marfan syndrome

1 Homocystinuria

Tumors

1 Osteoid osteoma

1 Eosinophilic granuloma

1 Spinal cord tumors

Traumatic

Postural Scoliosis

Postural scoliosis occurs in adolescent girls as a characteristically nonprogressive mild left thoracolumbar curve without vertebral rotation that corrects in recumbency. A similar flexible scoliosis can occur with limb length discrepancy and corrects with use of a lift on the short side. Treatment of postural scoliosis is generally with exercises only.

Congenital Scoliosis

Congenital scoliosis occurs as a result of developmental anomalies during embryonic resegmentation of the spine. Unilateral hemivertebrae, fusions of segments, and posterolateral bony bars are the most common abnormalities. When progression of the scoliosis develops, most frequently with unilateral bony bars, fusion of the involved area is necessary. Usually instrumentation is not used, as it increases the risk of paraplegia in these patients. About 20 percent of patients with congenital scoliosis have asymptomatic renal anomalies, and these must be evaluated by intravenous pyelography or abdominal CT or ultrasound studies.

Neuromuscular Scoliosis

Curves with neuromuscular causes such as spinal muscular atrophy, meningomyelocele, cerebral palsy, muscular dystrophies, poliomyelitis, or traumatic paralysis tend to be long, thoracolumbar C-shaped deformities with a high propensity for curve progression. Prognosis depends on the level of the lesion as well as on the age of the patient at the time of onset, with a greater likelihood of progression in younger patients and higher-level lesions. With neurofibromatosis, the curve can be sharp, relatively short, and associated with cutaneous nevi, neurofibromas, or café-au-lait spots. In addition, there may be scalloping of the vertebral bodies and narrowing of the proximal ribs because of neurofibromas of the nerve roots; these curves also have a strong tendency to progress.

Infantile and Juvenile Idiopathic Scoliosis

Infantile scoliosis presents from birth to age three, while juvenile scoliosis may present up to age ten. The cause is unclear, and the tendency for curve progression is variable. The difference between the rib–vertebral angles on the convex and concave sides of the curve has been correlated predictively with progression, with values greater than 20 degrees indicating a high probability, and less than 10 degrees, a low probability. Early management of the infantile or the early-onset juvenile patients consists of casting in patients up to the age of 4 or 5 years, and bracing thereafter. In patients with progression despite appropriate casting or bracing, surgical intervention is indicated. Generally, anterior and posterior surgery is necessary, with anterior fusion over four or five apical segments only. Subsequent posterior instrumentation can be undertaken without fusion using expandable rods with repeat distractions performed annually until the child is old enough for posterior fusion. In the late-onset juvenile cases about 50 percent will require surgical intervention because of progression with bracing alone. Surgery can be delayed until curves reach a magnitude of 55 to 60 degrees, in an effort to allow maximal spinal growth before fusion. Anterior and posterior fusion is necessary.

Idiopathic Scoliosis

Pathology

The underlying cause of idiopathic scoliosis is unknown. There is a familial tendency, particularly in females. Abnormalities of vestibular function have been demonstrated in scoliotic patients as well as defects in posterior column functions of vibratory and proprioceptive sense. Abnormalities in collagen synthesis and proteoglycan content in the intervertebral discs also have been identified in scoliosis but are thought to be secondary changes. Thus a defect in neurologic function has been hypothesized as a possible cause. Progression of the curvature occurs during growth in the majority of cases. The most important factors in determining the risk of progression are age of onset, location of the primary curve, and skeletal or physiologic age. Curves tend to stabilize at skeletal maturity (determined radiographically by fusion of the iliac apophyses), but progression in adulthood of approximately 1 degree per year can be expected.

Clinical Manifestations

Scoliosis is rarely associated with back pain and usually presents as an asymptomatic deformity characterized by asymmetry of the back and chest, accentuated by forward bending. In older adults degenerative changes and pain can occur. Severe scoliosis can result in restrictive pulmonary disease and cor pulmonale. The most common type of idiopathic scoliosis is a right thoracic curve, usually seen in girls (Fig. 41-17). Patients frequently have asymmetrical height of the pelvis when standing, or asymmetrical shoulder height or scapular prominence. Forward bending causes prominence of the paraspinous or rib area on the convex side of the curve. Subtle degrees of asymmetry can be visualized without radiographs by using a Moiré pattern and appropriate illumination to visualize topography of the back; this also has been used as a screening tool.

Radiographic Evaluation

Full-length standing anteroposterior (AP) and lateral radiographs of the thoracic and lumbar spine are necessary, along with supine lateral bending radiographs to assess the flexibility of the curve(s). The primary curve is defined as the longest curve with the greatest degree of angulation and/or the least flexible curve. The degree of angulation (Cobb angle) is measured between perpendicular lines drawn to the end plates of the superior and inferior vertebrae in the curve (see Fig. 41-17).

Treatment

The goal of treatment is to prevent worsening of the deformity. In this regard, careful periodic follow-up is essential to monitor curve progression. For idiopathic scoliosis, nomograms have been developed to predict the risk of progression on the basis of magnitude of curve, skeletal maturity, and age. Observation is the mainstay of conservative treatment, and general spinal exercises to maintain flexibility are advised. For curves greater than 30 degrees with documented progression, bracing with a Milwaukee-type brace or a thoracolumbosacral orthosis is indicated. An alternative but controversial treatment is electrical muscle stimulation on the convexity of the curve, which is used by the patient at night. For curves over 40 degrees with evidence of progression, fusion with instrumentation is indicated. In general, Harrington and Luque rodding posteriorly with sublaminar or transspinous process wiring (segmental stabilization) are the most commonly used techniques. Cotrel-Dubousset posterior instrumentation, which has cross-linkage between the rods, affording greater stability and better preservation of anatomic sagittal contours, also may be used (Fig. 41-18). Spinal cord monitoring (somatosensory evoked potentials) often is used during surgical correction of spinal deformities to decrease the incidence of neurologic deficit. Electrical stimulation of the lower extremities transmits impulses through the spinal cord, causing changes on the electroencephalogram (EEG). Computerized equipment analyzes and monitors the EEG patterns and can give early warning of conduction abnormalities in the cord, allowing preventive modification of the surgical procedure intraoperatively.

Anterior (Dwyer or Zielke) instrumentation occasionally is used in treatment of lumbar or thoracolumbar curves, particularly those of neuromuscular origin, or in adults. Most neuromuscular curves are treated surgically fairly early because of the strong propensity for progression; generally a long fusion from the upper thoracic spine to the sacrum is needed with segmental (Luque) instrumentation. The selection of the extent of the fusion is determined by identifying the vertebra in neutral rotation at the upper and at the lower end of the curve, judged by symmetry of the pedicles and midline position of the spinous process on an anteroposterior radiograph. With Harrington rodding and spinal fusion, postoperative ambulatory immobilization in a body cast or orthosis is necessary, while the more rigid fixation with the Cotrel-Dubousset or Luque technique can obviate the need for this.

Knee Deformities

Genu Valgum

Angular deviations about the knees often are physiologic and rarely require treatment. From birth until about 3 years of age, children have a physiologic varus orientation of the knees (bowleg appearance), which generally corrects spontaneously and converts to a valgus orientation from ages 4 to 8 (Fig. 41-19). The normal adult valgus angle of the knees of 6 degrees is obtained during these years. Physiologic angulations in children of as great as 30 degrees of varus and 20 degrees of valgus can spontaneously correct with growth. Thus the most common treatment approach is parental reassurance and clinical follow-up.

Occasionally progression of the apparent deformity occurs. Causes include vitamin D deficiency (rickets), rheumatoid arthritis, and trauma as well as idiopathic causes. Genu valgum also may be seen in some genetic diseases, such as Morquio's syndrome (mucopolysaccharidosis type IV) and Engelmann's (or Camurati-Engelmann) disease (diaphyseal dysplasia). With progressive deformity, treatment with long leg braces or corrective osteotomy of the proximal tibia or distal femur may be needed.

Genu Varum

Genu varum occasionally fails to resolve spontaneously in infancy, and progression can occur. Physiologic genu varum in an infant is demonstrated in Fig. 41-19. The most common form of genu varum is Blount's disease, a disturbance of the medial proximal tibial growth plate that is seen in infantile and adolescent forms. Persistent excessive medial weight bearing is thought to traumatize the medial growth plate, slowing medial growth and causing the deformity to worsen. Measurement of the metaphyseal-diaphyseal angle

provides a means of diagnosis (Fig. 41-20). An angle greater than 11 degrees is consistent with a diagnosis of Blount's disease. In cases with significant deformity or pain, osteotomy of the tibia may be indicated. Stapling of the lateral growth plate is an alternative but less predictable method. Genu varum also may be associated with renal osteodystrophy or achondroplasia, or it may follow trauma to the growth plates about the knee.

Foot and Ankle Deformities

Flatfoot (Pes Planus)

Hypermobile or flexible flatfoot is a benign congenital condition that ordinarily is not painful and does not require treatment. Flexibility can be demonstrated by having the child stand on the toes, which usually will reconstitute a normal arch contour, particularly if the flatfoot is in association with Achilles tendon tightness. Treatment with an arch support or orthosis is indicated only for foot pain, excessive shoe wear, or family history of symptomatic flatfoot.

Peroneal Spastic Flatfoot

Although peroneal spastic flatfoot may be caused by inflammatory and traumatic involvement of the tarsal joints, the most common cause is abnormal congenital fusion of tarsal bones, referred to as *tarsal bar* or *tarsal coalition*. The bar may be cartilaginous, particularly in younger children, and may be difficult to visualize on routine radiographs. The most common coalition is the calcaneonavicular bar, followed by talocalcaneal and calcaneocuboid fusions (Fig. 41-21). The oblique radiograph will best demonstrate a calcaneonavicular bar, and an axial radiograph of the hindfoot is most useful for identifying talocalcaneal coalition. CT scans are also quite helpful when plain-film radiographs are negative. Conservative treatment with anti-inflammatory medications and orthotics can be helpful, but if symptoms persist, resection of the bar with fat or silicone interposition sometimes is necessary. In older adolescents and adults triple arthrodesis may be indicated, especially if degenerative arthritis is present in the adjacent tarsal joints.

Acquired Flatfoot

Flatfoot can be caused by osseous disruption (trauma or infection), ligamentous or tendinous rupture (foot sprain or posterior tibialis tendon rupture), or neuromuscular disorders (poliomyelitis, cerebral palsy, or Charcot neuropathic arthropathy). Acquired flatfoot involves medial and plantar deviation of the head of the talus, frequently accompanied by valgus deformity of the hindfoot (pes planovalgus). Treatment involves supportive measures and orthotics. In cases with severe deformity, pain, or secondary degenerative changes, options for surgical intervention include Grice subtalar or triple arthrodesis, calcaneal or midfoot osteotomy, and posterior tibialis advancement.

Contracture

The term *contracture* implies a permanent shortening and loss of flexibility of muscles, joints, tendons, or fascia. Contractures can be congenital or acquired. Examples of congenital contractures include talipes equinovarus (clubfoot), torticollis, and arthrogryposis multiplex congenita. Acquired contractures of joints can result from periarticular trauma, arthritis, muscle imbalance (as with cerebral palsy), burns, muscle injury, or idiopathic conditions such as Dupuytren's contracture of the palmar fascia. Contractures of muscles or joints also may occur postoperatively and can be avoided by appropriate strengthening and range-of-motion exercises with the assistance of a physical therapist. The use of continuous-passive- motion machines postoperatively also can help to prevent joint contractures; these have been particularly useful for the knee and also are used occasionally after surgery on the hip or elbow.

Ischemic Contractures (Volkmann's Contracture; Compartment Syndrome)

Volkmann first described contractures of the forearm muscles that followed tight bandaging after an elbow fracture in 1875. The contractures result from ischemic muscle necrosis, now recognized as the phenomenon called *compartment syndrome*. All muscle groups in the extremities are bounded by tough, fibrous fascial envelopes called *compartments*. The fascia is relatively unyielding to muscle swelling, which can result from fractures, bleeding, surgery, blunt trauma, or prolonged ischemia (Fig. 41-22 A). Elevated pressure within a compartment further compromises arterial inflow, and progressive ischemia ensues. Nerves and muscles within the compartment suffer ischemic injury that can rapidly become irreversible, and severe pain and loss of neuromuscular function occur.

The cardinal signs of compartment syndrome include (1) pain out of proportion to the injury; (2) pain on passive stretch of involved muscles; (3) paresthesias, hypesthesia, or loss of motor function, and (4) diminished or absent pulses, along with coolness or pallor of the extremity consistent with decreased perfusion. Circumferential dressings such as tight bandages or casts can contribute to the development of a compartment syndrome by compromising perfusion, and they must be split or removed at the first signs of a compartment syndrome. If symptoms fail to resolve immediately, compartment pressure measurements must be made. Direct measurement of intracompartmental pressures using a needle or slit catheter with a mercury or electronic manometer is the method of definitive diagnosis of compartment syndrome (Fig. 41-22 B). Normal compartment pressures are less than 10 mmHg. Sustained pressures above 30 mmHg indicate impending compartment syndrome, and sustained pressures above 40 mmHg indicate the definite presence of a compartment syndrome.

If untreated, the sustained progressive ischemia causes muscle necrosis and loss of voluntary muscle function. The muscle is replaced by fibrous connective tissue, which leads to severe contractures of the joints powered by the involved muscles. Supracondylar fractures of the humerus and tibial fractures are common causes of compartment syndromes. An exercise- induced compartment syndrome occurs with exercise, causing pain and paresthesias, but is self-limited and resolves after the activity is discontinued. Diagnosis is with postexercise compartment pressure measurements, and treatment is by elective surgical fasciotomy. Contractures and permanent muscle damage generally do not occur in this form of compartment syndrome.

Compartment syndrome is an orthopaedic emergency and requires immediate surgical fasciotomy to prevent irreversible nerve and muscle damage. Severe damage can occur within a few hours. Arteriography often is indicated to rule out arterial damage in cases of posttraumatic compartment syndrome. The compartment pressures that will be tolerated by the extremity without irreversible damage depend on the perfusion pressure; hence with relative hypotension a lower level of compartment pressure will necessitate surgery, and slightly higher levels can be tolerated in a hypertensive patient. Compartment pressures less than 20 mmHg below diastolic blood pressure necessitate surgical intervention in hypotensive patients. The presence of symptoms with compartment pressures above normal is an indication for surgery. The skin and fascia must be left open, and pressures within all compartments in the extremity must be measured and all involved compartments thoroughly released.

EPIPHYSEAL DISORDERS (OSTEOCHONDROSES)

The epiphysis comprises the cartilaginous end of a long bone between the growth plate (physis) and the articular surface. Epiphyses develop secondary centers of ossification, which expand from a central location to replace the cartilaginous epiphysis with trabecular bone through the process of endochondral ossification. At maturity, the only portion of the epiphysis that remains as cartilage is the articular surface. During childhood, a number of disorders of the epiphysis can occur, the etiology of which remains poorly understood. There are several forms of epiphyses, including articular epiphyses, traction epiphyses (apophyses), and atavistic epiphyses. Disorders of the epiphyses were formerly referred to as osteochondritis, but the lack of inflammatory changes histopathologically led to the use of the term *osteochondrosis* as a more accurate description of the various derangements of growth or ossification occurring in epiphyses.

Osteochondroses affect primary and secondary ossification centers. Primary ossification centers (and associated disorders) include:

Scaphoid (Preiser's disease)

Lunate, in adults (Kienböck's disease)

Patella (Köhler's disease)

Talus (Mouchet's disease)

Tarsal navicular (Köhler's disease)

Secondary ossification centers include:

Vertebral epiphysis (Scheuermann's kyphosis)

Humeral head (Hass's disease)

Capitellum (Panner's disease)

Radial head (Brailsford's disease)

Pubic symphysis (Van Neck's disease)

Ischiopubic junction (Oldsberg's disease)

Femoral head (Legg-Calvé-Perthes disease)

Patella (Sinding-Larsen-Johansson syndrome)

Tibial tubercle (Osgood-Schlatter disease)

Calcaneus (Sever's disease)

Metatarsal head (Freiberg's disease)

Cause

The etiology of osteochondrosis remains controversial. In general, vascular disturbance, possibly secondary to trauma, is thought to result in avascular necrosis of the involved

epiphysis. In some osteochondroses, such as Osgood-Schlatter disease, there is clear evidence that a traction injury to the epiphysis initiates the pathologic changes. There is some evidence also for an underlying genetic defect in epiphyseal cartilage as a predisposing factor.

Clinical Manifestations

The condition may be unilateral or bilateral, with gradual onset and sometimes a history of antecedent trauma. Pain, joint effusion, limp, limitation of range of motion, and muscle spasm may be present, but the symptoms are often mild.

Radiographic Findings

Osteopenia of the epiphyseal ossification center is present, with areas of increased density. With more severe involvement, fragmentation of the ossification center and deformity and flattening of the epiphysis can occur. The areas of increased density correspond to necrotic bone. The course of the disease usually involves spontaneous revascularization and reconstitution of the ossification center, but there may be residual deformity or growth disturbance.

Legg-Calvé-Perthes Disease

Osteochondrosis involving the hip is primarily seen in boys between the ages of 5 and 9 years and is bilateral in about 10 percent of patients. The pathology is that of osteonecrosis of the ossification center of the femoral epiphysis, but the cause is unknown. The disease has been reported in 1 to 3 percent of patients following transient synovitis of the hip, a disorder clinically mimicking septic arthritis in children but self-limited and resolving spontaneously. In most cases there is no identifiable antecedent event. Epidemiologically, some risk factors have been identified, including delayed bone growth, low socioeconomic status, breech delivery, and being a child of older parents. Coexistence with other osteochondroses such as Scheuermann's disease has been reported and suggests the possibility of an underlying systemic cartilaginous defect. Attention deficit hyperactive disorder has been identified in about one-third of patients with Perthes disease, providing evidence of an underlying genetic predisposing factor. Patients present in the prodromal stage with a limp and hip and/or referred knee pain and loss of motion in the affected hip secondary to muscle spasm initially but later due to deformity of the femoral head. During the active phase, radiographic progression of the disease occurs, but symptoms and muscle spasm resolve or become intermittent. Patients exhibit limitation of abduction and internal rotation, initially due to spasm and later due to flattening and deformity of the femoral head (coxa plana). With the restoration phase the clinical symptoms resolve and the ossific nucleus reconstitutes, although some permanent femoral head deformity and restriction of abduction and hip rotation occasionally remain.

Radiographic Findings

The Caterall classification of the radiographic extent of epiphyseal involvement is the staging system most widely used for prognostication. This staging system has been criticized as insufficiently prospective, since accurate radiographic classification often is not possible until 6 to 8 months after diagnosis. Caterall group I disease exhibits absence of metaphyseal changes or sequester formation, with involvement of less than half of the anterior epiphysis. Group II disease shows sequester involving the anterior half of the epiphysis, with anterolateral metaphyseal reaction and subchondral fracture line. Group III disease involves more than half of the epiphysis with sequester and metaphyseal reaction, and group IV exhibits whole-head involvement with diffuse metaphyseal reaction.

A new classification based on the lateral pillar of the femoral head has been proposed that demonstrates greater predictive value and less interobserver variability (Fig. 41-23). For children with a bone age less than 6 years and lateral pillar involvement of stage A or B, the outcome does not appear to be affected by treatment (i.e., containment). For skeletally older children with stage B involvement, containment by abduction bracing, or femoral or pelvic osteotomy improves outcome.

Waldenström has described four radiographic stages of the disease: (1) failure of ossification center growth and joint space widening; (2) fragmentation with areas of radiolucency; (3) reossification with areas of new bone formation; and (4) the healed phase with reconstitution of the epiphysis. The fragmentation stage in a patient with group III involvement is shown in Fig. 41-24.

Studies of the natural history of the disease indicate good, fair, and poor results in about one-third of patients each. Patients with partial (Caterall I and II) involvement have a much better prognosis than those with more extensive involvement (Caterall III and IV). Both clinical and radiologic "at risk" signs that are correlated with a worse prognosis have been identified. Clinical signs include age over 6 years, obesity, progressive loss of hip motion, and adduction contracture. Radiologic signs include defective ossification in the lateral epiphysis, calcification of the epiphysis lateral to the acetabulum, diffuse metaphyseal reaction, lateral subluxation of the hip, and a horizontal growth plate.

Treatment

The differential diagnosis of the disease must include consideration of other disorders that can have similar radiographic changes and clinical presentations, such as hypothyroidism, spondyloepiphyseal dysplasia, multiple epiphyseal dysplasia and sickle cell disease. Treatment is based on containment of the diseased epiphysis within the acetabulum until reossification is complete, which affords the best chance of obtaining a spherical femoral head. The skeletal age and degree of lateral pillar involvement of the epiphysis are important factors in identifying those patients most likely to benefit from containment therapy. For femoral heads in which the diseased portion of the epiphysis is contained within the acetabulum in both neutral and abducted positions, observation alone is needed and the prognosis is excellent. For hips with epiphyseal involvement lateral to the acetabulum, but which are contained with abduction, ambulatory treatment in a walking abduction brace for the duration of the healing (12 to 18 months) is generally recommended. For those few patients with noncontainable epiphyseal involvement with the hip abducted, surgical intervention is necessary to achieve containment. Surgical options include varus osteotomy of the proximal femur or osteotomy of the pelvis to extend acetabular coverage over the femoral head (Salter or Chiari osteotomy) (see Fig. 41-28).

Outcome depends on the staging, final sphericity of the femoral head, and congruity with the acetabulum. Symptoms of degenerative arthritis may present in the fourth to sixth decades in patients with residual femoral head deformity.

Osgood-Schlatter Disease

Osgood-Schlatter disease usually presents in patients 13 to 15 years of age whose history may reveal a precipitating injury. Avulsion or fatigue fracture of the epiphysis is thought to be the underlying cause, with secondary defective ossification and prominence of the tubercle. Pain that is aggravated by quadriceps contraction and tenderness and enlargement of the tibial tubercle are the usual presenting complaints. Radiographs may be normal, or irregularity and fragmentation of the ossification center of the tibial tubercle may be present (Fig. 41-25).

Treatment

Treatment is symptomatic and conservative. Immobilization of the knee in an extension or cylinder cast for 4 to 6 weeks may be helpful in particularly symptomatic cases. Symptoms will disappear after fusion of the tibial tubercle apophysis, although enlargement of the tubercle may be permanent. Occasionally a separate ossicle remains within the patellar tendon, and excision can be undertaken for persistent symptoms.

Sinding-Larsen-Johansson syndrome is a similar condition involving the inferior pole of the patella, causing sclerosis and fragmentation. This disorder too is caused by excessive tensile forces on the apophysis and is self-limited and treated symptomatically.

Other Osteochondroses

Köhler's disease of the tarsal navicular occurs in children, with swelling and pain in the area of the navicular. Radiographs show increased density in the navicular ossification center, and treatment is with casting and protected weight bearing until symptoms improve, followed by use of an orthotic arch support (Fig. 41-26). Sever's disease affects the posterior apophysis of the calcaneus in children and can be similar to Osgood-Schlatter disease after a chronic traction injury. Increased radiographic density of the apophysis and pain in the area occur, and treatment is with a heel lift and protected weight bearing when necessary. Freiberg's disease involves the epiphyses of the metatarsal heads, usually the second or third. The cause is unknown, but pathologically, like other osteochondroses, necrosis of bone with repair occurs, and the disease is self-limited and treated with a cast or orthosis with a metatarsal pad to protect the metatarsal head from weight bearing until healing occurs.

CONGENITAL DEFORMITIES

Congenital malformations can be caused by developmental aberrations in utero, environmental influences [radiation exposure, rubella, syphilis, human immunodeficiency virus (HIV), teratogenic drugs, etc.], or hereditary conditions. Overall incidence of congenital malformations diagnosed within 5 years of birth is 23 per 1,000 (Table 41-7). Some of the more common disorders are clubfoot (4.4 per 1,000), spina bifida (3 per 1,000), and hip dislocation (0.7 per 1,000). Metatarsus adductus, valgus hindfoot, unilateral externally rotated leg, internal tibial torsion, or an adducted thigh with external rotation of the leg are conditions thought to result from in utero position, which also may be a contributing factor in the pathogenesis of talipes equinovarus (clubfoot). All of these conditions except talipes equinovarus generally respond to passive stretching exercises, with corrective casts necessary only in refractory cases.

Hip Dislocation

Developmental dysplasia of the hip (DDH) consists of partial or complete displacement of the femoral head from the acetabulum. Epidemiologically, DDH is more common in certain ethnic groups (Navajo Indians, northern Italians, and Japanese), firstborn white females, and in association with other anomalies (torticollis, congenital knee hyperextension or dislocation, and foot deformities). Contributory causative factors include multifactorial inheritance, mechanical factors such as prenatal position, and hormonal influences such as maternal relaxins.

Pathology

In DDH the acetabulum is shallow with a more vertical orientation than normal. It is filled with remnants of ligamentum teres, fat, and fibrocartilage (pulvinar). Overgrowth of the lateral acetabular cartilage occurs in response to pressure of the femoral head (limbus), increasing the obstruction to reduction. A false acetabulum may be present as a depression on the ilium superior to the true acetabulum, lined with periosteum and a fold of the capsule. Ossification of the femoral head is delayed, and the capsule can develop an hourglass-shaped constriction crossed by the iliopsoas tendon. Within a few months of birth, tightness of the shortened adductor muscles develops, causing loss of hip abduction.

Clinical Manifestations

Examination of every newborn's hips in the nursery after birth is essential and effectively detects the majority of dislocated hips. Successful treatment hinges on early diagnosis and intervention. On examination, a hip click can be elicited as the femoral head reduces in the acetabulum with flexion and abduction of the thigh (Ortolani's sign). A similar click occurs with dislocation of the hip by adduction with gentle posterior pressure (Barlow's sign). The gluteal folds will be asymmetrical, and the affected thigh appears shortened when the knees and hips are flexed to 90 degrees (Galeazzi's sign). Limitation of abduction with the hip in flexion may be present at birth but becomes progressively apparent during the first 2 or 3 months. If the dislocation goes undetected the child will have a noticeable limp or waddle when walking begins. Periodic follow-up examinations of the hips in newborns during the first few months are important, and the hip click will disappear, with the limitation of abduction being the most obvious abnormality. Increased lumbar lordosis will be

present as the child begins to walk, especially in bilateral dislocations. Milder degrees of dysplasia, with subluxation or shallow acetabulum without frank dislocation, also can occur and are a cause of symptomatic degenerative hip arthritis in adolescents and young adults.

Radiologic Findings

The secondary center of ossification of the femoral head is absent in the newborn and usually appears by 4 to 6 months, although there is a fairly wide range (2 to 10 months). When present, the ossific nucleus should be within vertical lines from the acetabular margin (Perkins' lines) and below a horizontal line drawn through the triradiate growth plate cartilage of the acetabulum (Hilgenreiner's line, Fig. 41-27). The *acetabular index* is the angle of the acetabular roof with Hilgenreiner's line, and normally is 22 degrees. With DDH this angle is greater, and can be as high as 40 degrees. With the hip in abduction, the femoral neck fails to point toward the triradiate cartilage, and Shenton's line (a smooth arc from the lesser trochanter on the femur continuing along the neck and inferior cortex of the pubic bone) is discontinuous. If the ossific nucleus is not present, an arthrogram or MRI will readily demonstrate the dislocated position of the femoral head, although ultrasonography is also an accurate and a less invasive diagnostic modality.

Treatment

In the neonate with DDH, the hip usually can be reduced in flexion and abduction, and treatment is a splint or brace that can maintain this position. The Pavlik harness and Von Rosen splint are two of the more commonly used devices. Flexion of the hip should be 90 to 110 degrees, with slightly less than full abduction. Treatment is for 3 to 6 months, until the acetabular development is normal. In DDH diagnosed in the first few months of life, abduction bracing can also be successful, but radiographic evaluation is needed to be sure that the hip remains reduced. In children under 18 months of age, closed reduction can be attempted under anesthesia. The role of prereduction skin traction, which was advocated in the past, is currently controversial. If concentric reduction can be obtained as judged from radiographs or arthrogram, the patient is placed in a double hip spica cast with the hips flexed to approximately 100 degrees and slightly less than full abduction. Extremes of abduction can impair the blood supply to the femoral head and lead to avascular necrosis. Adductor tenotomy can facilitate the reduction. In children in whom closed reduction cannot be obtained, open reduction is performed using an anterolateral approach, or, in children less than 24 months of age, using a medial approach to the hip. The anterolateral approach in older children facilitates femoral shortening when necessary to achieve reduction. The child is placed in a bilateral hip spica cast with the hips in the position described above to maintain the reduction.

Persistent acetabular dysplasia in children 2 to 5 years old can be corrected by either pericapsular (Pemberton) or innominate (Salter) osteotomy. More extensive procedures such as the Chiari innominate osteotomy are occasionally necessary to obtain superior acetabular coverage of the femoral head in older children (Fig. 41-28).

Dislocation of the Knee

The three types of congenital dislocation of the knee are developmental (thought to be secondary to in utero position), defective embryogenesis (associated with other defects such as DDH), and quadriceps contracture caused by arthrogryposis. The knee is fixed in hyperextension with a variable degree of anterior subluxation or dislocation of the tibia on the distal femur. Treatment depends on the degree of severity and ranges from stretching exercises to surgical quadriceps lengthening followed by cast immobilization.

Pseudarthrosis of the Tibia

Children with this disorder may have a frank pseudarthrosis at birth (usually at the level of the distal and middle thirds of the tibia) or, in milder forms, anterolateral bowing that progresses to pseudarthrosis later. This condition must be differentiated from congenital posteromedial bowing of the tibia, which usually is self-limited. Congenital pseudarthrosis of the tibia is frequently associated with neurofibromatosis. With bowing alone, a total- contact orthosis is used to prevent progression, but once pseudarthrosis develops it can be very difficult to treat. Treatment methods for established pseudarthrosis include intramedullary rod fixation with bone grafting; more recently, free vascularized fibular grafts have been used successfully. Amputation occasionally is necessary in refractory cases.

Metatarsus Adductus

Metatarsus adductus is a common congenital foot deformity also known as metatarsus varus, the cause of which is thought to be related to intrauterine positioning. It occurs in 1 per 1,000 live births, and half of the cases are bilateral. The forefoot is angulated medially, but the deformity is not associated with equinus of the heel. When passively correctable, metatarsus adductus resolves spontaneously. In incompletely correctable deformities, serial manipulation and casting is the accepted treatment. Surgical treatment rarely is necessary and is indicated only in children over the age of 5 years with severe symptomatic residual adductus deformity. Osteotomies of the lateral column (cuboid), medial cuneiform, or metatarsals can accomplish correction.

Talipes Equinovarus (Clubfoot)

Talipes equinovarus is a complex deformity consisting of plantar flexion of the ankle, inversion of the foot, adduction of the forefoot, and internal rotation of the tibia (Fig. 41-29). The cause is unknown, and theories include neurologic and vascular defects. Inheritance is variable in pattern, with overall incidence of approximately 1 per 1,000 births, and a 2:1 male preponderance. Siblings have a 1 in 35 probability of the disorder, and identical twins a 1 in 3 chance. The prevailing theory of cause is a single gene predisposing mutation in conjunction with one or more unidentified environmental factors. Talipes equinovarus is associated with other disorders, including myelomeningocele, arthrogryposis, and diastrophic dwarfism.

Pathologic Anatomy

Three-dimensional computer modeling has shown the talus to be externally rotated in the ankle mortise, with 45 degrees of internal rotation of the talar neck (compared to slight internal rotation in the mortise and a 25-degree internal rotation of the neck in normals). The talus is also smaller than normal, and the calcaneus is internally rotated and in varus compared to slight external rotation in normal feet. The tibial tendon sheaths are fibrotic and thickened, and the navicular is subluxated medially on the head of the talus. Radiographically, the calcaneus and talus are more parallel than divergent on a lateral view (Fig. 41-30).

Treatment

Without treatment the condition is permanent and ambulation difficult. Treatment is started immediately in the neonatal period, with stretching exercises and serial plaster casts. Adduction of the forefoot is corrected first, then hindfoot varus, and correction of equinus is not undertaken until these deformities are completely corrected. Conservative treatment is successful in about 50 percent of patients. Surgical correction consists of posteromedial and subtalar release and usually is delayed until the patient reaches 6 to 8 months, but it should not be delayed beyond the age of 1 to 2 years. Complete release of the subtalar joint is best accomplished by a circumferential transverse incision extending from the lateral side of the hindfoot posteriorly to the medial side, known as the Cincinnati approach.

In recurrent cases, manipulation will not suffice for correction, and posteromedial release is necessary. Lateral transfer of the tibialis anterior tendon along with Achilles tendon lengthening may also be helpful in patients with mild degrees of recurrence. Incomplete release of the calcaneocuboid joint is believed to be the most common cause of residual adduction or cavovarus deformities. In severe refractory deformities such as those seen with arthrogryposis, talectomy can be necessary to achieve a plantigrade foot.

Convex Pes Valgus (Vertical Talus)

Congenital vertical talus (also known as rocker-bottom flatfoot) consists of a dorsal dislocation of the navicular on the head of the talus, with the talus in a more vertical position than normal (Fig. 41-31). The sole of the foot has a rocker-bottom deformity and is not passively correctable. Vertical talus may be an isolated anomaly, or it may be found in association with meningomyelocele, arthrogryposis, or chromosomal aberrations. Early manipulation and casting can be successful, particularly in milder deformities (referred to as *oblique talus*, in which the first metatarsal aligns with the talus in forced plantar flexion). However, the majority of cases of true vertical talus require surgical correction. Diagnosis is made from the lateral forced-plantar-flexion radiograph, in which the talus and the first metatarsal are not in alignment. Operation is delayed until the patient is 4 to 8 months old and consists of open reduction and pinning of the talonavicular joint, along with lengthening of the Achilles tendon, posterior tibialis, and, when necessary, the toe extensors, peronei, and anterior tibialis tendons. In older children, a Grice subtalar arthrodesis may be necessary. Undiagnosed cases in older children with adaptive bony changes are treated with triple arthrodesis (see Fig. 41-14).

Arthrogryposis Multiplex Congenita (Amyoplasia)

Arthrogryposis involves failure of development of muscles and resulting contractures of the extremities. It is not thought to have a genetic basis and is characterized by severe, symmetrical contractures with deformities, including scoliosis, talipes equinovarus, DDH, knee dislocation, internally rotated shoulders, and extended elbows. The individual deformities are treated as described above, but contractures tend to be quite severe, and occasionally osteotomies are needed to improve function and decrease deformity. A genetic variant called distal arthrogryposis has been identified, which affects the digits only.

High Scapula (Sprengel's Deformity)

High scapula is caused by embryonic failure of the scapula to migrate to its normal position. The scapula may be attached to the vertebrae by a pathologic band of fibrous tissue and cartilage (omovertebral mass). Other deformities of the spine and ribs are often associated. The scapula is 1 to 4 inches higher than usual, tilted anteriorly, and does not move normally with shoulder motion (Fig. 41-32). Usually it is the cosmetic appearance rather than any functional deficit that brings attention to the disorder. Mild cases need no treatment, but in more severe cases surgical correction can be undertaken, although it is generally delayed until the patient is 3 to 6 years old.

Klippel-Feil Syndrome

Also known as brevicollis and congenital short neck, this syndrome consists of multiple fusions of cervical vertebrae secondary to defective resegmentation of the cervical spine during embryogenesis and is usually not treatable.

Torticollis

Torticollis, also known as congenital wryneck, is caused by unilateral contracture of the sternocleidomastoid muscle, with tilt of the head toward the affected side, rotation to the opposite side, and varying degrees of facial dysplasia. The disorder is thought to be posttraumatic, with a tender swelling in the muscle preceding the deformity. Treatment consists of stretching exercises and, in refractory cases, surgical release of the muscle with cast immobilization or vigorous passive exercises postoperatively.

Cleidocranial Dysostosis

Cleidocranial dysostosis is an autosomal dominant hereditary syndrome exhibiting clavicular aplasia, widened diameter of the cranium, and delayed fontanelle closure. Disability is minimal and treatment usually not needed.

Radioulnar Synostosis

This disorder results from defective embryogenesis, with a fusion between the proximal radius and ulna and absence of pronation and supination. Most cases are bilateral, and the degree of deformity or hypoplasia of the proximal radius is variable. Nonfunctional extremes of pronation or supination can be corrected by osteotomy, but resection of the fusion

bar has given disappointing results. Resection followed by low-dose irradiation (8 to 10 Gy), a treatment that prevents heterotopic ossification after hip surgery, has been used with some success in prevention of bar reformation.

Madelung's Deformity

This disorder involves progressive volar and ulnar angulation of the distal radius, causing prominence of the distal ulna and abnormality of the shape of the distal radial articular surface. Although wrist pain can occur, most patients are treated conservatively with splinting. Rarely, osteotomy of the distal radius and shortening of the distal ulna is indicated.

Aplasia or Dysplasia of Long Bones

These rare deformities, which do not appear to be hereditary, consist of absence of a bone or part of a bone. The term *congenital skeletal limb deficiencies* is used to describe these disorders collectively. Skeletal deficiencies can be intercalary (an absent segment with bones distal to it intact) or terminal (all bones distal to the lesion absent), and transverse or longitudinal. As an example, absence of the fibula alone would be classified as an intercalary, longitudinal fibular hemimelia. Absence of a hand or foot would be described as a terminal transverse defect (congenital amputation).

Absence of the Radius

The deformity known as radial clubhand is caused by the absence of the radius, which is sometimes bilateral and can be associated with craniofacial anomalies, thrombocytopenia, or Fanconi's anemia (pancytopenia). If the defect is terminal rather than intercalary, the thumb may be absent. Early treatment involves manipulation and splinting and, by the age of 12 months, surgical centralization of the carpus on the ulna. Absence of the ulna is much less common and is associated with digital abnormalities, such as syndactyly, in 90 percent of cases. Treatment is generally nonoperative for the wrist and elbow and focuses on correction of digital abnormalities that impair function.

Fibular Hemimelia

In this disorder, the limb is shortened and the hindfoot usually in equinovalgus position. Involvement is bilateral in 30 percent of cases. For milder cases, limb length equalizations may be possible using the Ilizarov technique; more severe deficiencies may require amputation.

Femoral Hypoplasia

These deformities can range from mild hypoplasia to complete absence of the upper femur. Mild cases are treated by limb length equalization procedures. In more severe cases fusion of the femoral remnant to the ilium or amputation with knee fusion and prosthetic limb fitting has been performed.

GENERALIZED BONE DISORDERS

Composition

Bone is made up of organic and inorganic materials. The organic components include type I collagen (90 percent of the organic matrix), phosphoproteins, bone-specific proteoglycan, sialoprotein, osteonectin, osteocalcin, and growth factors such as *transforming growth factor-beta (TGF- b)* , fibroblast growth factors, insulinlike growth factors, and bone morphogenetic proteins. These factors and proteins incorporated into the matrix of bone are responsible for the maintenance of normal bone remodeling processes, and they give bone matrix its osteogenic properties, which make bone grafting possible. Before it is mineralized, the bone matrix secreted by osteoblasts is called *osteoid*. The inorganic phase of bone consists of hydroxyapatite, a crystalline form of calcium and phosphate, and water, which makes up 8 to 9 percent of bone. Citrate, carbonate, and magnesium also are present. The mineral phase of bone is responsible for its compressive strength, while collagen gives bone its tensile strength. Hydroxyapatite crystals are spontaneously deposited in the "hole zones" formed within the quarter-staggered overlapping array of the type I collagen molecules and later are spread along the fibers by secondary nucleation. Osteonectin, a glycoprotein in bone matrix, may have a role in binding the hydroxyapatite crystals to the collagen. Osteopontin, a phosphoprotein in bone matrix, may be involved in cell attachment to bone, and osteocalcin, a carboxyglutamic acid containing glycoprotein, is thought to be involved in osteoclastic recruitment for the resorption of bone matrix.

Cell Biology

Bone matrix is normally secreted by osteoblasts in discrete layers (lamellae), with the collagen in each successive layer oriented 90 degrees to the adjacent layer. This plywoodlike structure maximizes the mechanical strength of the material. Osteoblasts become entrapped in the secreted matrix and become more quiescent cells known as *osteocytes*. Osteocytes have numerous cell processes called *canaliculi*, which communicate with each other through the bone matrix. In injury and repair processes, bone- forming tumors, and diseases such as Paget's disease, osteoblasts initially secrete woven bone, in which the collagen fibrils have a random orientation. The three-dimensional structure of bone is similar to that of a sponge, with all the bony trabeculae interconnected and the pores containing the bone marrow.

Osteoclasts are multinucleated bone-resorbing cells (also called giant cells) of monocytic lineage that attach to bone surfaces and secrete enzymes that break down organic components of bone matrix. In addition, a proton pump in the plasma membrane of the osteoclast acidifies the region between the cell and the bone surface, dissolving the mineral phase. The cavity excavated by the osteoclast is referred to as a *Howship's lacuna*. Bone is constantly formed and resorbed in the skeleton throughout life. Bone formation and resorption are locally coupled, with osteoblasts always following in the wake of an osteoclast and replacing the resorbed bone with new lamellar bone. Coupling factors such as bone morphogenetic proteins and other members of the TGF-b family are thought to be activated by the acidic environment of the osteoclast and may serve to attract and stimulate osteoblast progenitor cells, accounting for the local coupling phenomenon (Fig. [41-33](#)). When cortical bone undergoes remodeling, osteoclasts tunnel longitudinally within the cortex, forming a cutting cone. Concentric layers of lamellar bone are subsequently deposited, leaving only a narrow central canal that contains vascular channels supplying blood to the bone. This structure is referred to as a *haversian system* or *osteon* (Fig. [41-34](#)).

Enzymes

Osteoblasts contain high levels of alkaline phosphatase, while osteoclasts contain glycolytic enzymes, collagenases, acid hydrolases, and acid phosphatase. Evidence suggests that some of these enzymes, which are glycosylated with mannose-6-phosphate, remain on the resorption surface and can site-direct bone formation through interaction of mannose-6- phosphate moieties with the insulinlike growth factor II receptor of osteoblasts or osteoblast precursor cells. Alkaline phosphatase is thought to be involved in the mineralization process, although its role is not yet understood. Serum levels of alkaline phosphatase derived from bone can be elevated in rickets, osteomalacia, Paget's disease, osteosarcoma, and metastatic carcinoma. Decreased levels can be found in achondroplasia, scurvy, hypothyroidism, and a hereditary deficiency of alkaline phosphatase called *hypophosphatasia*.

Ossification

Endochondral Ossification

The long bones of the skeleton, with the exception of the clavicle, are formed embryonically from a hyaline cartilage precursor. In the center of the precursor, or anlage, the chondrocytes undergo hypertrophy and begin to calcify the surrounding cartilaginous matrix; this constitutes the primary ossification center. The process extends toward the ends of the long bone, becoming an orderly front of proliferating, hypertrophying, and calcifying cells known as the *growth plate* or *physis*. With ingrowth of the nutrient vessel into the central portion of the forming bone, the calcified cartilage is converted into trabecular bone. Concomitantly, formation of the cortical portion of the bone begins with osteoblastic bone synthesis from the periosteum directly (membranous bone formation), giving rise to the primary bone collar, which also extends from the center toward the ends of the long bone (Fig. [41-35](#)).

During fetal and childhood osteogenesis, endochondral ossification continues in the growth plate, accounting for growth in length of the bone; ossification also occurs within the epiphysis when the secondary centers of ossification appear. Latitudinal growth also occurs at the level of the growth plate by recruitment of cells from the perichondral ring of Lacroix. The diameter of the shaft of the long bones increases through periosteal accretion of new cortical bone with endosteal resorption. The process of endochondral ossification is an integral part of fracture healing by callus formation. Endochondral ossification as it occurs in the growth plate is depicted in Fig. [41-36](#). The growth plate exhibits a gradient of oxygen and nutrients from the upper proliferating zone to the zone of provisional calcification, and the chondrocytes gradually shift from an aerobic metabolism in the proliferating zone to an anaerobic metabolism in the hypertrophic zone. The hypertrophic chondrocytes acquire high levels of alkaline phosphatase, accumulate calcium, and secrete calcium-containing matrix vesicles from the plasma membrane that are thought to initially nucleate hydroxyapatite crystals in the matrix.

The proliferation and maturation of the chondrocytes in endochondral calcification are controlled systemically by growth hormone through insulinlike growth factor I, and locally by autocrine growth factors. The most important growth factors known to control chondrocyte proliferation and hypertrophy are TGF-b, basic fibroblast growth factor (bFGF), and parathyroid hormone-related protein (PTHrP). bFGF is produced in the proliferating and resting zones, and defects in FGF receptors are responsible for disorders such as achondroplasia and thanatophoric dysplasia (FGFR3 mutations). PTHrP is produced in the lower proliferating and upper hypertrophic zone and stimulates proliferation while suppressing hypertrophy. PTHrP or PTHrP receptor deletion causes severe lethal disruptions of the growth plate, with loss of the proliferating zone, premature hypertrophy, and defective transformation of calcified cartilage into bone.

The zone of provisional calcification is mechanically the weakest part of the growth plate, accounting for the propensity of children's fractures near the ends of long bones to involve this region. In addition, the hypertrophic chondrocytes produce a unique collagen (type X collagen) that facilitates conversion of the calcified cartilage into bone in the metaphysis. Defects in endochondral ossification may be inherited (achondroplasia) or acquired (vitamin D deficiency rickets).

Intramembranous Ossification

This form of ossification does not involve a cartilage precursor tissue; rather, osteoblasts secrete osteoid and mineralize it directly. Intramembranous or osteoblastic ossification occurs in the calvaria and, as described above, in the periosteum of the long bones. This process is a part of the later phases of fracture healing, in which it gradually supplants endochondral ossification.

Remodeling

As described earlier, bone is continuously forming and resorbing in a locally coupled manner. During growth, remodeling also changes the shape of the long bones, with activity of osteoclasts and osteoblasts particularly in the cone-shaped metaphyseal area resulting in a narrowing of the diameter of the bone toward the diaphysis. In addition, bone formation and resorption are strongly influenced by mechanical stresses in bone, with net resorption in unloaded bone and increased accretion in loaded bone. Bone is an anisotropic material and develops piezoelectric charges on its surfaces in response to mechanical stresses. Bone cells are thought to respond to these electrical signals, modulating formative and resorptive activities.

Bone Grafting

Mechanisms

Bone grafting, the transplantation of bone from one site to another (autologous graft) or from one individual to another (allograft), is an essential part of many orthopaedic surgical procedures. Bone grafts work through one or both of two mechanisms—osteoconduction or osteoinduction. With osteoconduction, the graft may serve to provide a mechanical support and serves as a scaffolding on which host osteoblasts can form new bone and through remodeling eventually replace the graft. This mechanism predominates in cortical bone and large segmental allografts. Osteoinduction is the property of release and activation of bone- inducing matrix proteins from the graft during osteoclastic resorption. Bone morphogenetic protein and other members of the TGF-βfamily are among the most important factors in this process. When demineralized bone matrix devoid of bone cells is implanted into muscle or subcutaneous tissue, endochondral-sequence bone formation occurs through this osteoinduction phenomenon.

Sites

The iliac crest is generally used as a donor site for cancellous grafting, and the tibia or fibula for cortical bone grafts. A recent development is the use of vascularized bone grafts, wherein an expendable bone such as the midportion of the fibula is removed along with its nutrient vessels. The vessels are then anastomosed to blood vessels in the area of the recipient site using microsurgical technique. This allows bridging of significant bony defects with living bone, which rapidly incorporates. This technique is being used to speed healing of avascular necrosis of the femoral head by inserting a vascularized graft in the femoral neck and head, and for repair of bony defects caused by trauma or tumor resections.

Allografts

Bone allografts are well tolerated immunologically because antigenic cellular proteins make up only a minute fraction of bone. Freezing or freeze-drying further diminishes the antigenicity of bone without impairing its osteoconductive and osteoinductive properties. As a result bone allografts rarely incite rejection reactions and immunosuppression is unnecessary. Osteochondral allografts are used to reconstruct joint surfaces and contain both bone and articular cartilage. They generally are frozen in the presence of a cryoprotectant such as dimethylsulfoxide, which allows preservation of a proportion of the chondrocytes. The collagen and proteoglycan matrix of articular cartilage does not allow penetration of lymphocytes or antibodies; thus cartilage is also an immunologically privileged tissue.

Synthetic Graft Substitutes

Synthetic materials are being used experimentally as bone graft substitutes and show significant promise. Tricalcium phosphate and hydroxyapatite ceramics and corals have been used successfully to fill bony defects and function as osteoconductive substrates. Resorbable synthetic materials such as polyglycolic and polylactic acid are under investigation as vehicles for growth factors. Purified and recombinant bone morphogenetic proteins and demineralized bone matrix have both been used successfully as bone graft substitutes. Table 41-8 lists bone grafting materials with summaries of their properties.

Classifications

Developmental disordersAchondroplasia Enchondromatosis Multiple exostoses Polyostotic fibrous dysplasia Osteogenesis imperfecta Osteopetrosis Osteochondrodystrophies
Metabolic disordersMineralization defects

1 Vitamin D deficiency (rickets or osteomalacia)

1a Nutritional

1a Malabsorption syndromes

1a Renal disease

1a Hepatobiliary disease

1 Vitamin D-resistant hypophosphatemic rickets

1 Aluminum-induced osteomalacia

Hormonal disorders

1 Hyperthyroidism

1 Hyperparathyroidism

1 Hypercortisolism (Cushing's disease)

1 Postmenopausal osteoporosis

1 Hypogonadism

1 Pituitary disorders

Bone marrow diseasesReticuloendothelial system

1 Histiocytosis X

1a Letterer-Siwe disease

1a Eosinophilic granuloma

1a Hand-Schuller-Christian disease

1 Lipoid granulomatosis

1a Gaucher's disease

1a Niemann-Pick disease

1a Xanthomatosis

Lymphomas of bone

1 Hodgkin's disease

1 Histiocytic lymphoma

Hematopoietic system

1 Leukemia

1 Multiple myeloma

1 Hemolytic anemias

1a Thalassemia (Cooley's anemia)

1a Sickle cell anemia

1a Erythroblastosis fetalis

1 Myelofibrosis

1 Mastocytosis

Vascular disordersPaget's disease Sudeck's atrophy Massive osteolysis (Gorham's disease, or disappearing bone disease)

Developmental Disorders

Defective cartilage proliferation, calcification, collagen synthesis, and bone remodeling are the most common causes of developmental disorders of bone. Some of these disorders are inherited conditions, and recent advances in molecular biology have allowed identification of specific defects in an increasing number of diseases. Table 41-9 lists musculoskeletal diseases for which the specific genetic defect has been identified.

Achondroplasia

The basic defect in this inherited autosomal dominant disorder is in the proliferation of cartilage due to FGFR3 mutation. Membranous bone development (calvaria, ribs, sternum) is normal, but endochondral ossification is defective. The proliferating zone of the growth plates is shortened and disordered. At birth, the infant has a normal-sized body, large head with a depression of the base of the nose, and short limbs. Shortness of stature because of deficient long-bone growth becomes progressively evident during childhood (Fig. 41-37). Trunk growth remains relatively normal, while the limbs are short, particularly proximally. The hands are short and broad, with digits of equal length. Increased lumbar lordosis may be present, as well as thoracolumbar kyphosis. Intelligence is normal, although life expectancy may be diminished. Genu varum is a common finding, and later in life patients may develop spinal stenosis as a result of decreased spinal canal diameter and interpedicular distance.

Enchondromatosis (Ollier's disease or Dyschondroplasia)

This condition, first described by Ollier in 1899, consists of abnormal foci of cartilage within the metaphyses of the long bones. These cartilage rests are referred to as *enchondromas*. The cause is unknown, and the disease is not hereditary. Disturbance of the growth near the ends of the involved bones results in foreshortening, with broadening and cystic changes in the metaphyses. The degree of deformity varies considerably, with some mild cases of enchondromatosis exhibiting no significant bony distortions. Involvement of the bones of the hand is common. When associated with hemangiomas, the disease is called Maffucci syndrome. The incidence of malignant degeneration of individual lesions into chondrosarcoma occurs in about 25 percent of individuals by age forty and is even more common in Maffucci syndrome. Pathologic fractures caused by weakening of bone by an enchondroma can heal, but the enchondroma will persist, and so curettage and bone grafting is recommended. Malignant degeneration of lesions is managed by local resection and reconstruction when necessary.

Multiple Exostoses (Metaphyseal Aclasis, or Osteochondromatosis)

This autosomal dominant hereditary disorder involves outgrowths of cartilaginous lesions from the metaphyses of the long bones as well as occasionally from the pelvis, ribs, and spine (Fig. 41-38). Extent of the disease is variable, and it is one of the more common skeletal dysplasias. Hereditary multiple exostoses have been mapped to three different chromosomal locations, but the specific genes involved have not been identified. Cartilage-capped exostoses, or osteochondromas, arise from bony surfaces and increase in size during childhood growth. The lesions grow by a process of relatively normal endochondral ossification, forming trabecular bone in the base. They can cause mechanical symptoms as a result of impingement on nerves, joints, tendons, or muscles and, when symptomatic, can be removed surgically. Care must be taken to remove the cartilage cap in its entirety, particularly in children, or local recurrence can result. Genu valgum and progressive ulnar deviation of the hand as well as limb length discrepancies because of asymmetrical limb involvement are other problems that may arise. The most serious complication is malignant degeneration into chondrosarcoma, which has been reported in 1 to 15 percent of affected individuals. This is usually heralded by enlargement of the lesion in adulthood, sometimes in association with pain. Thickness of the cartilaginous cap as assessed by CT or MRI is a good index of malignant transformation, with a thickness of 1 cm suspicious and over 2 cm pathognomonic. Treatment is by wide surgical excision, and metastasis is uncommon.

Polyostotic Fibrous Dysplasia

This disease usually appears in childhood, with development of expansile lytic lesions in the metaphyses and diaphyses of long bones. The pelvis, skull, and bones of the hand also may be involved. Bending deformities from recurrent fractures of the long bones can occur, most commonly in the proximal femur, causing a “shepherd's crook” deformity (varus of the upper femur). Radiographically, the dysplastic bone has a homogeneous “ground glass” appearance as a result of the lack of formed trabeculae (Fig. 41-39). Lesions that are symptomatic or threatening pathologic fracture require curettage and bone grafting as well as internal fixation for stabilization. Cortical bone graft is superior to cancellous graft, which tends to be resorbed by the dysplastic bone. Association with café-au-lait pigmented skin lesions and precocious puberty is termed Albright syndrome. Malignant degeneration has been reported in these lesions but is extremely rare. Fibrous dysplasia occurs also in a monostotic form.

Osteogenesis Imperfecta

Osteogenesis imperfecta is a genetically determined disorder in the structure or processing of type I collagen, with a spectrum of expression. Bone fragility is the unifying clinical manifestation, and four clinical groups have been categorized by Sillence (Table 41-10). While the clinical classification is useful for general treatment and prognosis, recent identification of the specific mutations in the type I collagen gene indicates a wide variety of mutations with differing clinical manifestations. Patients may have blue sclerae, abnormal dentition, deafness, and growth failure. The fetal form (type II) is severe and lethal, the infantile form (type III) less severe but associated with severe deformities, and the adolescent form (tarda, type I) is the most common and least severe form.

Radiographically, the bones may appear normal in type I but tend to be more gracile with decreased trabecular pattern. With more severe involvement, fractures may be evident, and enlargement of the metaphysis and epiphysis may be present, with disorganization of the growth plate into multiple islands of cartilage.

Treatment

Osteogenesis imperfecta is treated with orthoses to prevent fractures and correction of deformities by multiple osteotomies with intramedullary stabilization using telescoping rods (Fig. 41-40). Fracture tendency usually decreases at puberty. Correction of scoliosis with posterior instrumentation and fusion is helpful in preventing respiratory compromise but can be difficult. A wide variety of systemic therapies, including calcium, phosphorus, vitamin C, vitamin D, fluoride, and calcitonin, have been used without demonstrable beneficial effect.

Osteopetrosis (Albers-Schönberg Disease)

This rare skeletal disease is associated with increased density of the bones (Fig. 41-41). It has two forms, an infantile severe form, which is inherited as an autosomal recessive, and an adult (or tarda) milder form, which is an autosomal dominant trait. The central defect is in osteoclastic function, with failure of functional bone resorption and therefore of remodeling. Multiple genetic defects in osteoclast function have been shown to cause osteopetrosis (see Table 41-9). Consequently the medullary cavities are narrowed, with resulting deficiencies of bone marrow activity causing anemia and thrombocytopenia. The bones show persistence of calcified cartilage and woven bone, and absence of osteonal cortical bone. Patients have increased fracture tendency, susceptibility to osteomyelitis, and cranial nerve deficits from skull involvement.

Several patients with infantile osteopetrosis have been cured by bone marrow transplantation, with resumption of bone resorption and remodeling attributable to donor osteoclasts. Bone density in these patients and hematopoietic function returned to normal.

Melorheostosis

This condition, which may involve one or more bones in an extremity, is characterized by endosteal cortical hyperostosis, often described as a “candle dripping” appearance radiographically. The disease is associated with pain and contracture of adjacent joints. The cause is unknown, and there is no specific therapy beyond symptomatic treatment.

Metabolic Diseases

Mineralization

Mineralization requires secretion by chondrocytes or osteoblasts of a suitable organic matrix (osteoid or chondroid), within which hydroxyapatite crystals can form. Mineralization defects can result from inadequate availability of calcium and phosphate, interference with crystal formation by drugs or heavy metals, disruption of vitamin D metabolism, or aberrations in the organic matrix.

Scurvy

Scurvy is the clinical condition resulting from a nutritional deficiency of vitamin C, or ascorbic acid. Ascorbate is crucial to the cross-linking process in fibrillar collagen formation, which determines mechanical properties of collagen fibers. Capillary walls are most obviously affected, with microscopic hemorrhage as the hallmark of scurvy. Subperiosteal hemorrhages occur adjacent to the growth plates in the metaphyses of the long bones (Fig. 41-42) and in extreme cases can cause epiphyseal separation. Mineralization in the growth plate is defective, with an irregular and dense mineralization front where transition of the calcified cartilage to bone is impaired. Hemorrhage from the gastrointestinal tract or mucous membranes and subperiosteal hemorrhages present with pain and swelling that can mimic osteomyelitis. Treatment with vitamin C rapidly cures the disease, and within 24 h pain subsides. Protection from weight bearing until reossification occurs is recommended.

Vitamin D Metabolism

Vitamin D (cholecalciferol) is essential for normal bone metabolism and mineralization. Vitamin D is formed from dietary precursor forms, or from ultraviolet irradiation of 7-dehydrocholesterol in the skin. The active form of vitamin D is 1,25-dihydroxyvitamin D₃. 25-Hydroxylation occurs in the liver, and subsequent 1- α -hydroxylation occurs in the kidney. The major function of active vitamin D₃ is to enable absorption of calcium in the gut. Additional effects include a stimulation of renal tubular phosphate reabsorption and bone resorption. The physiologic significance of the bone resorbing effects is unknown, but the mechanism is through stimulation of differentiation of osteoclasts from precursor cells. The 1- α -hydroxylation of vitamin D is up-regulated in the kidney by *parathyroid hormone (PTH)*. Additional effects of PTH include stimulation of bone resorption and decreased renal tubular phosphate reabsorption. The stimulation of bone resorption involves osteoblasts as the target cell, with secondary secretion of an unknown factor that then induces osteoclasts to resorb bone. Calcitonin, on the other hand, activates osteoclast receptors directly to inhibit resorption. While this hormone has therapeutic uses in Paget's disease, hypercalcemia, and some forms of osteoporosis, it is not thought to play a significant role in normal human bone metabolism. Figure 41-43 summarizes the roles of vitamin D and PTH in the regulation of calcium and phosphate metabolism.

Rickets

Rickets is a relative deficiency of the active metabolite of vitamin D (1,25- dihydroxyvitamin D₃) in children, which leads to the inability of chondrocytes to actively mineralize matrix in the zone of provisional calcification of the growth plates. The defect in mineralization has been presumed to be secondary to inadequate calcium absorption in the gut, with inadequate local levels of calcium and phosphate at the mineralization front. This can be aggravated by decreased reabsorption of phosphate in the renal tubules, another effect of vitamin D. Some evidence suggests that chondrocytes in the growth plate have vitamin D receptors and can respond to this hormone in a direct fashion as well. Nutritional vitamin D deficiency in children is now rare in the United States, but disruptions of vitamin D metabolism (renal disease, anticonvulsant therapy, malabsorption syndromes) and inherited diseases (vitamin D-resistant hypophosphatemic rickets) can cause rickets.

Pathology

The normal growth plate is a well-defined plate of cartilage approximately 2 mm in thickness that separates the metaphysis from the epiphysis. In rickets, the growth plates are widened and irregular, with patchy decreased mineralization. The proliferative and hypertrophic zones are increased in height and somewhat disordered. The bony trabeculae in the metaphyses are thin, with diminished bone formation. In addition there can be widening and cupping of the metaphyses (Fig. 41-44) and deformities such as coxa vara, tibial bowing, and genu varum or valgus. Other clinical manifestations include gastrointestinal symptoms, irritability, open fontanelles, narrow chest, and prominence of the costochondral junctions (referred to as the “rachitic rosary”).

Treatment

Treatment of nutritional rickets is by supplementation with high doses of vitamin D (3,000 U daily; normal daily requirement 400 U). Calcium supplementation enhances the rate of healing. For persistent deformities, correction is delayed until the underlying metabolic defect has resolved, at which time osteotomies can be performed.

Enteropathic Rickets

The most common cause of malabsorptive rickets is celiac, or gluten- sensitive, enteropathy. Chronic diarrhea results in inadequate vitamin D absorption in the gut. The effects of the vitamin D deficiency are exacerbated by periods of rapid growth. The effects on bone and growth plates are similar to those described above, and treatment hinges on the elimination of gluten from the diet and vitamin D supplementation.

Renal Osteodystrophy

Renal disease impairs the 1-a-hydroxylation of vitamin D, causing a functional vitamin D deficiency. Inadequate calcium absorption in the gut causes slight hypocalcemia, which then leads to a secondary hyperparathyroidism. Severe osteopenia results, and in children the clinical picture resembles rickets. In patients on renal dialysis, aluminum contained in phosphate-binding agents (used chronically to decrease phosphate levels) binds to mineralization surfaces and impairs further mineral deposition, a form of osteomalacia. Both glomerular and tubular defects can cause renal osteodystrophy, including Fanconi syndrome (glycosuria, aminoaciduria), Lignac-Fanconi syndrome (impaired tubular phosphate reabsorption), and renal tubular acidosis.

Vitamin D-Resistant Hypophosphatemic Rickets

Vitamin D-resistant hypophosphatemic rickets is a familial disease associated with hypophosphatemia. The syndrome resembles nutritional vitamin D deficiency, except that very high doses of vitamin D in conjunction with phosphate supplementation are required to treat it. The patients are short in stature with disproportionately short limbs, and the disease is inherited as an X-linked dominant trait. The genetic cause of the disease has been identified as a neutral endopeptidase known as PEX, although its substrate is unknown. Oncogenic hypophosphatemia is an unrelated disorder associated with benign soft-tissue tumors that disappears with tumor excision.

Hypophosphatasia

Hypophosphatasia is a rare autosomal recessive hereditary disease characterized by a deficiency of alkaline phosphatase and urinary excretion of phosphoethanolamine. The disease, which is variable in severity, may present in childhood or adulthood, featuring stunting of growth and excessive bone fragility. Hyperphosphatasia is an extremely rare congenital disorder with markedly elevated levels of serum alkaline phosphatase. It is also known as juvenile Paget's disease. The skull and diaphyses of the long bones are thickened, and deformity and fractures can result. There is no effective systemic treatment for either of these disorders.

Osteomalacia

Osteomalacia is defined as a defect in mineralization of adult bone, generally resulting from abnormalities in vitamin D metabolism. As in rickets, causes of osteomalacia include nutritional deficiency, malabsorption, anticonvulsant therapy, and hepatic and renal diseases. In addition, aluminum can induce a mineralization defect, as can chronic use of diphosphonates (a treatment sometimes used for Paget's disease). Anticonvulsants can cause osteomalacia by interfering with vitamin D metabolism in the liver. Histologically, excessive unmineralized surfaces and thickness of the osteoid are observed. The bones become osteopenic, and fractures can result. Nutritional osteomalacia is more common in the elderly than previously recognized, aggravating involutional osteoporosis and the tendency for fractures. A characteristic radiographic finding is Looser's zones, which are stress fractures extending partly through the bone, usually on the concave side, secondary to mechanical failure in compression. Treatment depends on the underlying cause, but generally involves vitamin D supplementation.

Hyperparathyroid Bone Disease (Osteitis Fibrosa Cystica)

Excessive secretion of PTH can cause bone disease as a result of the resorptive effects of PTH on bone. Patients may present with bone pain and tenderness, hypercalcemia, and hypophosphatemia. Common causes are an underlying parathyroid adenoma or adenocarcinoma. Muscle weakness and calcium phosphate renal stones also may be present. Radiographically osteopenia is observed, with subperiosteal resorption in some sites, best seen along the radial borders of the digits and along the distal clavicle. Hemorrhagic cystic lesions (brown tumors) can occur in the long bones but usually resolve with treatment of the underlying hyperparathyroidism. Parathyroidectomy, the treatment of choice, results in spontaneous resolution of the bone disease.

Osteoporosis

In osteoporosis, a common disorder in the elderly, total bone mass and trabecular volume are decreased, but mineralization is normal. Skeletal bone mass reaches its peak at about the age of thirty. Both men and women subsequently lose bone throughout life, though women achieve a lower starting peak mass and have a hormonal acceleration of bone loss due to estrogen loss at menopause. Symptomatic osteoporosis with fractures is therefore much more common in women. Osteoporosis can be primary or secondary to some other process, usually a hormonal abnormality. Table [41-11](#) lists some of the types of osteoporosis. Type I and type II involutional osteoporosis are the most common. Osteoblasts have been shown to have estrogen receptors and respond to estrogen anabolically, which may explain the accelerated postmenopausal bone loss. Type I osteoporosis generally occurs in women 50 to 75 years of age and is characterized by loss primarily of trabecular bone. PTH levels are decreased in this group. Type II osteoporosis has a 2:1 female preponderance and is found in patients over 70 years of age, with proportionate loss of both trabecular and cortical bone. It is thought to result from depressed renal 1-a-hydroxylase activity associated with aging, leading to inadequate active vitamin D levels and decreased intestinal calcium absorption. Consequently these patients tend to have a mild secondary hyperparathyroidism contributing to their chronic bone loss. Type III patients are similar to type I except that their serum PTH levels are elevated; this group comprises about 10 percent of the postmenopausal form.

Secondary osteoporosis can result from a number of causes, as listed in Table [41-11](#), which must be ruled out by history, examination, and appropriate laboratory studies. Osteomalacia, with the numerous causes listed previously, must be ruled out as a cause of osteopenia and can be readily evaluated by histomorphometric analysis of a bone biopsy of the iliac crest. Additionally, accurate serum measurements of vitamin D metabolites and PTH levels are now widely available, facilitating evaluation of calcium metabolism. Patients with osteoporosis often present with fractures after minor trauma as the first indication of the disorder. Frequently backache with progressive kyphosis and loss of height are noted. Common locations for fractures are the distal radius, proximal femur, and vertebrae. With acute fracture, patients may have severe back pain, which gradually subsides as the fracture heals over 2 to 4 months. Often some residual aching pain persists, but neurologic deficits are extremely rare, even with severe compressions and kyphotic deformity.

Radiology

Radiographically, thinning of the cortices of long bones or vertebral bodies may be noted, with a loss of trabecular pattern and bulging of the discs into the vertebral end plates (Fig. [41-45](#)). Plain radiographs are notoriously inaccurate in the diagnosis of osteoporosis, since the apparent bone density is strongly dependent on radiographic technique. Accurate densitometric measurements have become widely available using quantitative CT scans of the spine, or dual-energy x-ray absorptiometry, which can measure the density of any bone. The accuracy of these techniques is within 1 percent. With dual-energy x-ray absorptiometry, measurements of both the spinal and femoral neck density usually are made to assess the appendicular and the axial skeletal mass. The analysis of spinal mass is demonstrated in Fig. [41-46](#), which also shows the age- dependent decline in the normal female population, with acceleration during menopause. By comparing the bone density of osteoporotic patients with that of normal control populations, an estimate of the risk of fractures can be made, which is valuable in guiding therapy. Sequential measurements are also helpful in assessing responses to therapeutic interventions.

Treatment

Acute vertebral fractures are treated with a lightweight extension brace, analgesics, and early mobilization. Bed rest aggravates the underlying osteopenia. Other fractures are treated as are those in nonsteoporotic patients (see , below). Endocrine causes, multiple myeloma, and osteomalacia must be ruled out as described above. Calcium supplementation (1,000 to 1,500mg daily) is generally recommended and can slow the rate of loss. Physiologic doses of vitamin D (400 U daily) help to ensure calcium absorption. In perimenopausal women, estrogen therapy is of proved benefit. Patients who have high-turnover osteoporosis (excessive bone resorption) as judged from histomorphometric or calcium balance studies may benefit from antiresorptive agents such as calcitonin. Bisphosphonates such as etidronate have been used cyclically (a 2-week course every 3 months) to inhibit bone resorption. Since bone is continually forming and resorbing, periodically inhibiting resorption can increase bone mass, since formation continues. Second-generation bisphosphonates, such as alendronate, have been used successfully to increase bone mass in postmenopausal osteoporotic patients by continuous daily oral administration. Fluoride has been used to increase bone formation and bone mass but remains controversial since the new bone formed is woven bone, which is structurally inferior. Prospective, randomized studies have demonstrated no decrease in fracture rates with fluoride therapy. Synthetic androgenic steroids also have been used to increase bone formation rates, as has human growth hormone, but these approaches are still experimental. All osteoporotic patients must be encouraged to begin a regular, progressive program of weight-bearing exercise such as walking, along with general spinal extension and strengthening exercises to help maintain bone mass and prevent fractures.

Pituitary Disturbances

Pituitary Short Stature

There are two main types of pituitary dwarfism, Fröhlich's adiposogenital type (short stature associated with obesity, genital hypoplasia, and mental retardation), and the Lorain-Lévi type (short stature without other mental or physical change). The usual cause is a tumor or cyst compressing the pituitary gland, although congenital aplasia also can occur. With the recent availability of recombinant human growth hormone, this form of dwarfism is now treatable.

Hyperpituitary Syndromes

Gigantism is caused by excessive growth hormone secretion during childhood and can be accompanied by subnormal mental development. The bones are increased in thickness and length. *Acromegaly* refers to the syndrome associated with excessive growth hormone secretion in the adult. Abnormal bone formation enlarges the alveolar margins of the jaws, leading to elongation of the face and projection of the chin. Prominence of the frontal region of the skull develops as well as increased size of the thorax. The ends of the long bones are enlarged, and the short bones in the hands and feet are elongated and thickened. The usual cause is a pituitary adenoma. Acromegalic patients also may develop arthritis of the spine or other joints that resembles osteoarthritis.

Hypothyroidism (Cretinism)

Congenital hypothyroidism leads to short stature, retardation of maturation, and developmental delay. The principal skeletal changes include decreased length of the long bones, thickened cortices, and delayed appearance of the secondary ossification centers in the epiphyses. Irregularity of the ossific nucleus resembles osteochondroses such as Legg-Calvé-Perthes disease in the hips. Closure of the growth plates also is delayed. Thyroid hormone replacement therapy, if started in infancy, cures the disease, and it can produce growth in stature even in adults because of the delayed closure of the growth plates.

Mucopolysaccharidoses

A series of 12 hereditary disorders of mucopolysaccharide metabolism has been described, with identification of the specific enzymatic defect in 10 of them. All affected patients have somewhat thickened, coarse facial features, joint stiffness, and short stature. Radiographic findings include oval vertebrae with anterior beaking, coxa valga, and a wide, flat pelvis. Many are associated also with thoracolumbar kyphosis. All forms of mucopolysaccharidosis are inherited as autosomal recessive disorders except type II (Hunter's syndrome), which is X-linked. Mucopolysaccharidosis patients have elevated urinary excretion of dermatan, heparan, or keratan sulfate, depending on the type. The most common forms are type I (Hurler's and Scheie's syndromes—deficiency of α -L-iduronidase) and type IV (Morquio syndrome; type IVA—deficiency of N -acetylgalactosamine-6-sulfatase; type IVB—deficiency of beta-galactosidase). Type I is associated with mental retardation and life expectancy of 10 to 15 years, while type IV patients have normal intelligence and survive well into adulthood.

Paget's Disease (Osteitis Deformans)

Osteitis deformans, first described by Sir James Paget in 1876, is a disorder of accelerated regional bone turnover. Ultrastructural studies have demonstrated viral-like inclusion particles in osteoclasts of affected bone, suggesting that the causative agent may be a slow virus. Despite numerous studies, no specific viral etiologic agent has yet been identified. Paget's disease may be monostotic (25 percent) or polyostotic (75 percent). Early in the disease there is excessive osteoclastic resorption and vascularity, followed by abnormal bone formation and sclerosis, with thickening of the trabeculae and cortical bone. In the late phase, dense sclerotic woven bone and marrow fibrosis predominate. The disease begins between the ages of 35 and 50 years and is painful in about 30 percent of patients. Often the diagnosis is made as an incidental finding on a radiograph taken for some other reason. Resorption and formation remain coupled in Paget's disease, so the excessive bone resorption that appears to be the primary defect is accompanied by excessive formation, causing the enlargement of involved bones with cortical and trabecular thickening. Bowing of involved bones such as the tibia and femur may occur, and arthritic changes may develop in adjacent joints. Common locations are the skull, pelvis, lumbar spine, femur, and tibia. Because of the disorganized collagen in the woven pagetic bone, tensile strength is poor. Therefore fractures can occur, more often starting on the tension or convex side of the long bone, and typically transverse in nature.

Skull involvement can lead to enlargement of the cranium and compression of cranial nerves, producing symptoms such as vertigo, deafness, or visual disturbances. Spinal involvement can cause back pain, ankylosis, and spinal cord compression with neurologic deficit. Serum calcium and phosphate levels usually are normal, but the alkaline phosphatase level is markedly elevated and is correlated with the activity of the disease. The urinary hydroxyproline level is elevated as a reflection of collagen breakdown during bone resorption. In a small percentage of patients with Paget's disease, sarcomatous degeneration, usually to an osteosarcoma, develops later in life. Signs of malignant degeneration include radiographic changes with bone lysis or destruction, soft-tissue mass, and progressive pain. The prognosis of pagetic sarcoma is poor.

Radiologic Findings

Early changes consist of lytic resorption of trabecular bone. In the skull this is referred to as *osteoporosis circumscripta*, and in the tibia the phenomenon resembles a flame-shaped area of advancing bone lysis. Later, coarsening of trabecular pattern with thick striations and cortical thickening and enlargement are observed (Fig. 41-47). Vertebral involvement is characterized by widening and squaring of the vertebral body and a thickened "picture frame" cortex. Associated joints exhibit degenerative changes with sclerosis and joint space narrowing.

Treatment

There is no cure for Paget's disease, although the symptoms and activity of the disease can be controlled pharmacologically. In asymptomatic patients in whom there is minimal concern about impending fracture, observation alone usually suffices. For symptomatic involvement, treatment is focused on antiresorptive agents (e.g., bisphosphonates such as etidronate or alendronate) or calcitonin. Treatment with calcitonin or diphosphonate is continued until biochemical parameters and symptoms improve, usually about 6 months.

Doses of etidronate higher than 5 mg/kg/day, or continuous treatment for periods longer than 6 to 12 months may be associated with an iatrogenic osteomalacia-like mineralization defect and predispose to fractures. Because bisphosphonates have an antiphosphaturic effect, a slight rise in serum phosphate indicates an effective dose of the medication, which varies depending on the extent of bony involvement. Patients who undergo major surgery should be treated in the perioperative period with calcitonin to prevent postoperative hypercalcemia secondary to immobilization. In cases of spinal involvement with neurologic deficit or impending paraplegia, mithramycin is the drug of choice, producing an immediate and profound inhibition of pagetic activity. It is not suitable for long-term therapy because of hepatic and renal toxicity.

When fractures occur, the incidence of nonunion is greater than normal, and healing more reliably obtains with surgical internal fixation in addition to antipagetic therapy. Sarcomas must be treated by radical resection and prosthetic limb reconstruction or amputation and, if the patient can tolerate the toxicity, chemotherapy.

Bone Marrow Diseases

Diseases of bone marrow constituents can have secondary effects on trabecular bone. The reticuloendothelial tissue of bone is found mainly at the ends of the long bones and in the cancellous bone of the axial skeleton (ribs, spine, skull, pelvis).

Lipoid Granulomatosis

These disorders result from disturbances in lipid metabolism within bone, causing accumulations that displace normal marrow elements. In Gaucher's disease a cerebroside lipoprotein accumulates in histiocytes in the liver, spleen, and bone marrow. Orthopaedic problems include pathologic fractures and avascular necrosis of the femoral head. Symptomatic Gaucher's disease has been treated successfully by exogenous parenteral recombinant alglucerase 3 times per week, with resolution of the bone lesions. Niemann-Pick disease and Tay-Sachs disease involve defective phosphatide lipid and cerebroside proteins, respectively, with primarily neurologic sequelae. All of these disorders can cause formation of tumorlike deposits within the bone marrow, with displacement of normal marrow and trabecular bone. The deposits consist primarily of lipid-laden histiocytes, or "foam cells," and the lesions cause bone destruction without much bony reaction, occasionally resulting in pathologic fractures.

Mastocytosis

Mastocytosis is a systemic disorder that infiltrates bone marrow by mast cells. Release of histamine, serotonin, and other mediators from the mast cells cause characteristic urticaria pigmentosa, a dermatologic condition, as well as pulmonary and a wide range of other symptoms. In the skeleton, mastocytosis can cause lytic or mixed lytic and sclerotic lesions, which may be localized or widespread. The cause of the disorder is unknown, and treatment generally is symptomatic.

Histiocytosis X

This term encompasses a spectrum of clinical disease, with three major forms. Letterer-Siwe disease is the infantile form, which involves hepatosplenomegaly and disseminated bony lesions, and runs a rapidly fatal course in most cases. Hand-Schuller-Christian disease usually occurs in children but occasionally presents in adulthood. The severity varies greatly, and manifestations include the triad of exophthalmos, diabetes insipidus, and skull lesions. In addition, hepatosplenomegaly and hypercholesterolemia can be present, and the disease tends to be progressive.

Eosinophilic granuloma, the least severe form of histiocytosis X, usually presents as a solitary bony lesion that may be painful or cause a pathologic fracture. Two-thirds of patients are under age twenty. Patients who develop multiple lesions usually do so within 2 years of onset, and may go on to develop the Hand-Schuller-Christian variant of the disease. Local tenderness and swelling may be present. Radiographically, eosinophilic granuloma has a highly variable appearance and may cause destructive or permeative bone lysis with periosteal reaction mimicking osteomyelitis or Ewing's sarcoma. In general, however, the radiograph shows well-circumscribed lytic lesions. Involvement of a vertebral body can cause flattening (vertebra plana, or Calvé's disease) and must be differentiated from osteomyelitis.

Treatment

Treatment of systemic forms of histiocytosis is with chemotherapeutic agents such as vinblastine and prednisone. Eosinophilic granuloma can be treated with low-dose radiation (2 to 10 Gy), although in long bones at risk of pathologic fracture, curettage with or without bone grafting (depending on the extent of the lesion) may be necessary. Local injection of steroids into symptomatic bony lesions has shown promise in inducing healing.

Chronic Multifocal Recurrent Osteomyelitis

This is a rare disorder of children that presents with swelling, pain, and radiographic findings of lysis, reactive sclerosis, and periosteal reaction resembling osteomyelitis, generally in a metaphyseal location adjacent to a growth plate. Biopsy shows acute and chronic inflammatory cells resembling osteomyelitis, but bacterial cultures are negative, and the disorder is self-limited. Antibiotics have not been proved to have any efficacy in the disorder, and multiple foci may become active at different times during childhood and resolve spontaneously. The disorder has a variable course, but it disappears by skeletal maturity. The disease resembles eosinophilic granuloma in radiographic appearance and clinical behavior, although eosinophils and hepatic or splenic involvement are absent.

Lymphatic and Hematopoietic Systems

Hodgkin's Disease

Hodgkin's disease may involve bone marrow, most commonly in the vertebrae or pelvis. Although bony involvement is frequent, presentation as a primary bone lesion without lymphatic involvement is unusual. The radiographic appearance may be lytic, blastic, or mixed. Generally, dull aching pain is the initial sign of bone involvement. The bony lesions of Hodgkin's disease are responsive to radiation treatment, and the systemic disease responds well to chemotherapy.

Leukemia

Lymphoblastic leukemia is the form that most frequently causes bony changes. The typical radiographic finding is a transverse zone of lucency in the metaphysis adjacent to the growth plate. Diffuse spotty osteopenia and vertebral compression fractures occur, and rarely large focal lytic lesions in long bones are seen. Treatment is aimed at controlling the systemic disease with chemotherapy.

Multiple Myeloma

Multiple myeloma, the most common primary malignancy of bone, is a malignant proliferation of plasma cells within the bone marrow. It usually affects individuals above the age of fifty. There is marrow replacement with tumor cells, and usually secretion of abnormal clonal immunoglobulins. The diffuse marrow involvement leads to anemia and punched out lytic lesions throughout the skeleton (Fig. 41-48). Because there is little reaction to the lesions, bone scans may not demonstrate all lesions, and hence this is one of the few skeletal malignancies in which a skeletal survey using plain radiographs can give a better assessment of the extent of the disease. The massive bone resorption can cause hypercalcemia, and pathologic fractures are common. Abnormal paraproteins produced by the plasma cells can cause renal glomerular damage and amyloidosis. In 50 percent of patients immunoglobulin can be detected in the urine (Bence-Jones protein). Serum electrophoresis may demonstrate the presence of an abnormal globulin, although immunoelectrophoresis is more reliable in establishing the presence of a monoclonal gammopathy.

Clinical Manifestations

Pain in areas of bony involvement is a common initial symptom, as is fatigue. Backache and pathologic vertebral compression fractures are also frequent, and the disease has a male preponderance. The multiple nature of the bony lesions, absence of pulmonary metastases, and osteolytic character of the lesions often suggest the diagnosis prior to bone marrow biopsy or demonstration of a serum or urine paraprotein. Multiple myeloma is a uniformly fatal disease, although it can be controlled for a number of years by systemic therapy.

Treatment

The usual regimen of chemotherapeutic agents includes cyclophosphamide, melphalan, prednisone, and vincristine. The possibility of bone marrow transplantation is under investigation as a potentially curative treatment but remains experimental. Orthopaedic surgical stabilization of pathologic fractures or impending fractures plays an important role in the management and maintenance of ambulatory function. Spinal stabilization and decompression of the cord are often necessary in cases with neurologic deficit that fails to respond promptly to radiation treatment.

Solitary Myeloma (Plasmacytoma)

Occasionally solitary plasma cell lesions occur in the long bones, spine, or pelvis. Seventy percent of these patients progress to multiple myeloma, and a circulating monoclonal paraprotein occasionally is found. The usual treatment involves radiation therapy with or without surgical resection, depending on the anticipated morbidity of the surgical procedure. Chemotherapy is usually reserved for those patients who progress to the multiple form.

Hemolytic Anemia

Both thalassemia (Cooley's anemia) and sickle cell anemia produce bone marrow changes in the axial and appendicular skeleton. In the skull, the expansion of the hematopoietic marrow can exhibit a “hair on end” or “sunray” appearance, and bone infarctions can be observed as serpiginous calcified densities in the metaphyses of the long bones. The femoral head may undergo avascular necrosis. Rarely, expansile pseudotumors consisting of hyperplastic bone marrow occur in association with thalassemia.

FRACTURES AND JOINT INJURIES

General Considerations

Definitions

A *fracture* is defined as a linear deformation or discontinuity of bone produced by forces that exceed the ultimate strength of the material. Deformation without fracture can occur with loads that exceed the elastic limit of the bone but not its ultimate strength. This is referred to as *plastic deformation* and is more common in children. *Pathologic fractures* occur when the strength of the bone is below normal, as in infections, tumors, or metabolic bone disease, or after the creation of surgical defects in bone. The direction and magnitude of force applied to a bone and the rate of loading all are important in determining the fracture pattern that will result.

Fractures are described anatomically according to location in the bone (intraarticular, epiphyseal, metaphyseal, diaphyseal), the plane of the fracture (transverse, oblique, spiral), the number and type of fragments, and whether the fracture is open (compound) or closed. Some fracture patterns are illustrated in Fig. [41-49](#). A spiral fracture is produced by torsional force, and *comminution* refers to presence of multiple fracture fragments.

In an undisplaced fracture a plane of cleavage exists between the fracture fragments without separation. For displaced fractures, the convention is to describe the direction of displacement of the distal fragment with reference to the proximal fragment (medial, lateral, posterior, etc.). Angulation refers to angular deformity between the long axes of the fracture fragments and is also described in terms of the distal fragment's relation to the proximal fragment. Rotational deformity also is expressed in terms of the movement of the distal fragment (internal or external) relative to the proximal fragment.

One of the most important distinctions is whether a fracture is *open* or *closed*. If a fracture communicates with the surface of the skin or mucous membranes, infection is a risk, and this constitutes an orthopaedic surgical emergency. A *stress fracture* occurs when a bone is subjected to repetitive stresses that individually are insufficient to cause fracture but cumulatively lead to fatigue failure. A *compression fracture* results from axial loading of bone with compaction of bony trabeculae; these are seen generally in vertebral bodies. A *greenstick fracture* is an incomplete fracture resulting from failure of a portion of the cortex under tension, with part of the opposing cortex still intact but plastically deformed; these usually are seen in children. A *torus fracture* also retains partial cortical continuity, but with buckling or failure in compression of the opposing cortex (Fig. [41-50](#)).

Diagnosis

The clinical manifestations of a fracture include pain, swelling, deformity, ecchymosis, instability, and crepitus. The diagnosis usually is confirmed radiographically with two radiographs taken at right angles to each other. Joints above and below the fracture site should be included in the radiograph to rule out associated injuries. Occasionally a completely nondisplaced fracture is not apparent on initial films; in this event immobilization with follow-up films 1 to 2 weeks later is indicated. The fracture line can be better visualized after some resorption and early periosteal repair reaction have taken place.

Evaluation of the Injured Patient

Immediate threat to a patient's life from an injured extremity is unusual but can be a consequence of hemorrhage and resulting shock. Associated injuries to the chest, head, and abdominal viscera are potentially more serious and require immediate evaluation and treatment priority. Multiple fractures, even when closed, can cause shock from internal hemorrhage, particularly if major pelvic injuries are present. A closed femoral fracture can readily result in 1 to 2 units of internal blood loss, and shock in adult patients can occur with hypovolemia of 1 to 2 L. Shock is treated by volume replacement emergently with crystalloid such as lactated Ringer's solution to restore blood pressure and perfusion, and as soon as possible with whole blood. Early aggressive treatment of hypovolemic shock greatly reduces the likelihood of morbidity and mortality. Use of pneumatic trousers (MAST—medical anti-shock trousers) during transport may help to maintain blood pressure and decrease blood loss from lower-extremity injuries. Assessment of possible spinal injury is imperative, and transport using a backboard and sandbags or other head supports helps to minimize the chances of causing additional injury. The patient should not be allowed to sit or stand until appropriate spinal radiographs have been taken to evaluate spinal stability.

Evaluation of the Injured Extremity

The injured extremity is evaluated as quickly as possible for neurovascular compromise, soft-tissue and bony injuries, and joint instability. Peripheral pulses and capillary refill are evaluated, and motor and sensory examinations are carried out to the extent of the patient's ability to cooperate. All findings must be carefully documented in the event of later changes.

Emergency Splinting of Fractures

After the neurovascular examination has been performed and soft-tissue trauma or wounds evaluated, fractured extremities are splinted to minimize further injury. Plaster splints, pillow splints, or air splints can be used to stabilize the extremity. Fractures involving the humerus or shoulder can be splinted with a sling. Fractures of the femur are best temporarily stabilized in a traction splint.

Open Fractures

Open fractures constitute an orthopaedic surgical emergency because of the risk of deep infection. Open fractures are generally higher-energy injuries, resulting in more comminution and soft-tissue injury and, consequently, greater impairment of bone blood supply. All these factors contribute to the increased risk of osteomyelitis. Infection in a fracture, once established, can be extremely difficult to eradicate and markedly increases the risk of nonunion. The major aim in treatment of open fractures is the prevention of infection. This is best accomplished by aggressive and immediate debridement of the wound and fracture site in a sterile operative environment, and initiation of empiric intravenous prophylactic antibiotic therapy. Debridement is optimally done within 8 h, and a repeat intraoperative wound inspection and debridement within 24 to 48 h is recommended if significant soft-tissue damage or loss is present. Primary closure of open fractures is rarely, if ever, indicated, and secondary closure after 5 to 7 days or plastic surgical soft-tissue coverage procedures are preferable.

Classification

Open fractures are classified as type I, II, or III, depending on the associated soft-tissue injury. A puncture wound or communication less than 1 cm in length is a type I open fracture. Type II fractures have a wound larger than 1 cm with moderate associated soft-tissue damage. Type III open fractures involve severe soft-tissue injury or loss and are subdivided into subtypes A—soft-tissue injury only; B—severe soft-tissue and bone injury/soft-tissue loss; and C—associated neurovascular injury.

Technique

The following is a general description of the initial management of open fractures.

- (1) The wound is cultured and covered with a sterile bandage, and the extremity is splinted.
- (2) Cephalosporin and, with type III or grossly contaminated wounds, penicillin and an aminoglycoside are administered intravenously. Tetanus toxoid or antitoxin is administered, depending on tetanus immunization status.
- (3) In a sterile operating room environment, the patient is anesthetized, and the extremity is prepared with antiseptic.
- (4) Skin edges of the wound are excised approximately 1 to 2 mm, and more if clearly avascular or crushed.
- (5) The wound is thoroughly irrigated with pulsatile lavage using several liters of saline solution.
- (6) Any devitalized muscle or debris is surgically excised, and the fracture site is exposed. The fracture surface is curetted to remove foreign material, followed again by copious irrigation of the wound with pulsatile jet lavage of saline solution for a total volume of at least 9L.
- (7) The fracture is reduced and the skin loosely approximated over a drain, leaving a significant area of the wound open, but covering exposed bone, neurovascular structures, or

tendons if possible.

(8) The fracture is stabilized with plaster immobilization or, more often, an external fixator to allow wound access. Some type II open fractures can be managed by primary internal fixation with intramedullary devices in the lower extremity or plating in the upper extremity at the time of primary debridement. Additionally, disrupted articular surface fragments can be provisionally stabilized with judicious use of pins or screws.

(9) Antibiotics are continued for a minimum of 10 to 14 days postoperatively. Repeat wound inspection and debridement are recommended after 24 to 48 h, especially for type III injuries.

(10) If secondary wound closure to achieve bone coverage is not feasible, skin grafts, free tissue transfers, or muscle flaps can be used to obtain adequate soft-tissue coverage for the fracture.

Vascular Injury

Major arterial injury should be suspected in any fracture-dislocation or significant trauma to an extremity. If perfusion to the extremity is disrupted, the maximum time that can elapse before onset of irreversible ischemic damage to muscle and other tissues is 6 to 8 h. Certain injuries are more likely to have associated vascular damage, including supracondylar humeral fractures, knee dislocations, femoral shaft fractures, type III open tibial fractures, and gunshot wounds.

The diagnosis of vascular compromise is not necessarily straightforward, since patients often are unconscious, and peripheral vasoconstriction or arterial spasm may be present. Capillary refill must be assessed, and if pulses are not palpable Doppler examination can be helpful. If there is any doubt about vascular integrity, arteriography as well as appropriate surgical exploration should be carried out.

Compartment Syndrome

One of the most serious complications of extremity trauma or ischemic injury is compartment syndrome, which was discussed earlier (see Contracture). A brief review is appropriate given the extreme importance of early diagnosis and treatment of this complication. The cardinal signs of pain, pallor, pulselessness, and paresthesias are present to variable degrees. Pain with passive stretch of muscles is one of the more reliable indicators of compartment syndrome, and accurate diagnosis is readily made by measurement of intracompartmental pressures using a slit catheter. Pressures in the range of 30 to 40 mmHg constitute an indication for fasciotomy. In patients with prolonged ischemia due to arterial compromise, prophylactic fasciotomies of all compartments distal to the vascular injury should be done concomitantly with reestablishment of perfusion, regardless of whether signs of compartment syndrome are present. In the presence of compartment syndrome, skin closure is contraindicated. The swollen muscle will cause gaping of the fasciotomy incisions, which can be treated with dressing changes and secondary split- thickness skin grafting or with gradual reapproximation of the wound edges using wire sutures or tape strips sequentially tightened daily.

Fat Embolism and the Acute Respiratory Distress Syndrome

Patients who sustain multiple fractures are at high risk for subsequent fat embolism, in which fat droplets from bone marrow enter the systemic circulation and impair pulmonary capillary perfusion via a complex mechanism. The final common pathway of fat embolism and other injuries that result in pulmonary parenchymal dysfunction after multiple-system trauma is severe hypoxemia, or the acute respiratory distress syndrome (ARDS). Fat embolism generally occurs within 24 to 72 h of injury and presents with hypoxemia, tachycardia, tachypnea, fever, restlessness, and confusion. The syndrome is fatal in 10 to 15 percent of cases. Chest radiographic findings are similar to those of other causes of ARDS, with bilateral patchy infiltrates. Petechiae may be present transiently in the axilla, chest, and conjunctiva and thrombocytopenia may occur, with fat droplets visible occasionally in blood specimens and in the urine.

Fat embolism syndrome occurs after total hip and total knee replacement as well as after trauma, but is seen most often following femoral fracture. Treatment of fat embolism syndrome is similar to treatment of ARDS, with administration of oxygen, ventilatory support, and positive end-expiratory pressure as needed to maintain a partial pressure of oxygen of 60 mmHg or better. Moderate-dose corticosteroids (methylprednisolone 9/mg/kg) given prophylactically after trauma has been shown to reduce the incidence of fat embolism from 28.8 to 2.5 percent. The use of corticosteroids in ARDS from causes other than fat embolism remains controversial, however. The other important factor in the management of the multiple- trauma patient that decreases the incidence and severity of fat embolism and ARDS is stabilization of the fractures within the first 24 h of injury. Specific medications such as low-molecular-weight dextran, heparin, and alcohol have not been proved to alter the outcome.

Peripheral Nerve Injuries

Extremity trauma is sometimes accompanied by injuries to peripheral nerves. In the least severe type of injury, neurapraxia, there is interruption of nerve conduction, which will ultimately recover, manifested by a transient complete or partial loss of motor and sensory function. The mechanism of injury is stretch or contusion, and resolution usually occurs within 2 to 3 months. With more severe stretch injuries, axonotmesis, or disruption of the axons with retention of the Schwann cell sheath, can occur. Axonal regeneration may occur with this injury, but only slowly (approximately 1 mm daily), and recovery may be incomplete. In neurotmesis, or complete division of the nerve, regeneration will not occur spontaneously, and surgical repair is necessary.

Assessment of the degree of nerve damage can be difficult unless open reduction of the fracture is needed, in which case the nerve can be explored. Careful documentation of nerve function is essential, as loss of function in a nerve after closed reduction of a fracture is one indication for surgical exploration. In cases of neurotmesis, microsurgical reapproximation of the nerve with fine epineural sutures gives the best chances of recovery. In cases of segmental nerve loss, cable grafting with several lengths of an expendable nerve such as the sural nerve allows some degree of regeneration. Nerve repair can be technically easier after 7 to 10 days, when some thickening of the epineurium due to scarring has developed. Persistent motor deficits in extremities can be managed with orthoses or, later, tendon transfers, depending on the functional loss and remaining available innervated muscles.

Fracture Healing

After fracture, hematoma develops at the fracture site and a clot is formed. Local mediators incite an inflammatory response, and necrosis of bone adjacent to the fracture site occurs as a consequence of disruption of its blood supply. In fractures with severe soft-tissue injury or loss, periosteal stripping, or comminution, the extent of bone necrosis can be significant and the fracture may be delayed in healing or fail to unite.

Stages

Fracture healing occurs in several stages or phases. At the time of fracture (stage of impact), the energy absorbed to failure determines the degree of comminution, soft-tissue injury, and disruption of the bone blood supply. The hematoma, which organizes into a fibrin clot, releases cytokines that attract inflammatory cells from the circulation, initiating the inflammatory stage of fracture healing, which lasts from a few days to 2 weeks. The cytokines released from platelets in the clot, such as platelet-derived growth factor (PDGF) and TGF- β , as well as other factors released from inflammatory cells, are probably involved in stimulating undifferentiated mesenchymal cells (which appear to be derived in large part from the periosteum) to undergo differentiation into fibroblasts, osteoblasts, and chondrocytes. This is accompanied by proliferation of fibrovascular tissue (granulation tissue) in the area of the fracture gap. These events signify the onset of the early reparative stage (or soft callus stage). The osteoblasts form adjacent to the periosteum, while in the more hypoxic area of the fracture gap, differentiation into hyaline cartilage is favored (Fig. 41-51). This fracture callus begins to stabilize the fracture ends and limit motion, leading to progressive vascular ingrowth. The cartilaginous callus, which tolerates hypoxia well and in fact undergoes cellular hypertrophy and matrix mineralization under hypoxic conditions (see Endochondral Ossification, above), begins to go through the endochondral sequence of mineralization. This late reparative stage (or hard callus stage) results in increasing stability of the fracture, and increasing membranous or osteoblastic ossification from the periphery of the callus gradually replaces the endochondral process.

The woven bone is later remodeled into true lamellar bone, a stage that can last from months to years. During this phase the limb is mechanically functional and remodels along lines of stress toward its original shape, with reconstitution of the medullary cavity. Table 41-12 summarizes the events in fracture healing by callus formation. Teleologically it appears that the function of the endochondral calcification is to stabilize the bone, allowing subsequent vascular growth across the fracture gap to support the more aerobically dependent osteoblastic bone formation. If a fracture is rigidly internally fixed with a metal plate under compression, no callus forms, and the fracture heals by primary remodeling through the activity of osteoclasts and osteoblasts without endochondral ossification.

Factors Influencing Healing

The importance of the periosteum in fracture healing is supported by the observation that fracture healing is accelerated in children, who have much thicker and more cellular periosteal bony coverings. In adults the periosteum is noticeably thinner, and healing is slower. Rates of fracture healing in young and old adults are similar, barring that presence of metabolic bone disease or nutritional deficiencies. Many investigations in animals and human beings have attempted to evaluate substances that stimulate fracture healing, including growth hormone, PTH, various vitamins, and prostaglandins. At this point the evidence is controversial at best. There is no strong evidence that normal fracture healing is accelerated with these types of treatment, although some beneficial effect on delayed union or nonunion remains a possibility. Head-injured patients, who often form abnormal bone and cartilage in muscle or other inappropriate tissues (heterotopic ossification), have accelerated healing of fractures; a factor in the serum of these patients is an anabolic stimulator of osteoblasts in culture, though this factor has not yet been identified.

Articular Cartilage Healing

Unlike bone, articular cartilage has a very limited ability to undergo repair after posttraumatic damage. Articular cartilage has a very orderly structure, with an organic matrix composed of 40 percent proteoglycans, 40 percent type II collagen, and 20 percent glycoproteins, growth factors, and minor collagens such as type IX and type XI (Table 41-13). The collagen fibers in the most superficial layers of cartilage are oriented parallel to the joint surface and gradually change direction to a radial orientation in the deeper layers, with a somewhat random orientation in the transitional zone (Fig. 41-52 A). The proteoglycan is a large macromolecular aggregate of subunits consisting of a protein core (aggrecan) with covalently bound branching sulfated carbohydrate groups called glycosaminoglycans (Fig. 41-52 B). The predominant glycosaminoglycans are chondroitin and keratan sulfate. These monomer subunits are noncovalently bound to a filament of hyaluronic acid by a small glycoprotein called a link protein. Proteoglycans are responsible for the compressibility, hydration, and mechanical integrity of cartilage.

The earliest responses to an injury such as a laceration of the articular surface are loss of matrix proteoglycans and an attempt, usually unsuccessful, by chondrocytes adjacent to the injury to resynthesize the matrix. Progressive degradation of matrix macromolecules ensues, and the cartilage may fibrillate or split along the radially oriented deeper collagen fibers. If a defect is near the synovial attachment of the joint, cells from this area may migrate into the defect and form fibrocartilage repair tissue. Usually, however, the defect remains and can trigger further areas of progressive degeneration.

If the cartilage defect penetrates the subchondral bony plate, cells from the bone marrow can migrate into the defect and will also allow fibrocartilaginous repair. Fibrocartilage,

unlike hyaline cartilage, is mostly composed of type I rather than type II collagen and has inferior wear characteristics, gradually degenerating over a period of years. Fibrocartilaginous repair can be augmented and some type II collagen formation stimulated by continuous passive motion of the joint. These properties form the basis for current approaches to treatment of chondral defects as well as the rationale for abrasion arthroplasty, or abrading damaged areas of cartilage down to a bleeding bony surface followed by passive motion and protection from weight bearing.

Experimental techniques have evolved to enhance hyaline cartilage repair in animal models as well as in human beings. An approach developed by Caplan and coworkers involves isolation of undifferentiated mesenchymal stem cells from bone marrow aspirates, which are then expanded in tissue culture and implanted in a collagen gel into an articular defect, and results in differentiation into hyaline cartilage. A second approach, developed by O'Driscoll and Salter, uses a periosteal graft attached to the chondral defect by sutures with the cambium layer of the periosteum facing the synovial fluid. This leads to resurfacing of the defect with hyaline cartilage, which can be further enhanced by incubation of the periosteum briefly with TGF- β before reimplantation. Peterson and coworkers have used autologous chondrocyte transplantation to resurface defects in human knee joints. Cartilage is removed arthroscopically from nonweight-bearing regions of the articular surface, and chondrocytes are isolated by enzymatic digestion and expanded in tissue culture. The chondrocytes are then injected in a collagen gel beneath a periosteal flap sutured to the articular surface over the defect. Early clinical results are encouraging, but these procedures are controversial and the long-term outcome uncertain.

Ligament Healing

Bony injuries to the extremities are frequently accompanied by ligamentous injuries. Ligament injuries also can occur independently as a primary injury to a joint; in general these injuries are referred to as sprains. Ligaments heal by progressive scar formation and contracture, and recent studies indicate that repair is improved by early motion in the absence of gross instability. Ligament injuries are graded as type I (stretch but no disruption of fibers), type II (tear of some of ligament fibers), and type III (complete mechanical discontinuity of the ligament). Type I and type II ligament injuries generally are treated by immobilization, or protected motion. Motion and mechanical strain enhance ligament healing. Treatment of type III ligament injuries depends on many factors and can range from nonintervention (as for a type III acromioclavicular separation) to immediate surgical repair or reconstruction (as for a torn anterior cruciate ligament in a competitive athlete).

Delayed Union and Nonunion

Delayed union is a somewhat arbitrarily defined term applied to fractures that take longer than average to heal and must be considered in context of the type of fracture and age of the patient. *Nonunion* refers to a condition in which a fracture fails to show progression toward union and in which healing is not expected even with prolonged immobilization. Radiographic examination can show sclerotic bone ends, a persistent fracture gap, and rounding off of the fracture ends. Gross motion can be demonstrable clinically or under fluoroscopic stress testing. The fracture gap usually contains nonmineralized fibrocartilage, although at times a true synovial pseudarthrosis, or false joint, is present. Factors predisposing to nonunion include excessive motion or inadequate immobilization of the fracture, interposed soft tissues, extensive soft-tissue damage, periosteal stripping, devascularization of the bone, and infection. Certain bones have more of a predisposition to nonunion, probably attributable to suboptimal orientation of the local bone vascular supply (such as the tibia, femoral neck, or carpal scaphoid).

Nonunion usually requires surgical intervention such as bone grafting with autogenous or vascularized fibular bone. A growing body of evidence supports the use of electromagnetic stimulation, which can induce a certain proportion of nonunions to heal without surgery. Also, one report suggests that administration of levodopa, which is known to cause a sustained increased level of growth hormone in the serum, can be beneficial in stimulation of healing of ununited fractures. During a 6-month treatment period, over 80 percent of a group of multiply operated patients with nonunion and failed bone grafts healed with this regimen.

Pathologic Fractures

Fractures that occur through bone abnormally weakened by a preexisting condition are termed pathologic. Often these fractures result from significantly less than the usual degree of force needed to cause a fracture. The underlying process may be either systemic (as in osteoporosis) or local (as in a bone tumor or cyst). An example of a pathologic fracture secondary to metastatic carcinoma is shown in Fig. [41-53](#). Pathologic fractures often require surgical internal fixation to achieve healing, and in patients with carcinomas that commonly metastasize to bone (e.g., breast, prostate, lung, kidney, and thyroid tumors), prophylactic fixation of large lesions in weight-bearing bones decreases morbidity and improves quality of life. Pain on weight bearing is a reliable clinical sign of an impending fracture.

Stress Fractures

Stress fractures are the eventual result of repeated stress to a bone, ultimately causing fatigue failure. The metatarsals, tibia, fibula, and calcaneus are common sites. Initially radiographs may be negative, but later periosteal reaction will become evident. Nuclear bone scans or MRI can readily demonstrate the lesion in case the radiographs are negative (Fig. [41-54](#)). Treatment usually is conservative, with immobilization and discontinuation of the causative activity.

Growth Plate Injuries

Longitudinal bone growth occurs as previously described in the growth plates, which lie between the epiphyses and metaphyses of the long bones. In children, injuries involving the growth plate are relatively common, and the fractures usually involve the zone of provisional calcification, the mechanically weakest region. Since the germinal cells in the upper growth plate usually are not damaged, fractures that do not cross the plate tend to heal without growth disturbances. The most commonly used classification of growth plate injuries is the Salter-Harris classification, depicted in Fig. [41-55](#). Type I is a separation of the growth plate and epiphysis from the metaphysis. Type II is similar, with a metaphyseal fragment remaining with the epiphysis. Since the growth plate is intact in these injuries, treatment is with closed reduction, and growth disturbance or premature closure of the growth plate is rare. Fortunately, the majority of epiphyseal fractures are of these two types. Because of the proximity of the fracture to the growth plate, a great degree of remodeling capacity exists, and significant angulations and displacements often heal and remodel uneventfully. Type III injuries are intraarticular, traverse the epiphysis and growth plate, and exit through the zone of provisional calcification. Type IV injuries are similar, but exit through the metaphysis. Since both these types of fractures cross and disrupt the growth plate, nothing less than anatomic reduction is acceptable, or growth disturbance will result. These fractures are usually treated by accurate open reduction and internal fixation. Type V injuries are difficult to recognize, and result from an axial load or crush injury to the growth plate that later results in premature fusion. Rang has described an additional physeal injury known as type VI, consisting of damage to the perichondral ring on one side of the growth plate. This can lead to a bony bar on one side of the physis acting as a tether and leading to a progressive angular deformity. Careful follow-up and education of the parents is necessary with these injuries to ensure the best clinical outcome, allow planning of any corrective procedures should growth disturbance occur, and minimize the possibility of misunderstanding.

Fractures in Children

Children's fractures present different problems from similar injuries in adults. Nonunion is extremely rare, and fracture healing is more rapid. The challenge is in recognizing and understanding what degree of displacement, angulation, and shortening is acceptable in relation to remaining growth potential. Fracture healing in a young child stimulates the blood flow to the extremity, and this in turn stimulates increased activity of the growth plates, causing the limb to grow at a faster rate than the uninjured extremity. Furthermore, most fractures in children are treated conservatively, whereas a greater degree of surgical intervention is used in fractures in adults.

Closed versus Open Reduction

The optimal method for handling a specific fracture should allow rapid union, reestablish length and alignment of the extremity, restore complete motion in adjacent joints, and return the patient to functional activity with a minimum of morbidity. Furthermore, cost concerns are an increasing pressure on the health care system, and prolonged hospitalization must be at least a distant consideration, obviously with the welfare of the patient taking precedence. Fortunately technical advances in fracture fixation allow treatment of a wider range of fractures surgically, and prolonged traction in the hospital is much less common than previously. The closed reduction of fractures has the advantages of minimal risk of infection and no further disruption of bone blood supply, and the disadvantages of less precise reduction and prolonged immobilization. Treatment must be individualized to the injury and patient.

Application of Plaster Casts

Circumferential rigid dressings of plaster or fiberglass are an important tool in the treatment of fractures. Plaster is composed of anhydrous calcium sulfate, which solidifies during an exothermic hydration reaction with water. Plaster sets up in a few minutes after addition of water but does not dry completely for 36 to 48 h.

Complications of Cast Treatment

The maximum temperature achieved during setting depends on the water temperature, and use of hot water can result in burns, particularly if the cast is thick or placed on a surface that reflects heat, such as rubber or plastic, during setting. Water used in casting should be lukewarm. After casting for acute injury or surgery, the extremity should be elevated, as swelling within a rigid container can lead to compartment syndrome, as previously described. This is of particular concern in patients who are unconscious or have sensory impairment in the limb. Unrelieved pain, pallor, loss of sensation, development of paresthesias, poor capillary refill, or diminishing motor function are all indications to bivalve the cast. This should be accompanied by cutting the underlying soft padding layer or dressing gauze, which can also become constrictive in the face of swelling. If the symptoms are not promptly relieved, the cast must be removed and the compartment pressures measured as previously described.

Pressure sores, or decubitus ulcers, can occur rapidly in a cast, and tend to occur over bony prominences such as the heel, olecranon, patella, and ischium. Proper molding of the cast, as well as prompt attention to any complaints of pressure or burning pain in an area beneath the cast, are essential to their prevention. The cast also can be windowed and padded to relieve the pressure and prevent skin breakdown; the window is then replaced to prevent window edema.

Types of Cast

When a cast is applied for an acute fracture, it should include the joints above and below the fracture. A long leg cast extends from the upper thigh to the metatarsal bases, and generally the ankle is placed in neutral dorsiflexion and the knee in about 30 degrees of flexion to more easily allow the patient to clear the floor with the immobilized extremity when on crutches. Proper molding above the femoral condyles is important. A cylinder cast extends from the upper thigh to the ankle above the malleoli; it must be well padded above the malleoli and molded above the femoral condyles to prevent it from sliding downward and causing pressure areas around the ankle. A short leg cast extends from the tibial tubercle to the metatarsals; it should be well molded around the malleoli, above the calcaneal tuberosity, in the arch of the foot, and along the shaft of the tibia. A patellar tendon bearing cast is similar to a short leg cast except that it is extended over the knee anteriorly to the midpoint of the patella and is molded over the anterior femoral condyles and patellar tendon to provide partial weight relief of the tibia when weight bearing. Most lower- extremity casts are used with a cast shoe or boot when weight bearing is allowed.

A double hip spica cast, used for immobilization of the hips, is analogous to a plaster pair of pants, usually applied with the hips in some flexion and abduction, and the knees in slight flexion. A single hip spica includes only one leg, and a one and one-half hip spica includes one leg to the toes and the other to just above the knee. Body casts are applied for immobilization of the spine and extend from the groin to the sternum. With improved methods of surgical spinal fixation, molded plastic body jackets often can be used instead of the more cumbersome body cast.

A short arm cast extends from below the elbow to the proximal palmar crease, allowing free flexion and extension of the digits and thumb. If the thumb is included, the cast is referred to as a thumb spica cast. A long arm cast extends from the upper arm to the proximal palmar crease, usually with the elbow flexed 90 degrees. A shoulder spica cast is a body jacket that extends to include the shoulder and elbow, usually with the shoulder abducted and the elbow flexed.

Orthoses

Orthoses and fracture braces are widely used for nonacute treatment of fractures, allowing greater functional use as the fractures heal. These are made of lightweight polypropylene, Orthoplast, or fiberglass and can be custom molded or used in standard sizes. Humeral, tibial, and forearm fracture braces are readily available and can be tightened with Velcro straps as swelling subsides and muscles atrophy. A knee immobilizer with metal staves often can be used in place of a cylinder cast, and adjustable-range-of-motion knee braces that extend from the thigh to the ankle can be helpful in rehabilitating patients after knee, distal femur, or proximal tibial surgery.

External Fixation

External fixation has been used increasingly in trauma treatment in recent years. Threaded pins are inserted into the bone above and below the fracture site and secured to a rigid, adjustable frame to immobilize the fracture (Fig. [41-56](#)). External fixation is particularly helpful in severely comminuted fractures or those with severe soft-tissue damage or loss, since access to wounds is permitted. Pin tract infections are a common complication (incidence 10 percent) but usually respond to local wound care and antibiotics. Occasionally, persistent infection causes loosening of pins and necessitates removal or replacement of one or more pins. An additional advantage of external fixation is in allowing early motion of the joints above and below the fracture, decreasing joint stiffness. External fixation also can be used to help reduce and maintain reduction of comminuted intraarticular fractures by ligamentotaxis (providing traction across the joint, which reduces fracture fragments due to remaining ligamentous and capsular attachments). This is useful in severely comminuted distal radius and tibial plafond fractures.

Another important use of external fixation is in the treatment of acute multiple-trauma patients with severe pelvic fractures and dislocations. Hemorrhage into the pelvis and retroperitoneum can become a life-threatening problem. Application of an anterior external fixator to provisionally stabilize the pelvis and decrease the intrapelvic volume effectively controls intrapelvic hemorrhage. This can be done under local anesthesia in the emergency department, and involves placing two or three pins in each anterior iliac crest and then affixing them to a rigid anterior frame (Fig. [41-57](#)).

One further application of external fixation is in limb reconstruction using the Ilizarov distraction osteogenesis technique, as described earlier for the correction of limb length deficiency. Preliminary results suggest that external fixation may be useful in repairing bony defects caused by traumatic segmental bone loss. A corticotomy is made above the defect, and the segment is gradually transported distally until it abuts the distal fragment. The resulting proximal defect ossifies by the process of distraction osteogenesis (Fig. [41-58](#)).

Traction

Traction is used in orthopaedics to overcome muscle spasm and apply distraction to fracture fragments to reduce the fracture. Additional uses include immobilization of long bone fractures until early healing occurs, immobilization and distraction for painful joint conditions, and correction of joint deformities or contractures.

Skin Traction

Skin traction is applied by means of tapes attached to the skin. It is used only when low levels of force are needed, since 10 lbs is the maximum amount of force that the skin will tolerate with this method. Skin traction applied to the foot (Buck's traction) often is used to immobilize hip fractures temporarily before surgery. While skin tapes can be used, a foam boot with Velcro straps is more convenient and more widely used. Russell's traction, used for the treatment of femoral fractures in children, can be used as shown in Fig. [41-59](#). Alternatively, separate weights can be applied to the femoral sling and the longitudinal foot, a technique known as split Russell's traction. Bryant's traction can be used in infants and small children with femoral fractures. As originally described, the technique consists of skin traction on both lower legs vertically (hips flexed to 90 degrees with knees extended). However, instances of serious vascular compromise of the extremities have been reported with this method; nevertheless, safety can be ensured by limiting its use to children not older than 2½ years of age and applying the traction at a 45-degree inclination rather than directly vertically.

Skeletal Traction

This involves the placement of a Steinmann pin or Kirschner wire percutaneously through the distal femur or proximal tibia in a transverse direction. A traction bow is attached to the pin and a weight applied. The patient is then placed in balanced suspension as shown in Fig. [41-60](#). Other forms of skeletal traction include halo traction with pins affixed to the skull for cervical spine injuries and olecranon traction overhead for humeral fractures.

Electrical Stimulation

The initial concept that electromagnetic fields or electrical potentials might influence bone formation stemmed from the observed piezoelectric properties of bone, which develops surface charges when mechanically stressed because of its anisotropic nature. Areas of compression develop electronegative potentials, and areas of tension develop electropositive potentials. Increasing clinical and basic research data support the effectiveness of electromagnetic bone stimulation in the healing of nonunions, although the mechanism of action is unknown. The most commonly used form of electrical stimulation involves use of external coils that are attached to a cast and centered over the nonunion site. After 3 to 6 months, union rates of 70 to 80 percent have been reported, but the method remains somewhat controversial, with only a few controlled clinical studies. The biologic events appear to involve stimulation of fibrocartilage at the nonunion site to mineralize. Significant effects of electromagnetic fields on bone formation in various models have been demonstrated by Bassett, Brighton, Rubin, and others. This form of treatment is contraindicated in cases in which the fracture gap is more than 1 cm or half the bone diameter, or in the presence of a synovial pseudarthrosis. Synovial pseudarthrosis can be ruled out by bone scan or aspiration of the site, which will contain synovial-like fluid.

Upper-Extremity Injuries

Shoulder

Full abduction of the shoulder requires motion in the glenohumeral joint, scapulothoracic articulation, acromioclavicular joint, and sternoclavicular joint. Glenohumeral and scapulothoracic motion occur in approximately a 2:1 ratio. For every 10 degrees of forward elevation of the arm, 4 degrees of elevation of the clavicle occurs. Motion at the acromioclavicular joint of 20 degrees occurs during the first 30 degrees of abduction. Because of these complex interactive motions, any disruption of one of these articulations can limit shoulder motion.

Sternoclavicular Joint Injuries

The sternoclavicular joint can be dislocated anteriorly or posteriorly, although anterior dislocations are much more common. These injuries most often occur in motor vehicle accidents or contact sports when force is applied to the shoulder along the axis of the clavicle. While the sternoclavicular joint can be difficult to visualize radiographically, a 40-degree cephalad-angled anteroposterior radiograph, tomogram, or CT scan can aid in evaluation. Differentiation from a Salter type I epiphyseal separation is necessary in young patients, since the medial ossification center often does not appear until age eighteen and fuses by age twenty-five. Posterior dislocations are dangerous because there may be associated serious damage to retrosternal pulmonary or vascular structures; therefore a very careful examination of the patient should be performed. Closed reduction is accomplished by shoulder retraction with longitudinal traction; for posterior dislocations this must be done under general anesthesia. Subsequently the patient is immobilized in a figure-of-eight harness or plaster to maintain shoulder retraction. Chronic unreduced dislocations that are painful may require excision of the proximal clavicle.

Fractures of the Clavicle

Clavicle fractures are common in both children and adults and usually occur at the junction of the middle and distal thirds. The fracture is caused by either direct downward force on the shoulder or indirect force such as occurs in a fall on the extended arm. Clinically there is swelling and tenderness at the fracture site and pain with movement of the extremity.

Treatment

Although the clavicle is difficult to immobilize, nonunion is unusual (incidence 0.5 percent) with conservative treatment, presumably because of the excellent blood supply in this area (Fig. [41-61](#)). The trapezius and sternocleidomastoid muscles tend to pull the proximal fragment superiorly. Fractures are treated similarly in children and adults, with use of a figure-of-eight device to retract and elevate the shoulders (Fig. [41-62](#)). A study comparing figure-of-eight immobilization with a simple sling showed no difference in results. Immobilization is for 4 to 6 weeks, depending on the radiographic evidence of progression toward union. A cosmetic deformity with prominence of the fracture site is the rule, but internal fixation trades the bony prominence for a scar and increases the risk of nonunion. In the rare case in which nonunion does occur, internal fixation with plating and bone grafting usually is successful.

Distal Clavicle Fractures

Fractures of the distal clavicle are less common (incidence 15 percent) and have been classified into three types. Type I fractures, which are lateral to the coracoclavicular ligaments (conoid and trapezoid), are stable and treated with a sling. Type II (interligamentous) fractures exhibit superior displacement of the proximal fragment and detachment of the ligaments. These are unstable, may progress to nonunion, and are best treated surgically by open reduction and stabilization with one or more transacromial pins. Type III fractures involve the articular surface. While stable, they may cause later arthritic symptoms that will require distal clavicular excision.

Acromioclavicular Joint Injuries

Injuries to the acromioclavicular joint result from downward force on the shoulder and are frequent in contact sports. These injuries are primarily ligamentous sprains of three types. The patient presents with pain, tenderness, and swelling in the area of the joint. Shoulder movement is painful, and there can be palpable or visible prominence of the distal clavicle if there is disruption of the coracoclavicular ligaments, which are the major stabilizers of the joint. Type I sprains involve partial tear of the ligamentous capsule, with no injury to the coracoclavicular ligaments. Type II sprains involve complete tearing of the ligamentous capsule and a grade II sprain of the coracoclavicular ligaments, and hence they can exhibit slight superior subluxation of the distal clavicle radiographically. Type III sprains involve complete tearing of both the capsular and coracoclavicular ligaments, with superior subluxation of the distal clavicle. Radiographic diagnosis is aided by comparison views of the joint bilaterally with and without the patient holding a 10-lb weight, which will accentuate the subluxation. The rare Type IV, V, and VI sprains involve severe posterior or inferior displacements and generally require surgical treatment.

Treatment

Type I injuries are treated with a sling until comfortable and then progressive exercises. Type II injuries can be treated similarly. Type III injuries cannot be reduced nonoperatively and leave a cosmetic deformity if untreated. In less active individuals, conservative symptomatic treatment generally gives good functional results. In patients involved in high-demand activities or with persistent symptoms or major concerns about the deformity, surgical repair or reconstruction usually is indicated. Various approaches include stabilization of the clavicle to the coracoid process with a screw (Bosworth) and reconstruction of the coracoclavicular ligaments using the coracoacromial ligament (Weaver-Dunn procedure). In patients with persistent pain, excision of the distal clavicle usually provides satisfactory results.

Fractures of the Scapula

Scapular fractures are most commonly seen as a result of violent trauma such as motor vehicle accidents. Consequently associated injuries such as brachial plexus injury and cardiopulmonary trauma are a concern, and these injuries take precedence. Scapular body fractures rarely cause symptoms and usually heal uneventfully. Even with involvement of the glenoid, given the non-weight-bearing status of this joint results of conservative treatment are usually satisfactory, and therefore symptomatic immobilization of the shoulder generally is preferable to more aggressive intervention. In individuals with more than 10 mm of displacement of a glenoid rim fracture comprising more than one-fourth of the joint surface, or with severe separation, or with articular step-off of more than 5mm, open reduction and internal fixation may be considered.

Acute Anterior Shoulder Dislocations

Dislocation of the humeral head can be anterior, posterior, or inferior to the glenoid. Anteroinferior dislocations can be subcoracoid or subglenoid. Anterior or anteroinferior dislocations, the most common injuries, are caused by a combination of external rotation and abduction, which tears the anterior capsular structures. Anterior dislocation of the humeral head can tear the anterior glenoid labrum, and a compression fracture of the posterolateral humeral head (Hill-Sachs lesion) can result from impingement on the glenoid.

The patient usually holds the arm in slight abduction and is unable to lower it. A flattening of the deltoid prominence or indentation beneath the tip of the acromion is frequently apparent. Sensation over the lateral deltoid must always be assessed carefully before reduction, as injury to the axillary nerve can occur. Because of spasm, motor function of the deltoid cannot be evaluated until after reduction.

Radiographic Findings

These include inferior and medial displacement of the humeral head on the anteroposterior view (Fig. [41-63](#)). An axillary view may be difficult to obtain, but a transscapular lateral view (Y view) will demonstrate the humeral head lying anterior to the glenoid.

Treatment

The dislocation must be reduced as soon as possible, and unrecognized dislocations may still be reducible within the first 2 weeks. More chronic dislocations and those that cannot be reduced by closed means often require open reduction.

While a number of different reduction maneuvers are acceptable, the basic principle involves longitudinal traction and gentle internal rotation. This can be achieved either by having the patient lie supine while an assistant applies countertraction with a sheet around the chest, or by having the patient lie prone with the arm over the edge of the table with a traction weight attached to the forearm. Adequate sedation and muscle relaxation are essential to counteract muscle spasm. After reduction, the patient is placed in a sling and swathe or shoulder immobilizer with the elbow at 90 degrees and the forearm across the abdomen. Postreduction anteroposterior and transscapular lateral radiographs are essential. Range-of-motion exercises are started after 3 to 4 weeks in younger individuals, and sooner in the elderly because of the predisposition to loss of motion.

Recurrent Anterior Dislocations

A small percentage of patients are subject to recurrent dislocations, sometimes with minor trauma or even with active abduction and external rotation of the shoulder. Occasionally patients complain of recurrent sensation of subluxation without frank dislocation. When spontaneous reduction does not occur, the reduction maneuver is as described above. After a brief period of immobilization until comfortable, appropriate rehabilitation focusing on strengthening the internal rotator muscles is begun. Continued episodes lead to articular damage and are an indication for reconstructive surgery. MRI or arthroscopy can be helpful in surgical planning in that tears of the anterior labrum, which lead to persistent instability, can be identified. Surgical repair focuses on repair of labral tears (Bankhart procedure), anterior capsular and subscapularis repair or tightening (Putti-Platt procedure), or a combination of these techniques. Capsulorrhaphy in conjunction with Bankhart repair has been advocated instead of subscapularis shortening procedures because of restoration of more normal shoulder biomechanics.

Arthroscopic repair of labral tears (Bankhart lesions) associated with pain or instability is increasingly common. Recurrence rates for dislocation are somewhat higher than with open repair, but the surgical procedure is minimally invasive. Shoulder arthroscopy also is useful in diagnosis and in decompression of the rotator cuff by acromioplasty. A superior labral anterior and posterior (SLAP) lesion has been described that typically results from a fall on an outstretched arm and can cause pain and clicking in the shoulder with overhead activities. Depending on the size of the labral tear, arthroscopic debridement or reattachment have been successful forms of treatment.

Chronic Dislocations

Patients with painful chronic dislocations may benefit from arthrodesis of the shoulder in a functional position. Occasionally these injuries are seen in elderly individuals with surprisingly few symptoms and may not warrant surgical intervention.

Posterior Shoulder Dislocations

Posterior dislocations are less common and are seen after seizures or motor vehicle accidents. Physical findings include an inability to externally rotate or abduct the arm, which is held tightly at the side and in internal rotation. The posterior shoulder can exhibit prominence compared to the opposite side, and the coracoid process often is more obvious on the affected side.

Radiographic Findings

On the anteroposterior view findings may be subtle, with loss of the distinct “half moon” overlap shadow seen on in a normal shoulder (Fig. [41-64](#)). Axillary or transscapular lateral views will confirm the posterior displacement. Associated lesser tubercle fractures are common, because of avulsion injury of the subscapularis muscle.

Treatment

Reduction is accomplished by the same methods described for anterior dislocation, except that gentle external rotation is used along with anterior pressure on the posterior humeral head. The shoulder can be immobilized in a shoulder spica cast or brace in 30 degrees of abduction and some external rotation for 3 to 4 weeks. Recurrent posterior dislocations can require posterior capsular reefing or osteotomy of the glenoid with bone block placement.

Fracture-Dislocations of the Shoulder

Dislocation of the shoulder can be associated with fracture of the proximal humerus. Reduction is with maneuvers similar to those for simple dislocations, but usually requires general anesthesia. If reduction cannot be obtained by closed means, open reduction and internal fixation is indicated.

Humerus Fractures of the Proximal Humerus

Fractures of the surgical neck (proximal metaphysis) of the humerus are common injuries in adults, while in children the anatomic neck (junction of the epiphysis and metaphysis) is involved, usually in a Salter-Harris type II configuration (Fig. [41-65](#)). Treatment is with closed reduction using longitudinal manual traction, followed by a sling or hanging cast application. A significant degree of angulation can be accepted because of the large range of motion of the shoulder joint. In instances of unacceptable angulation or displacement, percutaneous pinning or open reduction is indicated.

In adults fractures of the proximal humerus may be comminuted, in which case they are classified according to the number of displaced segments (Neer classification). The segments consist of the head, shaft, and greater and lesser tubercles, and displacement is defined as more than 45 degrees of angulation or 1 cm of displacement. Thus fractures are three- or four-part if one or both tubercles are fractured and displaced. Displaced greater tubercle fractures require reduction to maintain function and are repaired using Hawkins's tension band wiring technique. In four-part fractures there is a high incidence of avascular necrosis of the humeral head, and generally replacement of the humeral head with a Neer-type endoprosthesis is advocated, with reattachment of the tubercles to the prosthesis by tension band wiring or heavy nonabsorbable sutures (Fig. [41-66](#)).

In all proximal humerus fractures, early motion is essential to prevent stiffness. Physical therapy is an important part of the treatment, and gentle passive pendulum exercises usually should be started within 1 or 2 weeks.

Fractures of the Humeral Shaft

Humeral shaft fractures may be transverse, comminuted, or spiral, depending on the mechanism of injury (direct versus torsional force). Swelling, pain, crepitus, and instability may be present, and occasionally the fracture is open, particularly with high-energy trauma such as a motor vehicle accident. Careful evaluation and documentation of the neurologic status of the limb is critical, as associated nerve injuries, especially to the radial nerve, are common. Radial nerve injury most commonly occurs with oblique fractures of the distal to middle thirds of the humerus.

Treatment

Generally humeral fractures are reduced by gentle longitudinal traction with adequate sedation or local anesthesia and the patient placed in a coaptation splint (from the axilla medially, around the elbow, and up over the shoulder laterally), although a long arm splint or plastic humeral fracture brace also may be used. The patient is then placed in a sling and swathe (Fig. [41-67](#)). Neurapraxia occurs in 95 percent of patients with radial nerve injury, but will resolve within a few weeks to 3 months. Therefore the presence of a radial nerve palsy is not an indication for surgical intervention. Loss of the nerve function after closed reduction is indicative of entrapment in the fracture site and warrants exploration. Late exploration is indicated if recovery does not occur within 3 months. If satisfactory reduction cannot be obtained or maintained, surgical treatment with compression plating or intramedullary rodding of the humerus is appropriate. This is also preferred in multiple-trauma patients, who require rapid mobilization to prevent pulmonary and other complications.

Elbow

The elbow functions as a hinge joint, and the proximal and distal radioulnar articulations are important for pronation and supination. Elbow injuries are notorious for causing loss of motion due to capsular fibrosis or ossification (myositis ossificans or, more accurately, heterotopic ossification). In addition to the routine anteroposterior and lateral views, oblique radiographs are helpful in delineating subtle injuries. Because the multiple ossification centers about the elbow appear at differing ages (capitellum—1 to 2 years; medial epicondyle—5 to 6 years; radial head—5 to 6 years; trochlea—9 to 10 years; lateral epicondyle—9 to 12 years), interpretation of radiographs in the context of acute trauma in children can be difficult. Comparison views of the opposite elbow often are helpful.

Fractures of the Radial Head and Neck

These injuries usually result from a fall on the outstretched hand. Patients have limited range of motion of the elbow, localized swelling and tenderness, particularly laterally, and pain with attempted pronation and supination. The greatest problem with these injuries is subsequent loss of motion. Aspiration of the joint and instillation of a small amount of lidocaine will improve range of motion and relieve pain. If the fracture is not severely angulated or displaced, the patient can be treated with a sling and active motion begun within 2 weeks to minimize occurrence of flexion contracture. If the radial head is comminuted, excision may be indicated. Long-term follow-up of radial head excision demonstrates proximal

subluxation of the radius but minimal symptoms. In cases of associated elbow dislocation with instability, a silicone spacer can be inserted until healing occurs, but it should be removed later to prevent silicone synovitis and fragmentation.

In children a fall on the outstretched hand causes a fracture of the radial neck, as opposed to the radial head fracture that occurs in adults. Closed reduction with varus stress on the elbow and pressure directed laterally on the proximal radius is performed, and the elbow is splinted. Angulation of up to 45 degrees is well tolerated and will remodel, but if greater angulation is present, open reduction may be necessary.

Fractures of the Proximal Ulna

Olecranon fractures occur primarily in adults, either as a result of forced flexion against an actively contracting triceps, or more commonly, from a direct force to the olecranon. In children the usual lack of displacement allows conservative treatment with immobilization in a relatively extended position. In adults, however, the fractures usually are displaced and require open reduction and internal fixation. The most common approach involves the use of tension band wiring, which converts the pull of the triceps into compressive force at the fracture site (Fig. 41-68). More comminuted fractures require excision of fragments and advancement of the triceps aponeurosis to the remaining olecranon, or plating with fixation of the multiple fragments.

Supracondylar Fractures of the Humerus

Supracondylar fractures are the most common fractures about the elbow in children but are rare in adults. The fracture is caused by a fall on the extended elbow, with hyperextension and posterior angulation of the distal condylar fragment. Less than 5 percent of these fractures are of the flexion type. Supracondylar fractures are especially hazardous because of the potential for neurovascular injuries, compartment syndromes, and sequelae such as Volkmann's contracture (discussed earlier under Contracture) and malunion. The brachial artery can be injured or lacerated by the anterior projection of the proximal fragment, as can the median nerve.

Children with displaced supracondylar fractures usually present with marked swelling about the elbow, deformity, and ecchymosis. Circulatory integrity of the forearm and hand as well as neurologic function must be assessed. The distal fragment may be displaced posteromedially (most common) or posterolaterally, and damage to the median or radial nerve can occur. Vigilance regarding development of compartment syndrome is essential.

Treatment

Undisplaced fractures can be treated with a posterior splint with 90 degrees of elbow flexion. Further flexion can be hazardous because of swelling, and all patients should be observed closely after immobilization. Displaced fractures in the past were treated with manipulative reduction and immobilization in as much flexion as tolerated without circulatory embarrassment, or by overhead olecranon skeletal traction. Today, however, the method of fluoroscopically guided percutaneous pinning after closed reduction under general anesthesia is the most widely accepted method. This allows maintenance of more anatomic reduction without the need for excessive flexion of the elbow (Fig. 41-69). Postoperatively the patient is maintained in a posterior splint with not more than 90 degrees of elbow flexion. Circulatory compromise or evidence of compartment syndrome mandates immediate brachial artery exploration and forearm fasciotomies. Pins are removed at 3 to 4 weeks, at which time active range- of-motion exercises are started.

Fractures of the Lateral Epicondyle in Children

These injuries result from elbow hyperextension with associated valgus stress. Fractures of the lateral condyle can be difficult to diagnose radiographically, and comparison views are useful (Fig. 41-70). Three stages of displacement have been described: stage I—undisplaced; stage II—minimally displaced; less than 2 mm; and stage III—displaced and rotated. These injuries are not only intraarticular but also cross the growth plate, and therefore any displacement warrants percutaneous pinning, or open reduction if needed. Complications include nonunion and growth disturbance.

Fractures of the Medial Epicondyle in Children

Valgus injury to the elbow can avulse the medial epicondyle. The treatment of this injury is controversial, and there are proponents of accurate reduction and pinning as well as of conservative treatment. However, if the fragment is entrapped in the joint or if ulnar nerve symptoms are present, open reduction and pinning is needed.

Comminuted Fractures of the Distal Humerus

Comminuted fractures of the distal humerus usually result from a direct blow to the flexed elbow. Discontinuity of the articular surface necessitates open reduction and internal fixation with both interfragmentary screws and medial and lateral buttress plates through a posterior approach. Best visualization of the articular surface is achieved by chevron osteotomy of the olecranon, reflecting the triceps proximally with this fragment, and fixing the olecranon later using an intramedullary screw. The major long- term complication is loss of elbow motion.

Dislocations of the Elbow

These injuries occur more commonly in children than adults and usually result from a fall on the outstretched forearm with posterior translocation of the radius and ulna. The anterior capsule is torn, as is the medial collateral ligament. These injuries usually are stable after reduction when associated fractures are not present. Immobilization is with a posterior splint and 90 degrees of elbow flexion, with early institution of motion. Careful neurologic assessment is necessary, since about 10 percent of patients have associated ulnar or median nerve injury.

Fracture-Dislocations of the Elbow

Posterior elbow dislocations in conjunction with fractures of the ulnar coronoid process, radial head, or lateral condyle occur in adults and are much less stable injuries than simple posterior dislocations. Open reduction may be necessary, and if the radial head is fractured, fixation or temporary prosthetic replacement often is necessary to ensure elbow stability in the healing period. Reattachment of an avulsed coronoid fragment also may improve stability by reconstituting the brachialis insertion.

Monteggia Fracture and Galeazzi Fracture-Dislocation

A Monteggia fracture is a fracture of the proximal ulna associated with dislocation of the radial head (Fig. 41-71). A Monteggia-equivalent fracture is fracture of the radial neck in addition and is seen more often in adults. The radial head dislocates posteriorly in 15 percent of patients and anteriorly in 85 percent. This injury illustrates the importance of the principle of obtaining radiographs of the joints above and below the level of a fracture. A single bone fracture in the forearm should always alert one to the possibility of injury to the radioulnar articulation distally or proximally. The Galeazzi fracture-dislocation is an analogous injury with fracture of the radial shaft and dislocation of the distal radioulnar joint.

Closed reduction may be possible, maintaining the elbow flexed and supinated. If the ulna fracture cannot be anatomically reduced, the injury will be unstable and the radial head likely to redislocate. Internal fixation of the ulna, usually by compression plating, may be necessary, but many of these injuries, particularly in children, can be treated closed.

The Galeazzi fracture-dislocation involves a distal-third radius fracture in combination with dorsal dislocation of the distal radioulnar joint. The resultant instability pattern is similar to that of the Monteggia lesion. In adults, plating of the radius and immobilization of the forearm in supination is necessary. This injury is uncommon in children, but closed reduction and immobilization in a long arm cast in full supination usually is successful.

Radial Head Subluxation in Children

Also known as "nursemaid's elbow," this injury results from pulling a child's elbow into sudden extension, and is seen in children under 5 years of age. The elbow can be flexed from 30 to 120 degrees without pain, and the radiographs are normal. The lesion is a tear of the attachment of the annular ligament, allowing escape of the radial head and resulting in entrapment of the annular ligament in the radiohumeral joint. It can be reduced by forceful supination with the elbow flexed 90 degrees, followed by sling immobilization for a week.

Forearm Fractures

Fractures of the forearm are common in children, but also occur in adults in motor vehicle accidents, falls, and contact sports. Deformity, swelling, and instability may be present. Radiographs must include the elbow and wrist, and careful neurovascular examination of the extremity is critical.

The posterolateral bow of the radius is critical to the maintenance of pronation and supination. In addition, rotational malalignment can occur from the pull of the supinator and biceps on the proximal radius. Anatomic reduction of the fractures with restoration of proper rotation and alignment is essential to a good functional outcome.

Treatment

Closed reduction can be undertaken under axillary nerve block, intravenous sedation, or general anesthesia. Longitudinal traction is applied with the elbow flexed to 90 degrees. After manipulative reduction, the fracture is immobilized in the position of maximum stability. Generally supination for proximal-third fractures, neutral rotation for middle-third fractures, and pronation for distal-third fractures are most appropriate. A long arm cast is applied, with care taken to mold volarly and dorsally along the interosseous membrane to help maintain position. In adults, results of closed treatment are usually inferior to open reduction and internal fixation, and nonunion and malunion rates are significantly higher. Therefore, displaced fractures in adults are ordinarily treated by operative reduction and rigid internal fixation using dynamic compression plates (Fig. 41-72). These plates have eccentrically placed holes that allow significant compression at the fracture site as the screws are seated in the plate. Postoperatively, a plaster long arm splint is applied, but if stable fixation is attained, early motion to restore supination and pronation is a major advantage of internal fixation. In cases of marked comminution or soft-tissue injury, early or primary bone grafting of the fractures is warranted.

Controversy remains over whether the hardware should be removed after fracture healing is achieved (1 to 2 years). In young adults, the stress shielding effect of hardware gradually causes weakening of the cortical bone beneath the plate. In time this results in a stress riser at the edge of the plate, which can lead to a subsequent fracture with less than normal force. On the other hand, incidences of fracture through screw holes (which are also stress risers for a number of months after hardware removal) have been as high as 20 percent after hardware removal. There is no consensus on plate removal at present. Often the plates cause tendinitis or other symptoms, and the patient requests removal despite the risk of refracture. A reasonable approach involves (1) using only the smaller dynamic compression plates in the forearm, which require 3.5-mm rather than 4.5- mm screws; (2) plate removal in symptomatic patients; (3) removal of hardware in young adults; and (4) prolonged protection after plate removal with an orthosis or removable fiberglass cast as well as activity restrictions for 6 months to 1 year.

Forearm Shaft Fractures in Children

Open reduction of these injuries is rarely indicated except when the injury occurs in a teenager without much growth remodeling potential remaining, in which case the fracture is handled as described above for adults.

Greenstick fractures (Fig. 41-73), in which one cortex is fractured but the opposing cortex has undergone plastic deformation and angulation, should be manipulated to complete

the fracture; otherwise reduction is hard to achieve. The reduction maneuvers are in other respects similar to those for the adult. Postreduction films are important, as is careful follow-up to ascertain that the reduction is maintained. Greater degrees of angulation can be accepted in fractures in younger children or fractures close to a growth plate, where the potential for remodeling is greater.

Distal Radius Colles' Fracture

Fracture of the distal radius with dorsal angulation and a “dinner fork” deformity of the wrist was described by Abraham Colles in 1814 (Fig. 41-74). This very common injury occurs as a result of a fall on the outstretched hand. The fracture is most common in patients over 50, often postmenopausal women with some degree of osteopenia. Patients have deformity, swelling, ecchymosis, and pain in the wrist area. Neurologic injury involving the median nerve occurs, and careful neurologic examination of the hand both before and after reduction is important. Patients occasionally develop an acute carpal tunnel syndrome after reduction, particularly if the wrist is immobilized in a position of extreme flexion. Radiographic evaluation should extend to the elbow to rule out other injuries.

Treatment

Goals of treatment include the minimizing of cosmetic deformity as well as restoration of good wrist and hand function. The usual approach is conservative treatment with closed reduction and plaster immobilization. Generally after regional anesthesia or sedation longitudinal traction is applied using Chinese finger traps with a counterweight to the humerus and the elbow at 90 degrees of flexion. Manipulation then consists of exaggeration of the deformity followed by volar flexion and ulnar deviation along with gentle pressure on the dorsum of the fragment to push it anteriorly. The wrist is then immobilized in a long arm plaster cast in ulnar deviation and flexion—avoiding full flexion because of the potential for median nerve compression. Elevation is continued for a few days to minimize swelling, and early active motion of all digits is encouraged; later shoulder range-of-motion exercises are begun to minimize any stiffness that may occur. In the event of the development of carpal tunnel syndrome, the cast is removed and the wrist placed in the neutral position. If the symptoms fail to resolve, surgical carpal tunnel release is indicated. There is no consensus on whether the wrist should be immobilized in pronation or supination, and hence the neutral position is most commonly used. At 3 to 4 weeks the cast is changed to a short arm cast, which is removed at 6 weeks from the initial injury.

If satisfactory reduction cannot be attained or is lost during the first week of treatment, other methods are used. Manipulation under anesthesia with percutaneous pinning has been used successfully in this situation, preventing recurrent loss of position.

In cases with articular surface disruption, open reduction of the fracture fragments with internal fixation is occasionally indicated to restore a congruous surface of the distal radius. More commonly, however, external fixation is used, allowing ligamentotaxis to achieve reduction of the fracture fragments. One disadvantage of external fixation of distal radius fractures is that marked stiffness of the wrist usually results. Figure 41-75 demonstrates the application of external fixation.

Smith Fracture

A fall on the dorsum of the wrist may cause a reversed Colles' fracture, which is given the eponym Smith fracture. These fractures are characterized by volar angulation and displacement of the distal fragment. Reduction is with longitudinal traction and cast immobilization, but position may be difficult to maintain, and percutaneous pin fixation or open reduction and internal fixation with a volar buttress plate occasionally is necessary.

Wrist

A fall on the outstretched wrist can result in ligamentous injuries, with resultant pain, limitation of motion, and swelling. Ligamentous disruption, particularly when it involves the scapholunate articulation, can result in intercarpal instability patterns and must be carefully evaluated. In addition, these injuries must be differentiated from bony disruptions such as scaphoid fracture, lunate dislocation, or transscaphoid perilunate dislocation.

Fractures of the Scaphoid (Carpal Navicular)

Young adults who fall on the outstretched hand frequently fracture the scaphoid (Fig. 41-76). Signs include tenderness in the “anatomic snuff box,” which is formed by the extensor pollicis longus and abductor pollicis longus tendons just over the tuberosity of the scaphoid. In addition to standard anteroposterior and lateral radiographs, an oblique 17-degree view (navicular view) is necessary to adequately visualize the scaphoid. Even with negative films, a nondisplaced fracture may be present when there is reproducible tenderness over the scaphoid. Because the blood supply to the scaphoid enters distally, fractures devascularize the proximal fragment to some extent; consequently healing of this fracture is slow and the nonunion incidence significant. Therefore, in the presence of clinical signs, the patient is placed in a thumb spica cast presumptively and repeat films are obtained out of plaster 3 weeks later (see Fig. 41-76). If radiographs are negative and symptoms resolved, mobilization of the wrist can begin.

Treatment

If the fracture is minimally displaced, thumb spica cast immobilization is initiated; authors differ as to whether a long or short arm cast is necessary, but most prefer a long arm cast at least initially. Average healing time is 12 to 16 weeks, and in some cases can take longer than 6 months.

In cases with displacement or angulation of the scaphoid, open reduction and internal fixation with pins or screws should be performed. In cases of late diagnosis (up to 6 months), union can still occur with prolonged immobilization. When nonunion is not associated with arthritic change in the wrist, autogenous bone grafting is the recommended approach and has a union rate of 92 percent. Electrical stimulation also has been reported to promote union. If degenerative changes are present, radial styloidectomy can provide pain relief and improved function.

Lunate and Perilunate Dislocations

These injuries are uncommon but can result from traumatic hyperextension of the wrist. Patients complain of pain, swelling, and limitation of motion, and median nerve symptoms may be present. The lunate is displaced anteriorly, best visualized on lateral radiographs. On the anteroposterior view the normal quadrilateral shape of the lunate is more triangular and larger than on the opposite side.

Treatment

Closed reduction can be accomplished by hyperextending the wrist and applying pressure volarly over the dislocated lunate. Afterward the wrist should be immobilized in flexion. If unsuccessful, open reduction through a volar approach is carried out with ligamentous repair and pin fixation. A similar approach is taken with the treatment of transscaphoid perilunate dislocations, in which the lunate remains attached to the distal radius via the radiolunate ligament and the carpus dislocates. Closed reduction of these injuries can be difficult, and open reduction often is necessary to achieve anatomic relationships of the carpal bones. Residual scapholunate dissociation resulting in dorsal or volar intercalated segment instability (depending on the direction of relative flexion/extension of the lunate and scaphoid) can cause late progressive carpal collapse and degenerative arthritis.

Hand Metacarpal Fractures

Metacarpal fractures can occur from direct trauma, crush injuries, or striking the hand against an object. Fractures of the neck of the fourth and fifth metacarpals commonly result from fistfights. Local swelling masks the degree of volar angulation of the metacarpal head. The angular deformity must be assessed on the lateral radiograph, and rotational deformity must be judged clinically by flexing the digit, which should point toward the scaphoid.

Treatment

Metacarpal neck fractures usually are treated by closed manipulation under local anesthesia. For the fourth and fifth metacarpals, an ulnar gutter splint is applied with as much *metacarpophalangeal (MCP)* joint flexion as possible and the *interphalangeal (IP)* joints in extension. An important point in immobilization of any injury to the hand is that the collateral ligaments of the MCP joints are lengthened in flexion, and therefore immobilization in extension leads to contracture of the collaterals and difficulty in regaining flexion. Up to 40 degrees of residual angulation can be accepted in the fourth and fifth metacarpals because of the increased mobility of these rays in the anteroposterior direction. If angulation is excessive or involves the more radial rays, percutaneous pinning is necessary to maintain acceptable reduction.

Metacarpal shaft fractures usually can be treated conservatively, but if excessive angulation, displacement, or rotatory malalignment is present, open reduction and Kirschner-wire fixation or plating are occasionally necessary. Comminuted fractures of the metacarpal heads are difficult to treat and should be splinted in a functional position and early motion initiated at 2 to 3 weeks.

Fracture-dislocations of the *carpometacarpal (CMC)* joint are rare and usually are treated with closed reduction and splinting.

Bennett Fracture

This is an intraarticular fracture of the thumb metacarpal caused by an axial force against the partially flexed bone. The fracture line typically leaves a small volar fragment with the trapezium, while the rest of the metacarpal is pulled dorsally and radially by the abductor pollicis longus muscle. This leads to instability and subluxation or dislocation of the CMC joint. Variants include a rare comminuted type (Rolando fracture) and extraarticular fractures of the proximal metaphysis, which are the most common type and are more stable.

Minimally displaced fractures can be treated by closed reduction and thumb spica casting but must be followed closely because of the propensity for displacement. Because of the inherent instability of the intraarticular fractures, the most common method of treatment involves closed reduction with percutaneous pinning of the metacarpal to the trapezium (Fig. 41-77). If significant incongruity of the joint surface persists despite attempted closed reduction, open reduction and pin or screw fixation are recommended. Rolando fractures are difficult to treat because they can have significant comminution. If the articular surface can be restored by pinning after closed or open reduction, this is the method of choice, but because of the small sizes of the fragments it often is not feasible. In these cases brief immobilization followed by early motion is suggested.

Ulnar Collateral Ligament Injury of the Thumb (Gamekeeper's Thumb)

Sudden abduction stress to the thumb can lead to rupture of the ulnar collateral ligament of the MCP joint. The injury was first described in British gamekeepers, who were noted to have a chronic laxity of the ligament caused by chronic repetitive abduction stress when killing game. Falls and ski pole injuries are some of the more common causes today. Stener described a phenomenon that makes conservative treatment of this injury difficult, namely, interposition of the adductor insertion between the ends of the ligament. This occurs in approximately 50 percent of these injuries and precludes healing of the ligament, leading to chronic instability of the joint. The patient usually presents with a painful, swollen thumb MCP joint. The ulnar aspect of the joint is tender, and radiographs usually are normal, although sometimes an avulsion fragment at the base of the proximal phalanx is visible. If standard radiographs are normal, stress radiographs are recommended to demonstrate the instability. This can be appreciated on physical examination as well.

Treatment

For grades I and II ligament sprains, immobilization for 3 to 6 weeks in a thumb spica cast with the thumb in slight adduction is advised. Because of the prevalence of the Stener lesion, for grade III (complete) tears, operative repair is indicated. For ruptures diagnosed later than 2 to 3 weeks after injury, primary repairs are not feasible, and reconstructive

surgery is necessary, with tendon grafts or capsular reefing and adductor tendon advancement.

Phalangeal Fractures

Nondisplaced phalangeal shaft fractures are treated by splinting in a position of function (MCP joints flexed approximately 70 degrees and IP joints in only slight flexion). Generally, malleable aluminum splints with foam padding held in place with adhesive tape can be used. For displaced fractures, manipulative reduction may be necessary before splinting.

Small avulsion fractures of the IP joints are treated symptomatically by splinting for 2 weeks followed by range-of-motion exercises. Interphalangeal dislocations are usually dorsal but also may be volar or rotatory. Dorsal dislocations are accompanied by tearing of the volar plate, often with an associated chip avulsion. Rotatory dislocations resulting from twisting injuries are unusual, and they can be irreducible because of soft-tissue interpositions such as buttonhole capture of one condyle through the extensor aponeurosis. Occasionally dorsal IP dislocations are associated with intraarticular fracture of the base of the phalanx.

Interphalangeal Dislocations

Treatment of IP dislocations is with closed reduction followed by buddy taping for 3 to 6 weeks, or by extension block splinting (a dorsal splint that limits full extension but allows active flexion) to allow healing of the volar plate and prevent late hyperextension deformity. Fracture-dislocations involving a portion of the articular surface can be treated by closed reduction with subsequent extension block splinting. If closed reduction is unsatisfactory, open reduction with fixation of the articular fragments can be performed. In cases in which the fragments are small or comminution is present, a volar plate (Eaton) arthroplasty is recommended, with excision of the fragments and advancement of the volar plate into the defect with a pullout wire.

Distal Phalanx Fractures

Crush injuries to the distal phalanx are the most common fractures in the hand. Often there are associated injuries of the nail and nailbed. The most frequent types are comminuted tuft fractures of the distal portion of the phalanx, and transverse fractures of the more proximal base of the phalanx, which may be angulated. The nailbed injuries, which are often neglected, may actually be more of a problem than the fractures themselves.

Initial care is aimed at cleansing any wounds to prevent infection. Drainage of any subungual hematoma by piercing the nail with a hot paper clip or a battery-powered hot wire cautery affords significant pain relief. Late nail deformity is an unavoidable complication of crush injuries, and the patient should be advised of this. Careful repair of any nailbed lacerations with fine absorbable sutures helps considerably in diminishing these problems. If the nail has been avulsed, it is replaced beneath the cuticle after nailbed repair to act as a splint to protect the nailbed and decrease local tenderness. The nail eventually will be pushed off by growth of the new nail. The fracture is then splinted with a protective metal splint for 3 to 4 weeks. Angulated fractures of the distal phalanx should be reduced and, if unstable, pinned with a Kirschner wire before splinting.

Mallet Finger

Forcible flexion of the distal phalanx against active contraction of the extensor mechanism avulses the distal insertion of the extensor tendon from the dorsum of the phalanx, usually with a small fragment of bone (Fig. 41-78). The patient is unable to fully extend the distal phalanx. The injury is generally treated conservatively, with a volar or dorsal splint across the *distal interphalangeal (DIP)* joint holding it in hyperextension. Most authors believe that immobilization of the *proximal interphalangeal (PIP)* joint is not necessary. The splinting must be maintained for 6 weeks, and occasionally longer if active extension has not been regained. Some extensor lag may remain after treatment but is usually not problematic. Even if seen late (2 to 3 months after injury), most mallet finger injuries will improve with prolonged (8 weeks) splinting.

If the avulsed fragment of bone comprises more than one-third of the articular surface, operation is indicated, although for small amounts of displacement, conservative treatment as described above is appropriate. With volar subluxation of the distal phalanx, however, repair with a pull-out wire or Kirschner-wire fixation is recommended.

Markedly comminuted intraarticular IP joint fractures are best treated by splinting in a position of function and early arthrodesis if symptoms warrant. Open fractures are treated with irrigation, debridement, and antibiotics, following the general principles of treatment of other open fractures described previously.

Lower-Extremity Injuries

Femur

Before the advent of internal fixation for fractures of the proximal femur, patients were treated with bed rest and traction, and mortality rates were high. Although mortality and morbidity have significantly improved with modern methods of management, hip fractures represent a major challenge to the health care system, with over 275,000 such injuries annually in the United States at an estimated cost of over \$3 billion. These fractures generally occur in the 60- to 70-year-old population, and the incidence continues to increase as the number of elderly increases. Femoral neck and intertrochanteric fractures occur with approximately equal frequency and similar epidemiology. The incidence increases with increasing age, with over one-third of women and one-sixth of men over the age of 90 years having sustained a hip fracture. Osteoporosis is a contributing factor to hip fractures, causing decreased mechanical strength of the proximal femur. Mild nutritional osteomalacia also can be a factor but has been found in only about 2 percent of patients with femoral neck fractures. The mortality of hip fractures in the elderly has been reported to be as high as 20 to 50 percent within the first year.

These fractures occur much less commonly in young adults or children, in whom they usually result from high-energy trauma such as motor vehicle accidents. The most serious complications of these fractures are osteonecrosis of the femoral head and nonunion.

Femoral Neck Fractures

Femoral neck fractures are most commonly produced by a fall. The patient usually complains of pain in the groin or thigh and is unable to bear weight on the extremity. The leg usually appears shortened and externally rotated, and any attempt at motion causes severe pain. Trochanteric ecchymosis may be evident. Diagnosis is confirmed by anteroposterior and lateral radiographs of the hip, and careful physical examination is necessary to rule out other injuries to the ribs, upper extremities, or knee. Discontinuity of the trabeculae is seen in both views radiographically, and usually the femoral head is angulated posteriorly on the lateral view.

The most widely used classification of femoral neck fractures is the Garden classification (Fig. 41-79):

Type I—Incomplete fracture (usually impacted in valgus)

Type II—Complete fracture without displacement

Type III—Complete fracture with partial displacement

Type IV—Complete displacement of fracture

Higher incidences of osteonecrosis and nonunion occur with the displaced type III and type IV fractures. The undisplaced type I and type II fractures can be difficult to diagnose, and initial radiographs may be negative. Some patients with undisplaced fractures have surprisingly little pain and are even able to bear weight. Usually internal rotation of the hip will cause pain, and ecchymosis over the trochanter should raise suspicion. A bone scan or MRI should be performed in any patient with negative radiographs and unexplained hip pain after a fall and will readily demonstrate the fracture in essentially all cases, although the age of the fracture may remain difficult to ascertain.

Treatment

The femoral head has a more precarious blood supply than many bones, given the large surface covered with cartilage, which has no blood vessels. The majority of the blood supply comes in through the neck and through subcapital epiphyseal vessels that run in the inferior capsule, supplied from the medial and lateral circumflex vessels. A small contribution in adults can come from the artery of the ligamentum teres. Disruption of the blood supply occurs with fracture of the neck, accounting for the high incidence of osteonecrosis and nonunion. Mechanical factors such as the high loads across the femoral neck and the fact that little if any external callus formation occurs in these intraarticular fractures further contribute to nonunion. Aspiration of capsular hematoma has been advocated by some to decrease tamponade of capsular vessels, but the effect on development of osteonecrosis remains uncertain.

In elderly patients, surgical treatment allows early mobilization, significantly decreasing morbidity and mortality. Fixation within 24 h has been advocated with hip fractures to minimize osteonecrosis, but more recent data suggest that the most important consideration is achieving medical stability of the patient before surgery. It is generally agreed, however, that fracture fixation and mobilization of the patient as quickly as possible are beneficial. Internal fixation is indicated in nondisplaced fractures and in younger, active individuals even with significant displacement. Displaced fractures in the elderly are treated most commonly by endoprosthetic replacement of the femoral head because of the high incidence of osteonecrosis or nonunion. With type I and type II fractures, nonunion occurs in less than 5 percent of cases, and osteonecrosis in less than 8 percent, while 10 to 30 percent of patients with type III and type IV fractures develop nonunion and 15 to 33 percent osteonecrosis.

Nondisplaced Femoral Neck Fractures

Type I and type II fractures can be treated conservatively, particularly with fractures impacted in valgus. Because of the frequent inability of the elderly patient to fully cooperate with protected weight bearing and the continued potential for displacement, which would significantly worsen prognosis, internal fixation generally is recommended. In demented, nonambulatory patients with other medical problems, conservative treatment often is indicated. Even medically ill patients, however, usually can tolerate internal fixation, which can be done percutaneously with multiple pins under local anesthesia if necessary. Comparison between sliding compression screw with side plate fixation and multiple parallel threaded-pin fixation has indicated superior union rates with multiple pins or screws, with three pins providing adequate stability (Fig. 41-80). In young adults every effort should be made to preserve the femoral head, even in displaced fractures. Accurate reduction is important in minimizing the incidence of nonunion, and generally varus position or more than 15 degrees of angulation on the lateral should not be accepted. If closed reduction is unsuccessful, open reduction through an anterolateral or posterior approach should be undertaken. Muscle pedicle bone grafting has been advocated to increase union rate and decrease osteonecrosis but remains controversial.

Postoperative Management

Because of the high incidence of thromboembolic disease in elderly patients after hip fracture, many advocate some form of prophylaxis perioperatively. Approaches to this problem include low-dose aspirin, warfarin, antiembolism stockings, and pneumatic compression stockings. Mobilization of the patient is rapid, with sitting in a chair within 24 h of surgery a goal. Weight bearing is protected until the fracture is healed, usually within 4 to 6 months. Follow-up is needed for longer periods, as osteonecrosis may not become radiographically evident for up to 2 years. Viability of the femoral head before or after internal fixation can be assessed by bone scan or MRI. MRI changes of osteonecrosis are present within a few days of blood supply disruption, but after surgery the metal artifact introduced by the pins severely limits usefulness of MRI. In patients who develop

osteonecrosis, subsequent collapse necessitates prosthetic replacement. In patients with nonunion and a viable femoral head by conventional or tomographic (single photon emission computed tomography, or SPECT) bone scan, either prosthetic replacement or valgus osteotomy and bone grafting can be undertaken, depending on the age and activity level of the patient.

Endoprosthetic Hip Replacement

In elderly patients the relatively high incidence of nonunion and osteonecrosis with displaced fractures argues for endoprosthetic replacement of the femoral head. If the hip is arthritic, total hip replacement arthroplasty is indicated. Usually cemented arthroplasties are recommended, although in younger or more active patients with complications after attempted femoral head salvage, uncemented porous-coated prosthetic replacement may be considered. Bipolar endoprostheses have a movable articulating surface with a polyethylene liner capturing a smaller spherical ball that is continuous with the intramedullary stem of the device (Fig. 41-81). The motion at both articulating surfaces decreases wear on the acetabulum, and the problem of late acetabular migration (protrusion) is considerably less with bipolar than with earlier fixed-head endoprosthetic designs.

Additional indications for endoprosthetic replacement as opposed to internal fixation include neurologic conditions (parkinsonism, stroke, dementia), metabolic bone disease (renal osteodystrophy, hyperparathyroidism, Paget's disease), and pathologic fractures caused by metastatic carcinoma. With cemented arthroplasty, weight bearing is allowed as tolerated. In patients with preexisting degenerative changes in the joint, total joint replacement rather than bipolar arthroplasty is indicated.

Femoral Neck Fractures in Children

Femoral neck fractures usually occur in children only after high-energy trauma, such as pedestrian/motor vehicle accidents or falls from a great height. Fractures have been classified as transepiphyseal (with or without dislocation), transcervical, cervicotrochanteric, and intertrochanteric. Transcervical and cervicotrochanteric fractures account for the majority. Pediatric hip fractures are associated with a high incidence of complications, including physeal closure, nonunion, and osteonecrosis of both the femoral head and neck. Displaced fractures are treated by gentle closed reduction and pinning, and nondisplaced fractures by spica cast immobilization with close radiographic follow-up to monitor possible displacement. Intertrochanteric fractures usually can be treated conservatively with a spica cast.

Intertrochanteric and Subtrochanteric Fractures

These fractures occur in the elderly after falls and after significant trauma in younger individuals and generally are mechanically less stable than femoral neck fractures. The instability of the fractures usually results from comminution of the posteromedial cortex in the area of the calcar and lesser trochanter, with a tendency of the fracture to collapse into varus. Patients present with an inability to bear weight, shortening and external rotation of the lower part of the extremity, and often swelling or ecchymosis about the hip.

The considerations for surgical treatment of intertrochanteric and subtrochanteric femur fractures are similar to those described for femoral neck fractures, with internal fixation allowing early mobilization and decreasing significantly morbidity and mortality in the elderly. In younger individuals who can tolerate bed rest or in patients who have medical contraindications to surgical intervention, Russell's or skeletal traction and balanced suspension may be used, but internal fixation is the preferred method of treatment. Union of these fractures may require 4 months or more, however. Surgical intervention consists of open reduction and internal fixation using a sliding compression device with side plate (Fig. 41-82) or an intramedullary fixation device (Zickel nail, gamma nail, or reconstruction nail).

Intertrochanteric fractures have a lower incidence of nonunion than femoral neck fractures, and osteonecrosis of the femoral head is uncommon since its blood supply usually is not disrupted. Treatment of unstable fractures with medial comminution has been controversial, with anatomic reduction, medial displacement, or valgus osteotomies advocated to improve stability. Because a sliding hip screw allows controlled collapse at the fracture site, anatomically reduced fractures often spontaneously impact and medially displace to a more stable configuration. Fixation of a lesser trochanteric fragment, when possible, will also enhance stability. Intramedullary devices have been associated with a high rate of complications in intertrochanteric fractures but are preferred for subtrochanteric fractures. Subtrochanteric fractures have a significant incidence of delayed and nonunion, however, and in the presence of medial comminution primary bone grafting may be indicated.

Postoperative considerations include attention to rapid mobilization, pulmonary hygiene, and thromboembolic prophylaxis, as for femoral neck fractures. Generally, protected weight bearing is indicated until radiographic evidence of healing is present and will depend also on fracture pattern and fixation stability.

Femoral Shaft Fractures

Femoral fractures can occur at any age and generally result from violent trauma. Multiple-trauma patients require evaluation of any associated injuries of the head, abdomen, and chest. Patients exhibit instability of the extremity, pain with motion, rotational deformity, and shortening of the affected extremity. Neurovascular examination is essential, since there may be injury to the sciatic or femoral nerve or the femoral artery. Associated femoral fracture or knee ligament injuries occur in about 5 percent of patients. In the presence of signs of distal ischemia, arteriography is indicated, followed by immediate vascular exploration for repair or reconstruction and stabilization by internal fixation.

Treatment

Emergency treatment includes stabilization with a splint or traction splint. Open femoral fractures generally result from significant violence and should be handled as described earlier for general treatment of open fractures.

In children femoral shaft fractures are almost always treated conservatively, with modified Bryant's (in infants), Russell's (for ages 2 to 10 years), or distal femoral skeletal traction (for age 10 years and older). If fractures in young children are minimally displaced or shortened less than 1 to 2 cm, immediate spica cast treatment is appropriate. In children 1 cm of overriding is acceptable, since epiphyseal stimulation from the fracture will result in relative overgrowth of the extremity by about that amount. Angular or rotational deformities should be corrected as completely as possible, however (Fig. 41-83). In children treated initially in traction, callus formation will occur within 2 weeks, and the fracture will become nontender. The patient is then placed in a one and one-half hip spica cast for 2 to 4 months, until radiographic evidence of union is present. In children over 12 years of age with multiple trauma, internal fixation of the fracture to facilitate mobilization is often indicated.

The use of two 3- to 4-mm diameter titanium intramedullary nails inserted across the fracture site through the distal femoral metaphysis has been introduced by Rang for use in children as young as 5 years old to allow early mobilization and shorter hospitalization and to obviate the inconvenience of prolonged spica cast care (Fig. 41-84). One disadvantage of the method is the need for a second surgical procedure for nail removal, but the technique may prove to be a convenient and cost-effective method of treatment.

In adults with medical contraindications to surgery, traction until callus has formed followed by placement in a cast brace or hip spica cast may be appropriate. Conservative treatment of femoral fractures in adults results in a high rate of malunion and knee stiffness, and so the standard method of treatment of most adult femoral fractures is now intramedullary rod fixation. The advent of interlocking nails, which allow multiple screws to be placed through the femur and nail to give rotational and length control proximally and distally, has greatly extended the indications for surgery and is now the method of choice for treatment of comminuted femoral fractures (Fig. 41-85). This form of internal fixation is usually done "closed" under fluoroscopic image intensification, which avoids stripping of the periosteum and promotes more rapid healing. Whether treatment is surgical or conservative, skeletal traction and placement of a Steinmann pin through the distal femur or proximal tibia are needed. The distal femur should be used if knee injury is suspected, but distal locking is technically easier if the traction pin is in the tibia.

Intramedullary nailing facilitates mobilization of the multiple-trauma patient and is recommended within 24 h of injury. Fat embolism and other complications are decreased by early internal fixation. Infection rates with intramedullary fixation are 0.5 to 1 percent, and even with open fractures are in the range of 0 to 4 percent. Infections are treated by rod removal, overreaming to remove infected bone and granulation tissue, and repeat fixation with a larger rod in conjunction with appropriate antibiotic therapy. Union can be predictably expected in internally fixed femoral fractures in 16 to 20 weeks, and delayed union rates are approximately 1 to 2 percent overall. In the presence of nonunion, fatigue failure will eventually result, and the rod should be removed and replaced with a larger rod, which will result in healing in the majority of cases.

Supracondylar Femoral Fractures

Supracondylar fractures occur in all age groups as a result of motor vehicle accidents, falls, and other accidents. Patients usually present with pain, swelling, and deformity. Supracondylar fractures may be transverse, or T or Y shaped with an intraarticular split of the condyles. Knee stiffness is a major late problem with these fractures. Treatment typically is with skeletal traction through a tibial pin followed by cast-brace mobilization. Patients with multiple trauma should be treated by open reduction and internal fixation with a blade plate, dynamic compression screw and side plate, or intramedullary device. Extensive comminution frequently necessitates primary bone grafting, and emphasis should be on initial accurate fixation of the condyles to ensure a congruous articular surface.

Traumatic Distal Femoral Epiphyseal Separation

This injury results from hyperextension of the knee in the adolescent. Treatment is with closed reduction and plaster immobilization. Occasionally, percutaneous pin fixation to maintain an unstable reduction is indicated.

Knee

Injuries to the ligaments and menisci of the knee joint are common in athletic activities, including contact sports and skiing. Diagnosis is strongly dependent on clinical examination, since radiographs usually are normal.

Medial Collateral Ligament Injury

The medial collateral ligament (MCL) consists of both superficial and deep layers, which extend from the medial femoral epicondyle to the medial proximal tibia, blending with the capsule and retinacular structures anteriorly and posteriorly. The anterior portion is tight in flexion, the posterior portion tight in extension, and the middle portion tight throughout the range of motion. With valgus stress to the MCL, the deep capsular portion ruptures first, followed by the tibial collateral (superficial portion) and, with extreme force, the anterior cruciate ligament (ACL).

Clinical findings vary, but with isolated MCL tears they are generally localized to the medial aspect of the knee. Associated hemarthrosis may be noted, and usually there is local swelling and tenderness. With grade I and grade II partial tears, pain with valgus stress with the knee in 20 to 30 degrees of flexion, and slight opening of the medial joint line can occur. Abduction with opening of the joint line of more than 5 mm with valgus stress usually is diagnostic of a grade III rupture of the MCL. Comparison must be made to the other knee, since baseline laxity varies among patients.

Integrity of the other ligaments of the knee also must be established by careful examination to rule out combined injuries. Varus stress testing with the knee in 20 to 30 degrees of flexion, as described above for valgus testing, is used to determine competence of the lateral collateral ligament (LCL) and associated lateral complex. Anteroposterior translation of the tibia on the femur is tested with the knee in 60 to 90 degrees of flexion (anterior drawer test). Rupture of the ACL allows anterior translation and can be evaluated manually or with a device called the KT-1000, which more accurately measures displacements.

The Lachman test is a sensitive maneuver for evaluating anteroposterior motion. The test is performed with the knee in 20 degrees of flexion, with the examiner grasping the distal thigh with one hand and the proximal tibia with the other while applying translatory force. Again, comparison to the uninjured side is essential. Laxity in the Lachman test indicates ACL disruption, as does the pivot shift maneuver, which indicates the presence of anterolateral rotatory instability. This test is conducted by extending the knee with valgus stress applied; a “clunk” in the last 15 degrees of extension indicates anterior subluxation of the lateral tibia. The reverse test elicits reduction by flexing the knee from full extension, with a “clunk” again indicating reduction of the anterior subluxation. The posterior cruciate ligament (PCL) is examined using the reverse drawer test—forcing the tibia posteriorly with the knee in 60 to 90 degrees of flexion. An alternative method is to have the patient actively extend the knee with the foot fixed and the knee flexed. Anterior translation of the tibia will be observed if the PCL is torn.

Treatment

Isolated MCL tears will heal with conservative management consisting of a hinged cast or brace to prevent valgus stress for 6 weeks, followed by weight bearing as tolerated unless the patient has excessive valgus knee alignment. With combined injuries, surgical repair is warranted.

Anterior Cruciate Ligament Rupture

Isolated tears of the ACL can result from hyperextension, internal rotation of the tibia on the femur, or a direct blow on a flexed knee. Fifty percent of patients with acute isolated ACL tears have an associated meniscal injury, and MCL or LCL injuries are also common. ACL tears are the most common cause of chronic knee disability among athletes. The chronic rotatory subluxation of an ACL-deficient knee can lead to episodes of giving way and recurrent injury, and in the long term to early degenerative arthritis.

Treatment

Patients with isolated ACL injuries who do not engage in high-demand activity (e.g., older patients and nonathletes) are treated with a period of immobilization to allow healing of secondary restraints followed by an active rehabilitation program stressing hamstring strengthening and graduated quadriceps exercises. If symptomatic instability does not develop, further intervention is unnecessary. In cases of combined instability, surgical repair of the collateral ligament along with primary repair or reconstruction of the ACL is indicated, especially in patients involved in high-demand activities.

Midsubstance ACL ruptures, which are the most common, are difficult to repair primarily and usually require reconstruction with a tendon graft, either from the semitendinosus or the middle third of the patellar tendon. Patellar tendon grafts are also used for late reconstructions and have the advantage that bony blocks removed with each end of the graft can be securely attached within tunnels in the femur and tibia with screw fixation. Synthetic materials (dacron, carbon fiber, etc.) and allografts also have been used, but the autologous patellar tendon remains the most often used substitute. Isometric placement of the graft is essential to prevent tension or laxity throughout the range of motion.

Arthroscopic reconstruction of the ACL using a patellar tendon graft or other material is now a routine approach and allows precise placement of the graft tunnels in the femur and tibia. A small anterior incision is used to harvest the patellar tendon graft without requiring full arthrotomy. This less invasive procedure allows earlier recovery from surgery and earlier mobilization. Meniscal pathology can be treated arthroscopically at the same procedure.

Lateral Collateral Ligament Injury

The LCL is disrupted by varus stress to the knee, usually in association with injury to the associated lateral complex (posterolateral capsule, arcuate complex, popliteus tendon). The peroneal nerve also may be injured with this mechanism. With complete injuries, surgical repair is generally recommended.

Posterior Cruciate Ligament Injury

The PCL can be injured by a direct blow to the anterior knee, as in hitting a dashboard, or in combination with the ACL in hyperextension injuries. The ratio of PCL to ACL injuries is about 1:10, and isolated injuries do not necessarily cause significant instability. Midsubstance repairs have been associated with poor results, but reattachment of a bony avulsed fragment is recommended when present. Chronic patellofemoral symptoms can result from PCL insufficiency, and if symptoms are significant or if PCL rupture is associated with an LCL disruption, surgical reconstruction, usually with a patellar tendon graft, is advised. The same applies to combined PCL/ACL injuries, in which chronic instability usually follows conservative treatment.

Combined Injuries

Combined multiple-ligament injuries not only increase the likelihood that surgical repair will be necessary but also raise the possibility of knee dislocation. Dislocation can be associated with vascular injury to the popliteal artery with intimal tear or intraluminal thrombus. Vascular injury may be evident immediately or may become evident several days after the injury. Compartment syndrome in the leg also is a serious concern. Arteriography is recommended in all cases of injury to both cruciate ligaments in combination with a collateral ligament injury, in known knee dislocation, and in cases with any suggestion of vascular compromise.

Chronic Ligamentous Instability

Persistent laxity can cause pain and episodic giving way or locking and prevents participation in sports and other activities. Minor degrees of instability are manageable with an exercise program and limitation of high-demand activities. With symptomatic instability, ligamentous reconstruction is indicated. ACL and PCL instabilities are reconstructed as described for acute injuries. Stress radiographs are helpful in defining lax structures causing the instability with MCL and LCL insufficiency, and capsular reefing or other reconstructive procedures are helpful. Rehabilitation and postoperative care are extremely important in optimizing the results of knee ligament surgery, because of the tendency for knee stiffness to develop. For knees with inadequate motion after postoperative rehabilitation, manipulation under anesthesia, or arthroscopic lysis of adhesions followed by manipulation and postoperative continuous passive motion can significantly improve motion. Continuous-passive-motion machines also are useful in improving postoperative knee motion following fracture fixation and other routine knee surgeries.

Meniscal Injuries

The medial and lateral menisci of the knee are composed of fibrocartilage, triangular in cross-section, and serve functions of load transmission, stability, shock absorption, and lubrication. The peripheral 10 to 30 percent of the menisci and the anterior and posterior horn attachments are well vascularized and allow healing of tears. The inner two-thirds of the menisci, however, are avascular, precluding healing. The tibia rotates laterally on the femur with knee extension, and medially with flexion. If this rotation is blocked or forcibly reversed, the meniscus can be injured. Examples include catching a cleated shoe with weight applied to the flexed knee, or twisting the knee while in a squatting position.

The medial meniscus is injured more frequently than the lateral, because of its more constrained by capsular attachments. Several patterns of meniscal injury are illustrated in Fig. 41-86.

Meniscal tears are common as sports injuries and can present with effusion, a history of catching or locking, and knee pain. Patients may have joint line tenderness, limited knee motion, and a catching or “click” associated with pain during passive extension of the knee in conjunction with manual rotation of the tibia and varus or valgus stress (MacMurray’s test). Clinical diagnosis sometimes is difficult. Arthrography of the knee is fairly reliable in the diagnosis of medial meniscal tears (incidence 95 percent), but substantially less accurate in diagnosis of lateral meniscal tears (incidence 70 percent), owing to the distortion caused by the popliteus tendon. MRI, however, provides extremely accurate diagnosis of internal derangements of the knee of all types, including meniscal tears, osteochondral injuries, and ACL or PCL tears. High signal seen within the menisci also can indicate myxoid degenerative change, which is overinterpreted as a tear.

Treatment

Diagnostic arthroscopy provides another means of diagnosing meniscal injury accurately, and surgical arthroscopy may be performed to repair or excise the torn fragment simultaneously. The menisci are important stabilizers of the knee, and retention of as much well-contoured periphery of the meniscus as possible is desirable. Complete meniscectomy is associated with late degenerative arthritic changes in the affected compartment. For the same reason, peripheral tears in the vascularized portion of the menisci, which will heal, are treated by repair or immobilization. If the joint is unstable, operative repair of the tear with nonabsorbable sutures arthroscopically or by arthrotomy is indicated. If a peripheral tear is of partial thickness or the joint is stable, arthroscopic debridement of the tear followed by immobilization often results in healing. In either case the function of the meniscus can be retained. Meniscal tears in association with instability should not be repaired unless the instability is addressed also, such as by ACL reconstruction for anterolateral rotatory instability.

When intraarticular pathology has been localized by MRI or diagnostic arthroscopy, surgical intervention often can be carried out using arthroscopic technique, and the majority of meniscal lesions, osteochondral lesions, loose bodies, synovial bands, debridement of fibrillated articular cartilage, and abrasion arthroplasty procedures are now done arthroscopically. The advantages of arthroscopic surgery are minimal morbidity, decreased knee stiffness and postoperative pain, and more rapid rehabilitation and return to function. Using a small diagnostic arthroscope in conjunction with a video camera allows documentation of the pathology and viewing of the procedure by others. Small portals are used to introduce the arthroscope and surgical instruments into the joint. Using the technique of triangulation, the surgeon can manipulate objects within the knee, excise torn menisci, and shave or debride cartilage.

Most surgeons use anterolateral and anteromedial portals for diagnosis and most routine procedures. The knee is distended with saline solution injected through a portal in the suprapatellar pouch. Motorized shavers and cutting tools allow rapid resection of synovial, meniscal, or cartilaginous tissue and abrasion of bone to promote fibrocartilage repair. The general technique of surgical knee arthroscopy is illustrated in Fig. 41- 87. This technique has more recently been applied to a number of other joints, including the shoulder, hip, ankle, and elbow for diagnosis, and to a limited extent, treatment. For severe meniscal injuries, meniscal transplants (allografts) have been used, but this procedure is experimental and awaits further evaluation of outcome.

Patellar Fractures

Fractures of the patella usually occur as a result of direct trauma to the anterior knee but also can occur with forced flexion of the knee against active contraction of the quadriceps muscle. The patella is an integral part of the extensor mechanism, and disruption of the continuity of the mechanism results in inability to actively extend the knee. Ruptures of the quadriceps tendon or patellar tendon result from the same mechanism of injury and also present with lack of active extension and with tenderness and palpable defect over the ruptured tendon. Fractures may be transverse, comminuted, or stellate.

Treatment

Nondisplaced fractures are treated by immobilization in a cylinder cast in full extension for 6 to 8 weeks. If separation of the fragments is present at the time of presentation or during later treatment, surgical treatment is indicated. If the fracture is comminuted, excision of the smaller fragments, with attachment of the patellar or quadriceps tendon to the larger fragment using wires or heavy nonabsorbable sutures through drill holes, is performed. For transverse fractures, the tension band wiring technique is recommended, with two

parallel Kirschner wires placed longitudinally through the fragments and a figure-of-eight wire looped over these anteriorly. This arrangement converts tensile force of quadriceps contraction into compressive force at the fracture site. The retinaculum is repaired and the knee immobilized in a cylinder cast or knee immobilizer in extension. In cases of extreme comminution, primary excision of the patella with retinacular repair is recommended. In all injuries, during extension immobilization active quadriceps exercises are started immediately. Range-of-motion exercises are started at 4 to 6 weeks.

Tendon Ruptures

Ruptures of the quadriceps and patellar tendons are similarly treated, using heavy wire or nonabsorbable sutures through drill holes in the patella, followed by immobilization in extension. After 4 to 6 weeks, immobilization is discontinued and range-of-motion exercises started.

Patellar Dislocation

Lateral dislocation of the patella is relatively common, particularly in young women and girls. Medial dislocation is uncommon but occasionally occurs with severe injury or after poliomyelitis. The higher incidence in females is related to the increased valgus of the knee in women. Hypoplastic development of the lateral femoral condyle also can cause patellar dislocations.

The patient complains of the knee giving way as the patella dislocates laterally and usually then spontaneously reduces. Occasionally the patella becomes entrapped lateral to the femoral condyle and requires manipulative reduction. After reduction the patient may have effusion and mild pain. For an acute first-time dislocation, treatment is with immobilization in a cylinder cast in full extension to allow healing of the torn medial retinaculum, which may decrease the likelihood of recurrence. In some cases a small osteochondral fracture of the patella is present and should be repaired arthroscopically. For recurrent dislocations, initial efforts are directed toward strengthening the quadriceps mechanism with straight-leg-raising exercises. If symptoms of subluxation persist, either a bony or soft-tissue realignment procedure is indicated.

Soft-tissue procedures include lateral retinacular release with or without medial reefing of the retinaculum and Insall's proximal tube realignment procedure (lateral release with imbrication of the lateral retinaculum over the quadriceps tendon to the medial retinaculum). In skeletally mature individuals, the medial advancement of the tibial tubercle of Hauser, or in cases with patellofemoral pain and chondromalacia, the anteromedial elevation of the tibial tubercle of Maquet is applicable. While these procedures all improve the incidence of dislocation and subluxation, pain relief has not been nearly as predictable. The procedure of simple lateral retinacular release has come under greater scrutiny in light of experimental results suggesting that it may be less effective than previously purported. Postoperatively, appropriate quadriceps rehabilitative exercises are important. Progressive chondromalacia necessitates further surgical intervention such as arthroscopic chondral debridement or, in severe cases, patellectomy.

Tibial Plateau Fractures

Fractures of the tibial plateau most commonly occur in the middle-aged or elderly population, although falls from a height or motor vehicle accidents can cause this injury in younger persons. Most often the injury results from a valgus stress to the knee with axial loading. The lateral plateau is more frequently involved (60 percent; medial 15 percent; bicondylar 25 percent). Associated injuries to the medial collateral ligament often occur.

The major types include local compression, split fractures of the lateral plateau, depressed fractures of the lateral plateau, comminuted (split- depression) lateral plateau fractures, and bicondylar fractures. Tomograms or CT scans often are helpful in assessing the extent of the injury, particularly if internal fixation is contemplated.

Treatment

Patients with intact collateral ligaments and less than 10 degrees of instability (less than 10 mm joint line depression) can be treated conservatively, with brace or cast immobilization and protected weight bearing, especially in patients over 50 years of age and those with osteopenia, low-demand activity level, or medical contraindications to operation. Patients with more than 10 degrees of instability in extension or major ligamentous injuries, younger patients, and those with high-demand activity levels are candidates for surgical treatment.

Surgical intervention consists of elevation of depressed articular surfaces, with supportive bone grafting if necessary, and internal fixation ranging from one or more large cancellous screws to medial, lateral, or bilateral buttress plates in more severely comminuted fractures. Ruptured collateral ligaments are repaired primarily. Every effort should be made to avoid varus alignment, which is associated with a worse prognosis. Postoperatively, continuous passive motion aids in recovery of range of motion, and a hinged orthosis or cast brace is used in conjunction with protected weight bearing, which must be continued for 8 to 12 weeks, depending on the quality of the bone, comminution of the fracture, and stability of the internal fixation.

Lower Leg Tibial Shaft Fractures

Tibial shaft fractures result from direct trauma such as in motor vehicle and motorcycle accidents, sports, skiing, and falls. All age groups are affected, and about 30 percent are open injuries and 70 percent closed. Various fracture patterns occur, depending on the mechanism of injury (Fig. [41- 88](#)). The rate of healing is slower in fractures that are open, comminuted, or associated with significant soft-tissue injury.

Undisplaced fractures present with local pain, swelling, and the inability to bear weight, but without obvious deformity. Displaced or angulated fractures exhibit obvious deformity on physical examination. Attention must be given to careful neurovascular examination of the extremity, with especial attention to any signs of compartment syndrome.

Treatment

Closed tibial shaft fractures are best handled by conservative treatment, with closed reduction and long leg casting. Early weight bearing can expedite healing and function. In general, some angulation in the anteroposterior direction is acceptable given the compensatory motion of the knee and ankle in the same plane. Varus or valgus angulation of more than 5 degrees should be corrected if possible by manipulation or wedging of the cast. If adequate alignment cannot be achieved, internal fixation with plating or, preferably, with unreamed intramedullary locked or unlocked nailing is advisable. Internal fixation is indicated in the multiple-trauma patient to facilitate immediate mobilization. If significant swelling is present, the patient is admitted to the hospital for elevation and observation of the extremity. Progressive pain or neurovascular symptoms mandate splitting the cast and underlying soft padding, and if the symptoms fail to promptly resolve, measurement of compartment pressures as previously described. The majority of these fractures can readily be treated closed. The cast is changed to a patellar tendon bearing cast or fracture orthosis at 3 to 6 weeks, and weight bearing is permitted as tolerated. Closed tibial fractures generally unite in 12 to 16 weeks, depending on fracture pattern and patient age.

Open fractures of the tibia are treated by primary debridement as previously described. External fixation is the most common approach to fracture stabilization, particularly in grade III injuries (Fig. [41-89](#)). Grade I or grade II fractures can be treated by cast, but if satisfactory alignment cannot be obtained, external or internal fixation is appropriate. The recent development of unreamed locked tibial nails has greatly extended the indications for surgical treatment; several series have shown rates of infection to be no higher than with closed management of grade I and grade II injuries. In cases of significant soft-tissue loss, early coverage with a local gastrocnemius flap (in proximal fractures), a soleus flap (in midshaft fractures), or free vascularized tissue transfers (latissimus dorsi or rectus abdominis) is recommended and decreases the infection rate. Prophylactic fasciotomies at the time of primary debridement decrease the risk of development of compartment syndrome, and neurovascular status should be followed closely postoperatively. Second-look dressing change and debridement is recommended in all grade II and grade III injuries within 24 to 48 h, and intravenous antibiotics should be continued for 7 to 14 days.

In cases of delayed healing, which is relatively common with open tibial fractures, early posterolateral cancellous bone grafting or fibular osteotomy to improve load transmission across the fracture can be considered. Electrical stimulation is useful in delayed union and nonunion.

Ankle

The tibiotalar articulation functions as a hinge joint, with nearly all rotational motion in inversion and eversion occurring at the subtalar joint. The mortise of the ankle is formed by the downward projecting medial and lateral malleoli, which constrain the motion of the talus. The talus is largely covered with articular cartilage, with the blood supply primarily coming through a ring of vessels surrounding the neck. A normal mortise is required for stability of the ankle and painless motion. Additional stability of the talus in the mortise is provided by the distal tibiofibular ligaments (syndesmosis), interosseus membrane, calcaneofibular ligament laterally, and the deltoid ligament medially.

Ankle injuries are caused by a sudden force that exceeds the strength of either the malleoli or ligaments. Ankle injuries are classified according to the position of the ankle and foot at the time of the injury, based on experimental work by Lauge-Hansen. These positions include supination- adduction, supination-eversion (external rotation), pronation-abduction, and pronation-eversion. Supination-adduction fractures result in transverse fracture of the lateral malleolus with or without adduction fracture of the medial malleolus. Supination-eversion fractures are described in four stages of progressive injury involving tear of anterior tibiofibular ligament, spiral oblique fracture of the lateral malleolus, fracture of the posterior lip of the tibia, and fracture of the medial malleolus or deltoid ligament tear. Pronation-abduction fractures sequentially result in medial malleolar fracture or deltoid ligament rupture, syndesmosis disruption, posterior tibial lip fracture, and supramalleolar fracture of the fibula. Pronation-eversion injuries cause medial malleolar/deltoid disruption, anterior tibiofibular disruption, spiral supramalleolar fibular fracture, and posterior tibial lip fracture.

The injured ankle should be carefully examined for localization of tenderness, swelling, and deformity. Ability to bear weight does not rule out significant injury. The ankle can be splinted with a plaster splint or air splint until radiographs are obtained. Anteroposterior, lateral, and mortise radiographic views are necessary for adequate evaluation of ankle injuries. The mortise view is an AP film taken in 30 degrees of internal rotation, which places the malleoli in the same coronal plane.

Ligamentous Injuries

The most common injury is inversion sprain of the lateral ligaments of the ankle. The anterior talofibular ligament usually is involved, with local tenderness and ecchymosis over it. For grade I and grade II sprains, ambulatory treatment with taping or an air splint is adequate. For more severe grade III sprains, protected weight bearing and cast or splint immobilization are indicated for 2 to 4 weeks, followed by range-of-motion and muscle strengthening exercises. For recurrent inversion sprains, repair of the anterior talofibular and fibular collateral ligaments following the Brostrom-Gould method may provide satisfactory stability without excessive subtalar joint stiffness (Fig. [41-90](#)). Preoperatively, inversion instability is identified by inversion stress radiographs demonstrating more than 10 degrees of difference in the talar tilt as compared to the uninjured side. For patients with generalized ligamentous laxity or failed repairs, reconstruction of the calcaneofibular and anterior talofibular ligaments using a slip of the peroneus brevis tendon can be performed (Evans or Chrisman-Snooks procedures).

Distal Tibiofibular Diastasis

Eversion injuries to the ankle can result in disruption of the anterior and posterior tibiofibular ligaments (syndesmosis), with widening of the ankle mortise. Maintenance of an adequate closed reduction is difficult, since more than 1 to 2 mm of talar shift in a young person causes significant alteration of stresses in the articular cartilage of the ankle joint and is unacceptable. A Maisonneuve fracture is a fracture of the proximal fibula in association with syndesmotic diastasis, resulting from proximal propagation of the external rotation force from the talus along the interosseus membrane.

Surgical treatment of tibiofibular diastasis consists of closed or, rarely, open reduction with placement of a syndesmosis screw through the fibula just above the syndesmosis, affixing it to the tibia. The foot should be in the neutral position and not in plantar flexion, as the narrower width of the posterior talus would allow overreduction and subsequent limitation of dorsiflexion. A short leg cast is applied and weight bearing avoided for 8 weeks to prevent screw breakage, and then the screw is removed and weight bearing permitted as tolerated.

Malleolar Fractures and Dislocations

Undisplaced fractures of the malleoli require immobilization in a cast for approximately 8 weeks. The most common fracture is an isolated, minimally displaced oblique fracture of the lateral malleolus caused by external rotation. Displaced fractures are treated by closed reduction, reversing the direction of applied force that caused the injury. If anatomic restoration of the mortise is not accomplished, open reduction and internal fixation with screws or a small buttress plate is recommended.

Medial Malleolar Fractures

Undisplaced fractures are treated with plaster immobilization. Bimalleolar fractures with disruption of the mortise require accurate reduction, and the majority can be treated conservatively. Closed reduction is performed under local anesthesia or with intravenous sedation, with the leg hanging over the end of the table. Longitudinal manual traction is applied and the foot is inverted and internally rotated as it is brought up to 90 degrees. A well- molded long leg cast is applied and check radiographs obtained. More than 1 to 2 mm of displacement usually requires open reduction and internal fixation. The medial malleolus usually is fixed with one or two screws and the fibula with interfragmentary screws or a small buttress plate (Fig. [41- 91](#)).

Posterior Malleolar Fractures

The posterior malleolus usually is fractured in combination with the medial or lateral malleolus. If the fragment comprises less than 30 percent of the articular surface, accurate reduction is not important, but with larger fragments, anteroposterior instability will result from residual displacement. Therefore, with larger fragments, screw fixation after open or indirect closed reduction is indicated.

Pilon Fractures

Pilon fractures are severely comminuted fractures of the ankle and distal tibia, usually resulting from high-energy axial injury, and often involving primarily the anterior lip of the tibial plafond. Vertical compression drives the talus into the tibial plafond, and the entire distal tibia can be comminuted. If excessive comminution is not present, open reduction and internal fixation of the fragments is appropriate. In the face of osteopenia or severe comminution, external fixation, augmented in some cases by limited fixation of one or more larger fragments, is the treatment of choice. Problems with posttraumatic arthritis and ankle stiffness are common with these injuries, and later arthrodesis is often necessary.

Foot Fractures and Dislocations of the Talus

Talus fractures usually occur through the neck, although the body can be involved. The usual mechanism is forcible dorsiflexion of the foot, impinging the neck on the anterior lip of the distal tibia.

In more severe injuries, subtalar dislocation is associated with the talar fracture. The most severe fracture-dislocations involve dislocation of the talus from both the subtalar and ankle joints, usually with posterior extrusion of the talar body. Isolated subtalar dislocations can also occur without talar neck fractures, either medially or laterally from plantar flexion forces with inversion or eversion.

Hawkins has classified talar fractures as follows:

Type I—nondisplaced fracture of the neck of the talus

Type II—displaced fracture of the talar neck with subtalar dislocation

Type III—displaced talar neck fracture with subtalar and ankle dislocation of the body of the talus

Type IV—displaced talar neck fracture with subtalar, ankle, and talonavicular dislocation

The major complication of these injuries is osteonecrosis of the body of the talus, which is primarily supplied with blood through the neck of the talus. Osteonecrosis has been reported to occur in 0 to 13 percent of type I fractures, 20 to 50 percent of type II fractures, and 83 to 100 percent of type III fractures. The best indication of viability of the talus after treatment of the injury is subchondral lucency in the talus, which should be present 6 to 8 weeks because of disuse resorption.

Treatment

Type I fractures are treated closed, with short leg cast immobilization for 3 months. Type II fractures are treated similarly, but if anatomic reduction of the talar neck is not achieved, open reduction and internal fixation are recommended. All type II injuries require open reduction and internal fixation.

In cases in which osteonecrosis develops, prolonged protected weight bearing in a patellar tendon bearing brace has been advocated. If collapse of the body occurs, secondary arthritis will develop, and ankle arthrodesis is often necessary. The Blair-type fusion, in which a strut of anterior tibial cortex is slid down into the talar neck (with or without screw fixation), is used for this condition.

Isolated subtalar dislocations usually are treated by closed reduction. Occasionally the talar head is buttonholed through the anterior capsule, or interposition of the extensor brevis muscle (medial dislocations) or posterior tibial tendon (lateral dislocations) blocks closed reduction. The dislocations generally are fairly stable after closed or open reduction, and immobilization in a short leg cast for 3 to 4 weeks is followed by range-of- motion exercises.

Calcaneus Fractures

Fractures of the calcaneus usually result from falls from a height, with axial loading of the lower extremity. Ten percent of calcaneal fractures are associated with compression fractures in the thoracic or lumbar spine, and 26 percent with other lower-extremity injuries. The five major types of calcaneus fractures are extraarticular fractures, avulsion fractures, tongue type fractures, joint depression fractures, and comminuted fractures (Fig. [41-92](#)).

Treatment

Treatment is controversial, ranging from brief immobilization with early range-of-motion exercises to internal fixation with bone grafting. Problems caused by fractures of the calcaneus include widening of the heel, loss of height of the hindfoot, and disruption of the subtalar joint, which can result in loss of inversion/eversion and subtalar arthritis. In general all patients with significant calcaneus fractures should be hospitalized, as severe swelling and skin breakdown can occur. Nondisplaced fractures can be treated with 2 weeks of immobilization followed by active range-of-motion exercises, and protection of weight bearing for 6 to 8 weeks, depending on the fracture type. The most severe fractures involve the subtalar joint, with two major types, tongue and joint depression. In both types, incongruity of the subtalar joint can result, and advocates of no reduction, closed reduction, and open reduction and internal fixation disagree as to which approach to take with these injuries. Böhler's angle (the angle measured between a line across the calcaneal tuberosity and a line across the anterior and posterior subtalar joint; normal range is 25 to 40 degrees) is usually flattened by intraarticular fractures (see Fig. [41-92](#)). CT scans may help to better define the fracture pattern.

There has been renewed interest in internal fixation to anatomically restore the subtalar joint. In young, active patients with intraarticular fractures, current recommendations are for elevation of the subtalar joint with open reduction through a lateral approach, and bone grafting if necessary. The lateral wall is reduced and fixed with a buttress plate, which also stabilizes the medial sustentaculum and sagittal disruption of the calcaneus. Combined medial and lateral approaches can be used. Problems remain, though; lateral wound healing is the most significant early problem, and persistent long-term subtalar stiffness and pain may occur. For severe late subtalar arthritis, subtalar fusion or triple arthrodesis is appropriate. Intraarticular calcaneus fractures require strictly protected weight bearing for a full 3 months.

Metatarsal Fractures

Fifth Metatarsal Fractures

Avulsion fractures of the base of the fifth metatarsal are commonly associated with inversion injuries of the foot. The contraction of the peroneus brevis muscle causes the avulsion. The patient often has an associated lateral ankle ligament sprain and presents with pain and swelling laterally, local tenderness, and inability to bear weight.

Treatment ranges from symptomatic (elastic [Ace] bandage wrapping and protected weight bearing on crutches until comfortable) to short leg cast immobilization for 4 to 6 weeks. Fibrous union of the avulsed fragment is common but rarely causes problems. The diaphyseal or metaphyseal fracture of the fifth metatarsal (Jones fracture), usually caused by torsional force on the forefoot, can have markedly delayed healing, and nonunion is common. Some advocate primary internal fixation of the Jones fracture using an intramedullary screw or a small plate, while others endorse conservative treatment with cast immobilization and protected weight bearing for 2 to 3 months, with surgical intervention (internal fixation and local bone grafting) only if nonunion occurs.

Other Metatarsal Fractures

Single metatarsal fractures are readily treated with short leg cast immobilization for 6 weeks. Fracture-dislocations occur at the tarsometatarsal junction, often in association with metatarsal base fractures. Generally, closed reduction is recommended, but open reduction with Kirschner-wire fixation is indicated in some cases. Buckle fractures of the cuboid may be associated with dislocation or subluxation of the tarsometatarsal (Lisfranc) joint. These injuries are unstable and require open or closed reduction with Kirschner-wire fixation of both the medial and lateral columns of the foot.

Phalangeal Fractures

Phalangeal fractures commonly result from direct trauma, such as being struck by a dropped object or striking the toe against an object. Dislocation of the IP or MTP joints also can occur. Treatment is directed at alignment and pain relief. Generally closed reduction can be accomplished, and buddy taping of the affected toe to an adjacent toe is adequate for protection.

Pelvic Injuries

Acetabular Fractures and Hip Dislocations

Dislocation of the hip is caused by force applied to the femur and can be associated with fractures of the acetabulum or femoral head. The most common mechanism of injury is motor vehicle accidents, although falls from a significant height can cause hip dislocations. Position of the limb and direction of applied force determine the direction of the dislocation. Force applied to an abducted hip can result in anterior dislocation, and striking the knee on a car dashboard with the hip flexed and adducted causes posterior dislocations, often with fracture of the posterior wall of the acetabulum (Fig. 41-93). Direct trauma to the greater trochanter can result in medial wall fractures or central acetabular fracture-dislocations.

Thorough evaluation of acetabular fractures requires 45-degree oblique views (Judet view) of the pelvis to assess the integrity of the anterior (iliopubic) and posterior (ilioischial) columns and the anterior and posterior wall (acetabular rim). In addition, CT scans are helpful, particularly with the newer three-dimensional reconstructions, in fully delineating fracture patterns and demonstrating the presence of any intraarticular bone fragments. Acetabular fractures have been classified by Tile, as shown in Table 41-14, with transverse forms representing a dissociation of the anterior and posterior columns.

Patients with acetabular fractures often have other major injuries, and careful evaluation of the chest, abdomen, spine, and neurologic status is necessary. Prompt reduction of hip dislocations is essential in minimizing the incidence of osteonecrosis of the femoral head.

Anterior Hip Dislocations

These injuries result from forced abduction or anteroposterior force to an abducted thigh; they are much less common than posterior dislocations (10 to 18 percent of hip dislocations). There are two basic types—the superior or pubic type and the inferior or obturator type. The anteriorly displaced femoral head can compromise neurovascular structures. Both types result from abduction and external rotation, but with additional extension for the pubic type and flexion for the obturator type. Femoral head fractures occur in a significant percentage of cases, and late osteonecrosis occurs in about 10 percent. There is a high incidence of late posttraumatic arthritis. The patient presents with the lower extremity abducted and externally rotated. Reduction is accomplished, usually under general anesthesia, by longitudinal traction with subsequent flexion and internal rotation. If the CT scan or radiographs demonstrate intraarticular fragments, arthrotomy and fragment removal is indicated. Postoperatively the patient is maintained in light traction for 1 to 2 weeks, and range-of-motion exercises are initiated. The patient is mobilized on crutches, with protected weight bearing for 6 weeks.

Posterior Hip Dislocations

Posterior hip dislocations can be associated with posterior wall fractures (Fig. 41-94), which significantly impairs stability of the dislocation after reduction if the fragment is large. The patient presents with the thigh adducted, internally rotated, and flexed. Femoral head or neck fractures are sometimes associated, and sciatic nerve injuries are present in 10 percent. Reduction usually is accomplished by longitudinal traction, followed by gentle abduction and external rotation. If the reduction is unstable, fixation of the posterior wall is indicated (see Fig. 41-94). A CT scan to rule out the presence of intraarticular fragments is important, and if any are present, surgical removal is necessary. Promptness of reduction is important to minimize the chances of late osteonecrosis of the femoral head, which has been reported to range from 10 to 50 percent and is often not evident for several years after the injury. After reduction the patient is placed in light traction for 1 to 2 weeks, or until comfortable range of motion is recovered. Protected weight bearing and use of crutches are required for 4 to 6 weeks. Posttraumatic arthritis develops in a high percentage of patients from 5 to 30 years after injury.

Fractures of the Medial and Superior Acetabulum

If the superior dome of the acetabulum is intact, the fracture can be treated with skeletal traction for 8 to 12 weeks, although more aggressive treatment of displaced acetabular fractures is also practiced. In young, active patients, restoration of an anatomic acetabulum presents or decreases the severity of posttraumatic arthritis, and leaves better bone stock for later reconstructive options such as arthrodesis or total hip arthroplasty in the event that arthritic change does occur.

For open reduction and internal fixation of acetabular fractures, anterior (ilioinguinal), posterior (Kocher-Langenbeck), or combined approaches can be used, depending on the fracture pattern. Removal of the greater trochanter by osteotomy makes reduction easier, particularly if there is a transverse intercolumn fracture component, and internal fixation is facilitated by use of newer devices such as pelvic reduction forceps, a femoral distractor, and readily contourable reconstruction plates. Through these approaches, accurate reduction and internal fixation of complex, severely displaced fractures is possible (Fig. 41-95). Generally, the operation is delayed 5 to 7 days until the patient is stabilized, because immediate surgery can be associated with significant blood loss. Postoperative complications include thromboembolic disease, heterotopic bone formation, hip stiffness, and late osteonecrosis of either acetabular fragments or the femoral head.

Pelvic Fractures

The pelvis is an intact ring that protects the viscera and transmits mechanical weight-bearing force from the spine and axial skeleton to the lower extremities. Like acetabular fractures, pelvic fractures most commonly result from relatively high-energy trauma such as motor vehicle accidents, although a significant number of pubic ramus fractures occur from simple falls, particularly in the elderly.

Patients present with pain in the pelvic area, inability to bear weight, ecchymosis, and local tenderness. Often these individuals are victims of multiple trauma and are hemodynamically unstable. When the pelvis is fractured or disrupted, a large amount of bleeding into the pelvis and retroperitoneum can occur, aggravated by the fact that the intrapelvic volume is no longer constrained and the tamponade of venous bleeding that would occur with an intact pelvic ring cannot occur. Inlet and outlet views of the pelvis are essential to assess the integrity of the ring, and CT scans are extremely helpful in evaluating the extent of bony and ligamentous disruption.

Treatment

Stabilization of the patient and assessment of associated injuries are essential. If the patient is in shock and does not respond to standard volume replacement therapy, application of an anterior external fixator in the emergency department with two or three pins in each iliac crest can help significantly to contain hemorrhage and stabilize the patient for further diagnostic evaluation. Approximately 10 percent of patients have bladder or urethral injuries, necessitating urologic evaluation. Neurologic function also must be carefully assessed because of possible lumbosacral plexus injuries. Pelvic fractures have been classified by Young into four major types according to the mechanism of injury: lateral compression (LC), anteroposterior compression (APC), vertical shear (VS), and combined mechanism (CM). LC fractures involve pubic rami fractures, with or without sacral compression injury, iliac wing fracture, or disruption of the sacroiliac joint on one side. APC injuries involve symphysis diastasis, with varying degrees of sacroiliac joint disruption. VS injuries involve vertical displacement of one side of the pelvis. A similar classification system has been proposed by Tile, with type A (stable) equivalent to low-grade LC injuries; type B (rotationally unstable but vertically stable) equivalent to higher-grade LC injuries or APC injuries; and type C (rotationally and vertically unstable) equivalent to VS and CM.

The majority of pelvic fractures can be treated without surgical intervention. Type A or LC fractures usually are stable and can be treated conservatively with bed rest and subsequent protected weight bearing. APC injuries are vertically stable but not rotationally stable. Fractures with significant symphysis diastasis (greater than 3 cm) can be treated with placement of an anterior fixator or plating. In multiply injured patients, this approach facilitates rapid mobilization. In APC injuries with significant sacroiliac joint disruption or bilateral disruption, posterior internal fixation with plates or sacral bars often is indicated. VS injuries require skeletal traction, but if there is more than 1 to 2 cm of superior displacement, internal fixation anteriorly and posteriorly should be considered (Fig. 41-96). Sacral and coccygeal fractures generally are treated symptomatically. Neurologic deficits associated with sacral fractures usually occur in traction injuries, and recovery often is poor and not improved by surgical decompression.

Spinal Injuries

Fractures and dislocations of the cervical, lumbar, and thoracic spine most commonly follow major trauma such as motor vehicle accidents or falls from a height. Overall prognosis depends on associated spinal cord or nerve root injuries. Any patient who complains of neck or back pain or tenderness after an accident should be assumed to have an unstable spine until adequate radiographic evaluation can be carried out. For neck injuries, sandbags can be used for temporary stabilization and transport, followed by application of a hard cervical collar such as a Philadelphia collar. Patients with thoracolumbar injuries are placed on a rigid backboard for transport. Neurologic examination should be conducted as soon as possible, after attention to chest, abdominal, and other injuries. Anteroposterior and lateral radiographs of the cervical, thoracic, and lumbar spine are obtained before moving the patient from the backboard. Lateral radiographs of the cervical spine must include the C7 level, and generally abnormal findings on the plain radiographs are further delineated by CT scan of the injured area.

Cervical Fractures and Dislocations

Flexion injuries of the cervical spine occur most often at the level of C5 and C7, and can involve anterior compression fracture of the vertebral body, posterior longitudinal ligament tear, and unilateral or bilateral facet subluxation or dislocation. Facet dislocation or fracture presents radiographically with anterior displacement of up to 25 percent with a unilateral facet dislocation (associated also with a rotational difference in the vertebra above as compared to the vertebra below the lesion); bilateral facet dislocation can exhibit anterior displacement of up to 50 percent. Cord injury may be present, depending on the degree of canal compromise.

With an axial load to the cervical spine, burst fractures of the body can occur, with retropulsion of fragments into the spinal canal causing neurologic deficit (Fig. 41-97). Depending on the presence of rotation, flexion, and lateral bending, fractures of the pedicles or lamina occur. CT scans can readily delineate these injuries. Injuries of the spinous processes (clay shoveler's fractures) can occur as an avulsion caused by a sudden stress to the supraspinous ligament.

Axial compressive force can cause fractures of the ring of C1 (Jefferson's fracture). This injury, often subtle on plain radiographs, can be identified by asymmetric widening of the distance between the lateral masses and the odontoid process on the open-mouth anteroposterior view. Fractures of C2 involving the odontoid process or posterior elements (hangman's fracture) result from flexion or extension and rotational forces. In children under the age of 6 years, odontoid injuries can result in an epiphyseal displacement and associated displacement of the atlas upon the axis.

Treatment

Unstable fractures are defined as those with neurologic deficit or potential to cause a neurologic deficit. These injuries must be reduced promptly, usually using tong or halo skeletal traction. The halo consists of four pins inserted under local anesthesia into the outer table of the skull and affixed to a circular ring through which traction can be applied. Additionally, the halo can be attached to a body cast or plastic vest, allowing mobilization of the patient after treatment of the acute injury. Dislocations with fractures are reduced with skeletal traction and confirmed radiographically. Immobilization in a halo vest is maintained for 3 months. Before discontinuation of the vest, flexion/extension lateral radiographs are obtained to be sure that the injury is stable. For bony injuries, healing with immobilization can be anticipated. For severe ligamentous disruptions that do not become stable with immobilization, posterior surgical fusion with spinous process wiring is indicated. This may also be necessary in cases in which satisfactory closed reduction cannot be accomplished.

Patients who present with minimal but progressive neurologic loss may require anterior decompression (for disc herniation) or posterior decompression (for epidural hematoma or depressed laminar fracture). The use of posterior decompression for patients with neurologic deficit is controversial. Late anterior decompression for patients with a complete lesion

allows regaining of one root level in about 50 percent of patients. Unsuccessful reduction with persistence of fragments in the spinal canal constitutes another indication for operative exploration. MRI and CT scans can greatly assist in the diagnosis of operable lesions in such situations.

Thoracolumbar Fractures and Dislocations

Fractures and dislocations in the upper thoracic spine are unusual because of the stabilizing effect of the rib cage. Flexion forces can cause compression fractures; this is particularly common in osteopenic fractures, which often are the result of relatively minor trauma. Flexion/rotation injuries result in fracture-dislocations, most commonly in the vicinity of the thoracolumbar junction. Axial loading also can cause burst fractures of the thoracic or lumbar vertebrae, and retropulsion of fragments into the canal with neurologic deficit can result. Flexion/distraction injuries occur in seatbelted passengers in motor vehicle accidents and can cause transverse fractures through the posterior elements and body (Chance fracture).

Treatment

After the patient is stabilized and other acute injuries assessed, appropriate radiographs are obtained with the patient on a backboard as described above, and a careful neurologic examination is carried out. Treatment of the fracture depends on the degree of stability. Compression fractures are treated symptomatically, with a brace and early ambulation. Burst fractures without neurologic deficit are treated with recumbency and a body cast, and spontaneous fusion across the involved discs will often result. Indication of stability of thoracolumbar spine fractures is based in part on the three-column assessment as well as on the neurologic status. Stability is conferred by the vertebral bodies, anterior longitudinal ligament, and discs (anterior column); the posterior longitudinal ligament, posterior annulus, and posterior vertebral body (middle column); and the posterior elements (posterior column). The middle column is thought to be the key to stability, and disruption of this column increases the likelihood that surgical stabilization will be necessary. Which columns are involved plays a role in determining whether an anterior or posterior surgical approach will be used.

In unstable fractures or those with neurologic deficit and canal compromise, open reduction with Harrington distraction rodding and indirect reduction or direct anterior decompression and stabilization should be carried out. Patients with paraplegia persisting beyond the period of spinal shock (assessed by return of reflexes such as the bulbocavernosus, usually about 24 h after injury) require posterior stabilization with rodding and fusion, most often with Harrington rods (Fig. 41-98). Metal sleeves (Edwards sleeves) over the rods have been used to improve kyphotic correction, and recently the more rigid fixation of segmental instrumentation (Luque or segmentally wired Harrington rods) has been advocated, particularly in complete neurologic lesions. This allows rapid mobilization of the patient and minimal external support. The cross-linked Cotrel-Dubousset instrumentation also has been used increasingly in trauma, and pedicle screw-fixed rods have been used for lower lumbar fractures.

DISEASES OF JOINTS

Afflictions involving joints are widespread. Degenerative joint disease affects most individuals over the age of 50 with decreased range of motion, although only a small proportion are symptomatic. Injuries that cause fractures also frequently involve joints, and the immobilization associated with prolonged casting for fractures invariably causes some temporary joint dysfunction.

Anatomy

Joints can be diarthrodial (involving articulating cartilage surfaces), or synarthrodial (involving the fibrous junction). Most major joints—the shoulder, knee, and hip—are diarthrodial. Synarthroses include the intervertebral discs, symphysis pubis, and ankle syndesmosis.

Diarthrodial joints are surrounded by a fibrous capsule that is lined with synovium. The synovium secretes the joint fluid, which lubricates the joint and provides nutrition to the articular cartilage.

Articular Cartilage

Normal articular cartilage is blue-white, smooth, glistening, and slightly compressible. The matrix of articular cartilage consists mainly of type II collagen and proteoglycan, with a smaller amount of glycoproteins and minor collagens such as type IX (which coats the type II collagen fibrils and binds them to the proteoglycan in the matrix) and type XI, which regulates fibril diameter. One of the glycoprotein components, chondronectin, binds chondrocytes to the matrix type II collagen. Normal articular chondrocytes do not replicate, although they do synthesize and turn over matrix components. After injury or in degenerative disease, chondrocytes have a limited ability to attempt repair, and they replicate to a limited degree (cloning). Under such conditions they also increase production of proteoglycan. Articular cartilage has no blood vessels and must derive its nutrition from the synovial fluid. The collagen and proteoglycan are maintained in a constrained molecular configuration in cartilage, providing its viscoelastic compressibility and load transmission characteristics. The collagen has a specific orientation in articular cartilage, as has been previously discussed.

Synovium

The synovium in a normal joint is only a few cell layers thick. Over ligaments the synovium is quite thin, but in areas of fat pads and loose capsule there is considerable areolar tissue beneath the synovial layer. The synovium is well vascularized from a plexus of vessels beneath the synovial layer. The cells are of three types: type A (macrophage-like, secretory cells); type B (fibroblast-like cells), and type C (dendritic cells of uncertain function).

The secretory cells produce the hyaluronic acid of the synovial fluid, which contributes to its lubricative viscosity. Synovial fluid also contains some plasma proteins, water, and electrolytes. Normal synovial fluid is clear, yellowish, and viscous and contains about 200 nucleated cells (polymorphonuclear leukocytes, lymphocytes, and monocytes).

The synovium is innervated, as is the joint capsule, which in addition to unmyelinated sensory fibers contains proprioceptive fibers. Generally the nerve supply to a joint is from the nerves innervating muscles that act across the joint. Nerve fibers accompany arteries into the subchondral bone and provide the basis for bone pain in some inflammatory conditions of joints. The proprioceptive fibers in the capsule and associated ligaments are important in providing information regarding joint position. There are no nerves in articular cartilage.

Articular Cartilage Degeneration

As already noted, articular cartilage has a limited repair capacity. Loss of proteoglycans occurs early after any insult (trauma, chemical, inflammatory, infectious, autoimmune) to a joint. This causes loss of elasticity of the cartilage and swelling with pathologic hydration. Load transmission becomes defective, causing progressive damage to the articular chondrocytes, which respond by replicating and synthesizing matrix molecules. The cells also paradoxically synthesize matrix metalloproteinases (collagenase, stromelysin, and gelatinase), which mediate further matrix breakdown. The mechanical forces on the joint in the context of now abnormal mechanical properties result in fibrillation of the cartilage, and debris from the fragments of matrix induces an inflammatory response. The synovium responds to this with inflammatory infiltration and hypertrophy. Secretion of metalloproteinases from the synovial cells also occurs, further contributing to the matrix breakdown. The vicious cycle of cartilage degradation continues in a self-propagating manner. The final common pathway for all inflammatory, degenerative, and traumatic afflictions of joints is thus progressive cartilage degradation. New therapeutic approaches to the treatment of arthritis involves inhibitors of metalloproteinases or pharmacologic stimulation of the production of the endogenous tissue inhibitor of matrix metalloproteinases.

In response to the activation of latent endochondral calcification in the deep layer of the articular cartilage, the subchondral plate begins to thicken (subchondral sclerosis). This contributes to acceleration of the breakdown of matrix because of stiffening of the subchondral bone with increased loading of the cartilage. The sequence of events in cartilage degeneration is shown schematically in Fig. 41-99.

Examination

The diagnosis of joint disease depends on the following:

- (1) Clinical history
- (2) Physical examination of all joints
- (3) Radiographs
- (4) Synovial fluid analysis
- (5) Serologic tests
- (6) Other imaging modalities, such as MRI and bone scans

Localization of pain, identification of aggravating factors and traumatic incidents, and the presence of swelling, stiffness, and mechanical symptoms of locking are important aspects of the history. Other considerations include the presence of systemic diseases or symptoms.

Clinical Examination

Involved joints should be inspected for swelling, effusion, warmth, and erythema, and unilateral cases should be compared with the opposite side. Range of motion, actively and passively, is assessed and ligamentous stability tested. Presence of muscle atrophy is noted and measured, and regional adenopathy, neurologic examination, and vascular examination of the extremity are carried out. Other joints in both the upper and lower extremities also should be examined, and the range of motion of the cervical and thoracolumbar spine should be assessed.

Radiographic Examination

Standard anteroposterior and lateral radiographs of any involved joint(s) are obtained. If ligamentous instability is present, stress radiographs can be helpful. Joint space narrowing; presence of effusion, intraarticular calcified loose bodies, osteophytes, subluxation, subchondral sclerosis, articular cartilage or meniscal calcification (chondrocalcinosis); and subchondral cyst formation are some of the more important findings that can be detected radiographically. Arthrography and MRI are used to evaluate the possibility of meniscal or rotator cuff lesions. MRI is increasingly useful in the diagnosis of nonosseous lesions in joints, and it can allow assessment of chondral lesions and degenerative changes that are not apparent on plain radiographs.

Synovial Fluid Analysis

The abnormalities in the synovial fluid are specific or nonspecific, depending on the cause of the joint problem. A specific diagnosis can be obtained for infection, gout, and pseudogout (calcium pyrophosphate deposition disease). The technique of aspiration depends on the involved joint and local anatomy. Aspiration through any cellulitic or fluctuant areas suggestive of infection must be strictly avoided to prevent the introduction of bacteria into a joint. The knee is aspirated either medial or lateral to the patella, the ankle anterolaterally at the corner of the mortise, the shoulder anteriorly, the elbow laterally at the radiohumeral joint, the hip anterolaterally or anteriorly, and the wrist dorsally. Sterile preparation and draping and meticulous sterile technique are necessary to prevent infection or contamination of cultures if infection is already present. Fluid is collected in (1) a culture tube, (2) a tube with EDTA to prevent clotting, and (3) an empty tube for chemistry or immunologic analyses.

Normal synovial fluid is slightly straw colored and clear. In inflammatory conditions or infection the fluid ranges from turbid to frankly purulent. In inflammatory conditions, the hyaluronate in the fluid is depolymerized and has decreased viscosity. Adding synovial fluid to 5% acetic acid should form a firm mucin clot that does not separate or fragment with agitation. In inflammatory conditions, a clot might not form or might fragment as a result of breakdown of glycoproteins and mucopolysaccharides in the fluid. The synovial fluid analysis in various disorders is shown in Table [41-15](#).

Fluid should be cultured for aerobic and anaerobic organisms as well as mycobacteria, fungi, and gonococci, with primary plates, chocolate agar (for gonococcal infection), and broth cultures to detect fastidious organisms.

A total cell count and differential count is performed on fluid in an anticoagulated tube. Joint fluid is examined also for the presence of urate or calcium pyrophosphate crystals after collection in a tube without EDTA or oxalate. After centrifugation of the fluid to concentrate the cells, the sediment is examined using polarized light microscopy. Urate crystals are negatively birefringent and rod-shaped. Calcium pyrophosphate crystals are rhomboidal and weakly positively birefringent. Glucose in joint fluid will also be significantly lower than serum values if large numbers of inflammatory cells are present.

Pyogenic Arthritis

With early diagnosis of septic arthritis and appropriate treatment, the prognosis for maintenance of normal joint function is excellent. The lysosomal enzymes released from white cells in the joint, however, can permanently destroy the articular cartilage and lead to progressive degenerative change if inadequately treated. Joint infections can result from hematogenous spread from septicemia, direct infection from adjacent traumatic wounds or surgery, or extension of an adjacent metaphyseal osteomyelitis.

Staphylococcus aureus and hemolytic streptococci are the two most common organisms that cause pyogenic arthritis. Gonococcal and coliform organisms, *Hemophilus influenzae* (in infants), pneumococci, meningococci, and *Brucella* are other causative agents. With the advent of widespread *H. influenzae* vaccination in children, the incidence of infections because of this organism has dramatically declined.

Septic arthritis occurs in children and adults and is more common in debilitated patients or those undergoing steroid therapy or immunologic suppression.

In patients with hematogenous spread from bacteremia several joints may be involved. Complaints include fever, pain and swelling in the affected joint, chills, and malaise. Elevation of the white blood count and erythrocyte sedimentation rate are common. The affected joint(s) are swollen, tender, erythematous, warm to touch, and painful during range-of-motion maneuvers. The physical findings vary greatly and often depend on the virulence of the organism; indolent infections present with a nearly normal examination.

Radiographs reveal effusion. In more advanced infections, erosions and joint space narrowing occur as consequences of the destruction of the articular cartilage (Fig. [41-100](#)). With chronic infection or a large abscess, subluxation of the joint also may be evident. Diagnosis is made by aspiration of the joint with culture, Gram stain, and analysis of the synovial fluid. The presence of organisms and an elevated white cell count in the fluid are diagnostic. The white cell count in the fluid may range from 25,000/mm³ to more than 200,000/mm³, usually with a high percentage (greater than 90 percent) of leukocytes.

Treatment

In the presence of an elevated synovial white cell count and other signs and symptoms of septic arthritis, antibiotic therapy is started empirically, usually with coverage for *S. aureus* (cephalosporin), streptococci (cephalosporin or a penicillin), and, in children, *H. influenzae* (ampicillin).

Controversy exists as to whether surgical drainage, arthroscopic drainage, or repeated daily aspiration represents the best treatment option. Generally it is agreed that for hip infections, which can result in rapid destruction and secondary osteonecrosis and can be difficult to aspirate, emergent surgical drainage is indicated. Chronic infections with loculation or thick purulence also require surgical drainage. In chronic infections removal of hypertrophic infected synovium can be advantageous. For the knee, arthroscopic drainage may be appropriate. Shoulder and ankle infections can be managed either by sequential aspiration or surgical drainage. In general, patients who undergo surgical drainage undergo defervescence and improve clinically more rapidly. Depending on the organism, intravenous antibiotic therapy is indicated for 2 to 4 weeks. If the septic arthritis results from extension from an adjacent osteomyelitis, intravenous antibiotic therapy for 6 weeks or longer is needed. Early joint motion is encouraged to restore nutrition to the articular cartilage and prevent stiffness.

For cases in which articular cartilage destruction progresses despite treatment, late secondary degenerative arthritis can result. Options for treatment when symptoms are severe include arthrodesis, resection arthroplasty, and, in rare cases, joint replacement arthroplasty.

Bone and Joint Tuberculosis

Tuberculous Arthritis

Although skeletal tuberculosis has been rare in the United States since the advent of appropriate antibiotic therapy, the incidence is increasing among immigrants, patients with HIV infection (AIDS), and patients on chronic immunosuppressive therapy. In addition, drug-resistant strains of tuberculosis are now emerging.

The most common site of skeletal involvement is the spine. The infection starts adjacent to the disc space and spreads across the disc to involve contiguous vertebrae. In the lumbar spine the infection can dissect into the psoas muscles, with subsequent abscess formation. Peripheral joint involvement affects the synovium, bone, and cartilage. Hypertrophic infected synovium (pannus) gradually covers the cartilage surface and erodes the subchondral bone, eventually destroying the cartilage as well. The articular cartilage (radiographically, the joint space) is well preserved until late in the course, unlike in pyogenic arthritis, in the course of which articular cartilage destruction occurs early. The result is complete destruction of the joint with fibrous ankylosis.

Tuberculosis can affect any large joint, and any chronic monoarthritis should be suspected of being tuberculous. Pulmonary disease is not necessarily present. The clinical onset is insidious, and symptoms are often present for weeks or months. Limitation of motion and swelling can occur, but marked signs of inflammation generally are absent.

Radiographs usually show soft-tissue swelling and, later, marginal erosions, but the joint space is maintained initially (Fig. [41-101](#)). In children enlargement of the adjacent epiphysis occurs, and in adults with long-standing disease complete destruction of the joint results.

The diagnosis depends on recovery of the organisms from the joint. Demonstration of pulmonary lesions or a positive purified protein derivative test is helpful but not pathognomonic. Joint aspiration and/or synovial biopsy is essential, with demonstration of acid-fast bacilli on smears or positive cultures. Synovial fluid analysis usually demonstrates a leukocytosis of less than 20,000/mm³. If aspiration is not diagnostic, open synovial biopsy is necessary.

Treatment

Therapeutic measures include antibiotic therapy, symptomatic treatment of the involved joint, and surgical debridement. General supportive measures include hydration, rest, and proper caloric and protein intake. Antituberculous chemotherapy is begun, and the treatment is highly successful if initiated before significant necrosis and abscess formation have occurred.

Usually triple-drug therapy with rifampin, isoniazid, and ethambutol is used. In resistant infections, a fourth drug such as streptomycin or para-aminosalicylic acid (PAS) is added. Treatment is for 6 to 12 months. Involved joints should be immobilized by traction or splinting. If there is subchondral bone involvement or if the joint fails to respond to drug treatment, surgical debridement often is indicated. In joints with severe destruction, arthrodesis is the treatment of choice. Patients with disease that has been inactive for 10 years can be considered for total joint arthroplasty, although reactivation rates of infection may run as high as 25 percent.

Tuberculosis of the Spine

Tuberculous spondylitis (Pott's disease) is the most common form of skeletal involvement (Fig. [41-102](#)). The infection can involve the thoracic or lumbar spine. Vertebral destruction results, and multiple levels can be involved, often with kyphotic deformity. Neurologic deficit or paralysis results in severe cases. Treatment involves appropriate chemotherapy and anterior surgical debridement, with abscess drainage and stabilization of the spine with anterior rib strut bone grafts. Antibiotic therapy is continued for 6 to 12 months after operation, and patients are immobilized in plaster body casts until fusion has occurred.

Tuberculosis of the Hip

Initial treatment of tuberculous involvement of the hip joint consists of appropriate antibiotic therapy and light traction. If symptoms continue, open debridement and synovectomy, or in cases of severe destruction, arthrodesis, are applicable.

Gonococcal Arthritis

Gonococcal arthritis is more common in females and results from spread from cervicitis or vaginitis. Symptoms often begin with migratory polyarthralgias, followed by localization in one or two joints. The knee is commonly affected, and systemic symptoms are variable, as is the severity of the local findings. Diagnosis depends on identification of gonococcal organisms by joint aspiration and culture. Synovial fluid analysis is compatible with a moderate inflammatory process. Radiographs often reveal no abnormality.

Intramuscular penicillin G 1 to 2 million U/day for 2 weeks is sufficient to eradicate the infection. If persistent purulent effusions are present, serial aspiration may be helpful, but usually this is not needed. If the patient fails to respond to antibiotic treatment, then other diagnoses such as Reiter's syndrome or early rheumatoid arthritis, should be considered. With proper treatment, normal recovery of joint function usually results.

Lyme Arthritis

Lyme disease is a tick-borne (deer tick, *Ixodes dammini*) illness caused by the spirochete *Borrelia burgdorferi*. Affected patients may have intermittent attacks of swelling in one or more large joints, accompanied by a characteristic skin rash (erythema chronicum migrans) that often precedes the arthritis. The intermittent attacks and rash comprise stage I Lyme disease and have an onset 3 to 30 days after a tick bite. Stage II can involve cardiac manifestations and neurologic symptoms such as Bell's palsy. In stage III a chronic low-grade septic arthritis develops. The knee joint is most often involved. Treatment includes tetracycline, penicillin, or erythromycin for 4 weeks.

Rheumatoid Arthritis

Rheumatoid arthritis is a systemic disease that affects not only the joints but the cardiovascular, nervous, and respiratory systems as well. Rheumatoid arthritis affects diarthrodial joints and all the supporting structures, including synovium, tendons, tendon sheaths, and bursal tissues. The synovium characteristically undergoes inflammatory infiltration and hypertrophy. The hypertrophic synovium (pannus) creeps over the articular surface, destroying the cartilage. At the joint margins the inflamed synovium induces osteoclastic bone resorption and creates periarticular erosions. The cartilage undergoes symmetrical thinning, and the ligamentous structures and capsule of the joints also erode and become lax, allowing joint subluxation. The adjacent bone becomes osteopenic from a combination of increased blood flow, immobility due to pain, and accelerated osteoclastic activity possibly related to inflammatory cytokines. Ultimately the process in a given joint can become “burned out,” with subsidence of the inflammation and ankylosis of the joint.

In about 25 percent of patients subcutaneous nodules develop along the ulna, the olecranon, or the dorsal aspect of the fingers or feet. Tendon ruptures from attrition of involved tenosynovium are common, especially in the shoulder and hand. Pericarditis and valvular involvement can occur, and granulomatous lesions can occur in the lungs. Uveitis also is common in rheumatoid patients.

The diagnosis of rheumatoid arthritis in a patient with long-standing disease and deformity is not difficult, but in early stages it may not be straightforward. Arthritis associated with collagen vascular diseases, systemic lupus erythematosus, psoriasis, or dermatomyositis can have a similar presentation. Measurement of rheumatoid factor may be helpful. Rheumatoid arthritis may present with morning stiffness, one or more swollen, tender, or painful joints that persist continuously for weeks to months, subcutaneous nodules, iritis, and symmetrical joint swelling.

The differential diagnosis includes lupus, polyarteritis, erythema nodosum, rheumatic fever, gout, tuberculous arthritis, Reiter's syndrome, hypertrophic pulmonary osteoarthropathy, ochronosis, multiple myeloma, and lymphoma.

Ninety percent of rheumatoid patients have an immunoglobulin in the serum called rheumatoid factor, although in juvenile rheumatoid arthritis only 20 percent are positive. The erythrocyte sedimentation rate is a good index of the disease activity. Synovial fluid analysis reveals leukocytosis of greater than 50,000/mm³. Radiographic findings include joint space narrowing, periarticular erosions, soft-tissue swelling, osteopenia, and joint subluxations (Fig. [41-103](#)). The joint space narrowing tends to be symmetrical or concentric, as opposed to the eccentric narrowing seen with degenerative arthritis.

Treatment involves a team effort on the part of the rheumatologist, orthopaedist, physical therapist, occupational therapist, and social worker. The aim of orthopaedic treatment is to halt disease progression, restore or maintain function, and relieve pain.

Medical Management

Numerous anti-inflammatory medications are available for treatment of the rheumatoid patient. Drugs include analgesics (codeine, aspirin), mild anti-inflammatory/analgesics (ibuprofen, naproxen), high-potency anti-inflammatory/analgesics (phenylbutazone, indomethacin), corticosteroids, and drugs that appear to modify the immune response (gold salts, penicillamine, antimalarials, azathioprine, cyclophosphamide, methotrexate). Corticosteroids are widely used in the management of rheumatoid patients, but usually are reserved for acute exacerbations or life-threatening situations such as vasculitis.

Orthopaedic Management

Efforts should focus on maintenance of muscle strength and range of motion of joints and avoidance of deforming forces. Protection of symptomatic joints with walking aids or splints is helpful. Varus and valgus deformities of the lower extremities can be treated with orthoses. The initial approach should focus on optimization of medical management of the disease. For persistent symptomatic synovitis, surgical synovectomy is helpful in slowing progression of the disease. In the knee, arthroscopic synovectomy can be undertaken and provides similar results to those obtained by arthrotomy. Radiation synovectomy by intraarticular injection of short-half-life isotopes such as dysprosium-165 ferric hydroxide has provided results comparable to those of surgical synovectomy. External beam radiation synovectomy too has been reported to have some success in early rheumatoid arthritis.

Knee

With joint space narrowing and collateral ligament laxity in the knee, instability results. Synovectomy at this stage is ineffective, and generally total knee replacement arthroplasty is indicated. Total knee arthroplasty has evolved to a degree of success comparable to that of total hip arthroplasty (86 to 95 percent good and excellent results at 5 years in total knee and total hip arthroplasty, respectively). Total knee replacement relieves pain and improves ambulatory function (Fig. [41-104](#)). Complications include infection, deep venous thrombosis, and late loosening. Infection requires removal of the prosthesis and all cement and the institution of appropriate intravenous antibiotic therapy, after which exchange arthroplasty can be undertaken, although the success rate is substantially lower than for primary arthroplasty.

Hip

Synovectomy generally is not performed for hip involvement in rheumatoid arthritis. For the patient with disabling pain and stiffness, total hip replacement arthroplasty has been quite successful (see Total Hip Replacement Arthroplasty under Osteoarthritis, below). Complications include infection and loosening of the implants. Although most rheumatoid patients undergoing total joint replacement are treated using cemented components, in younger patients the use of uncemented components, which have porous coatings allowing bony ingrowth, may prove to be acceptable, especially on the acetabular side.

Ankle and Foot

Most rheumatoid deformities about the ankle can be controlled by appropriate bracing. Patients with severe pain or deformity can benefit from arthrodesis. Total ankle joint replacements have had a high rate of loosening and unsuccessful outcomes and are not currently recommended.

Involvement of the feet with rheumatoid arthritis is quite common. Clawtoe occurs, with dislocation of the MTP joints and flexion of the IP joints, and the resulting painful plantar prominence of the metatarsal heads interferes significantly with ambulation. The Hoffman procedure is used for this problem and has an excellent success rate. It consists of resection of all of the metatarsophalangeal joints, with realignment and temporary Kirschner-wire fixation. Correction of hallux valgus by arthrodesis of the first MTP joint also can be helpful, and in conjunction with the Hoffman procedure on the other toes helps to prevent recurrent lateral drift (Fig. [41-105](#)).

Hand

The hand is one of the most common areas of problematic rheumatoid involvement. The MCP, PIP, and carpal joints and the tenosynovium of both the flexor and extensor tendons can be affected. Laxity of capsules allows ulnar drift of the digits. Ruptures of involved tendons can occur, but can be prevented by early tenosynovectomy. Common tendon ruptures include the extensor digiti minimi, common extensors to the ring and little fingers, extensor pollicis longus, flexor pollicis longus, and finger flexors. Associated carpal tunnel syndrome can occur with flexor tenosynovitis in the carpal tunnel and frequently requires carpal tunnel release with tenosynovectomy. Severe involvement of the MCP joints necessitates silicone replacement arthroplasties. Thumb deformities also are common and require stabilization of the MCP or IP joint by arthrodesis. Similarly, painful involvement of the thumb CMC joint often requires excisional arthroplasty, with tendon or silicone interposition arthroplasty.

Other Upper Extremity Deformities

Elbow involvement can occur in rheumatoid arthritis, with pain, restriction of motion, and instability. Splinting is helpful along with appropriate medical management. In selected cases, arthroplasty is necessary. Fascial interposition arthroplasties have been somewhat successful, but recent advances with unconstrained total elbow replacement suggest improved results over prior devices.

Shoulder involvement also is common, as is involvement of the rotator cuff with the synovitis, often leading to attrition or rupture of the rotator cuff. Results of total shoulder arthroplasty are not as successful as with osteoarthritis, in which the cuff involvement tends to be less severe, but may nevertheless afford pain relief and improved function.

Spine

Back pain is common in rheumatoid patients, and compression fractures and kyphosis develop as a result of osteopenia, often aggravated by chronic steroid use. More serious potential problems result from involvement of the cervical spine, with instability and the potential for neurologic deficit. Destruction of joints and ligaments can result in basilar invagination or cranial settling, subaxial subluxation, and atlantoaxial subluxation. Cranial settling and atlantoaxial subluxation can result in neurologic deficit, and posterior fusion, decompression, and wiring may be necessary to prevent quadriplegia. Evaluation of the cervical spine in rheumatoid patients is particularly important when considering surgical intervention requiring general anesthesia and intubation, which can be dangerous with instability of the upper spine. Neurologic evaluation of the rheumatoid patient as a part of routine follow-up is also important for this reason.

Osteoarthritis

Osteoarthritis is a term used to describe degenerative changes in diarthrodial joints. The primary change is in the articular cartilage, which becomes soft, loses elasticity, and fibrillates, ultimately resulting in eburnated bone devoid of cartilage. It is the final common pathway of degeneration of joints from nearly all insults, and occurs secondarily in joints previously damaged by trauma, inflammation, or sepsis.

The early changes in articular cartilage degeneration include loss of metachromatic staining, which is a result of the loss of proteoglycans in the territorial matrix of the chondrocytes. Swelling of the cartilage and fibrillation occur subsequently, as previously described. The cartilage grossly appears more yellow and is softer than normal cartilage. Mechanical attrition then causes flaking of debris from the damaged cartilage, which induces mild inflammatory changes in the synovium, contributing to the progressive breakdown by secretion of cytokines and proteases.

With loss of cartilage and eburnation of the underlying subchondral bone, marginal osteophytes develop, often at sites of ligamentous attachments. Reactivation of endochondral ossification in the deep layer of the cartilage as well as the increased abnormal mechanical stresses in the subchondral bone contribute to subchondral thickening and sclerosis. This also supports the vicious cycle of degeneration by increasing mechanical stresses in the cartilage. Subchondral degenerative cysts develop where the subchondral bone marrow has undergone mucoid degeneration, but the mechanism of cyst formation is unknown.

Scarring and fibrosis of the capsule occur, with loss of range of motion. Mild villous synovitis may be present, but synovial fluid cell counts are only mildly elevated (less than 2,000/mm³). Because of the relatively noninflammatory nature of osteoarthritis, many prefer the term osteoarthrosis.

Osteoarthritis can result from derangement of a joint by any of a wide variety of mechanisms, including congenital hip dysplasia, Legg-Calvé- Perthes disease, slipped capital femoral epiphysis, aseptic necrosis, septic arthritis, hemophilia, gout, pseudogout, and trauma. If no predisposing factor exists, the degenerative process is considered primary. Structural mutations have been identified in the type II collagen gene in patients with osteoarthritis, which may explain at least some forms with familial tendencies. Primary osteoarthritis affects mainly the large weight-bearing joints such as the hips and knees (Fig. 41-106), but generalized forms exist, most often in postmenopausal females, involving multiple joints, including the DIP joints. Osteophytes along the DIP joints are called *Heberden's nodes*. In generalized primary osteoarthritis the course often is more rapidly progressive, with more pronounced clinical signs of inflammation.

Clinical Manifestations

Radiographically, the evidence of degenerative changes in joints increases with age, but only a small proportion of these changes are associated with symptoms. Only 5 percent of patients over the age of 50 years and 20 to 30 percent of patients over the age of 60 years are estimated to have clinical symptoms. The onset of osteoarthritis usually is insidious, with radiographic changes often preceding symptoms. Stiffness is noted after rest and resolves with mild exercise of the joint. Swelling and effusion are often absent. There is no specific diagnostic test for osteoarthritis, although the disease usually is diagnosed radiographically on the basis of eccentric joint space narrowing, osteophyte formation, subchondral sclerosis, and subchondral cyst formation (Fig. 41-107). The radiographic findings and clinical symptoms frequently are poorly correlated.

Treatment

Therapeutic measures include (1) modification of activities to avoid high- impact forces on the joint; (2) anti-inflammatory analgesic medications; (3) weight loss; (4) range-of-motion exercises to minimize contractures; (5) walking aids such as a cane (because of the moment arm of the hip abductors, use of a cane in the contralateral hand results in a fivefold reduction in the forces acting across the joint; (6) orthoses to control instability if present; and (7) avoidance of aggravating activities.

Osteoarthritis tends to be only very slowly progressive, and symptoms can be satisfactorily managed conservatively for many years. When symptoms become intolerable, surgical intervention might be indicated. Procedures include osteotomy, arthrodesis, joint replacement arthroplasty, and resection arthroplasty. Arthroplasties provide pain relief and improved range of motion, and osteotomy alters weight-bearing forces to decrease pressure on damaged cartilage and slows progression.

Thumb Carpometacarpal Joint

Osteoarthritis of the thumb CMC joint is much more common in women, particularly after menopause. Patients have pain with motion of the thumb and interference with grasp because of pain. Conservative measures include the use of an orthosis to restrict CMC joint motion in addition to the approaches mentioned above. Rarely, surgical intervention (such as arthrodesis) is necessary. Implant arthroplasty also is useful, but loosening and silicone synovitis can occur. If both the CMC and the intercarpal joints are involved, resection of the trapezium with interposition of a folded segment of flexor carpi radialis tendon has been successful.

Wrist, Elbow, and Shoulder

The need for surgical treatment of osteoarthritis in these joints is relatively uncommon. Wrist pain can be treated by arthrodesis, and for localized intercarpal osteoarthritis limited intercarpal arthrodeses are frequently used. Total replacement arthroplasty is rarely used. Development of unconstrained elbow prosthetic replacements has improved the results of total elbow replacement arthroplasty, and fascial interposition arthroplasty too has been successful in younger patients. Manual laborers can be treated by arthrodesis of the elbow, but this is less commonly done than previously. Posttraumatic degenerative change in the acromioclavicular or sternoclavicular joints can be treated by resection of the joint. Unconstrained total shoulder replacement arthroplasty has provided acceptable results in osteoarthritic patients, and arthrodesis is also quite successful for shoulder arthritis.

Hip

Osteoarthritis of the hip can be primary or secondary to causes mentioned above, including developmental abnormalities of the femoral head or acetabulum. Patients may present with stiffness and pain in the inguinal area or referred pain in the knee, usually aggravated by weight bearing. Patients often have an abductor limp (leaning toward the affected side on weight bearing to decrease hip abduction force), and the affected extremity can exhibit shortening from both loss of joint space and hip flexion contracture. Internal and external rotation are limited and provoke pain, particularly with the hip in flexion.

Conservative treatment involves the measures described previously, and use of a cane on the contralateral side may decrease pain and limp dramatically. If conservative measures fail, surgical intervention often is appropriate. Surgical options include osteotomy, arthrodesis, and total hip replacement arthroplasty.

Arthrodesis

For young patients with severe osteoarthritis, the increased risk of long- term failure of total hip replacement militates against this approach. Arthrodesis of the hip provides excellent pain relief and function. Patients require spica cast immobilization for 3 to 6 months. Possible long-term problems include development of low back pain and knee pain from increased stresses on these joints. Preexisting degenerative change or symptoms in these joints contraindicate hip arthrodesis. Bilateral hip involvement also is a contraindication to fusion. Arthrodesis is useful in cases of pyogenic or tuberculous infection of the hip, in which the risk of reactivation after total hip replacement is significant.

Femoral Osteotomy

A number of osteotomies about the hip have been used for osteoarthritis, including medial displacement osteotomy, and varus or valgus osteotomies. These procedures have been more popular in Europe than in the United States, and significant pain relief has been reported in approximately 80 percent of patients. One disadvantage of osteotomies is that subsequent total hip replacement is made technically much more difficult if disease progresses despite osteotomy.

Total Hip Replacement Arthroplasty

Reconstructive procedures about the hip must be compared with the excellent results obtained by total hip replacement. Total hip replacement has been used for primary or secondary osteoarthritis, as well as rheumatoid arthritis. Long-term studies have indicated up to 91 percent implant survival at 15 years for conventional cemented arthroplasties. The major problem with total joint arthroplasty is related to loosening of the prosthetic components, usually at the bone-cement interface (Fig. 41- 108). Failure and revision rates as high as 20 to 25 percent at 5 years have been reported in some series. Failure rates are significantly higher in active patients under 40 years of age. A significant factor contributing to late loosening is wear debris of polyethylene and methyl methacrylate cement particles, which incite an inflammatory and macrophage response. Secondary release of local cytokines is thought to lead to bone resorption and consequent loosening. The predominant bone-resorptive cytokines identified in reactive membranes and fluid surrounding loose arthroplasties are IL-1, IL-6, and tumor necrosis factor–alpha. Perioperative or late infection poses the most serious complication of hip arthroplasty, often necessitating removal of the prosthetic components and cement in order to treat the problem definitively.

Uncemented Prosthetic Joint Replacement

Because of the problems associated with prosthetic loosening, efforts have been focused on the development of uncemented, porous-coated prosthetic designs that allow bony ingrowth to provide permanent stability. Designs include sintered metal beads (cobalt-chrome or titanium), sintered titanium wire mesh, and plasma-sprayed titanium bead coatings. Pore sizes of 50 to 400 mm have demonstrated optimum bony ingrowth experimentally, but in retrieval studies, significant areas of porous prostheses have been found to have only fibrous ingrowth. Micromotion greater than 150 mm appears to be associated with fibrous rather than bony ingrowth.

Another problem is dissociation of beads from the device, which may be a result of loosening. In an effort to enhance fixation, hydroxyapatite coatings have been introduced, allowing direct bony bonding. The use of recombinant growth factors also has been explored. Several clinical studies of uncemented hip replacements have been published, and while overall results are acceptable, mild persistent thigh pain and limp appear to be more common than with cemented arthroplasty (Fig. 41-109). Radiolucent lines are frequent around these prostheses, but revision rates have been low. Uncemented knee arthroplasty has been fairly successful on the femoral side, but problems with loosening have occurred with tibial and patellar components.

Infections

The most serious complication of total hip replacement is infection, which may occur immediately after operation or after several months or years. In addition, hematogenous seeding by bacteremia from other causes can result in infection of any total joint arthroplasty. Prophylactic antibiotics in patients who undergo invasive procedures such as dental work, colonoscopy, or cystoscopy is recommended. The use of clean air laminar flow operating rooms, prophylactic perioperative antibiotics (cephalosporin), and ultraviolet lights during surgery all have been associated with a drop in infection rates to less than 1 percent. Acute infection is treated by immediate debridement, suction/irrigation drainage for 24 to 72 h, and a full 6-week course of intravenous antibiotics. If infection recurs or is chronic, single-stage or two-stage exchange arthroplasty is undertaken. This involves removal of hardware and cement, debridement, and intravenous antibiotic therapy. With primary exchange the arthroplasty is revised using antibiotic-impregnated cement at the time of the debridement. More commonly, secondary exchange is carried out with revision after antibiotic treatment for 6 weeks (for staphylococcal or streptococcal infections) to a year (for gram-negative organisms).

In infected total knees or hips, the placement of antibiotic-impregnated cement beads in the joint at the time of primary debridement can enhance eradication of the infection. A cement spacer has been used in knee arthroplasties also to maintain the joint space until secondary revision is carried out. In some cases, loss of bone stock or presence of virulent organisms precludes revision surgery. In these instances resection arthroplasty (Girdlestone procedure), leaving a pseudarthrosis, provides satisfactory function and pain relief, although walking aids usually are needed permanently.

Heterotopic Ossification

Heterotopic ossification (myositis ossificans) is a frequent complication of total hip arthroplasty (reported incidence varies from 8 to 90 percent) and consists of abnormal bone and cartilage formation in the soft tissues adjacent to the joint. This can limit joint motion and progress to ankylosis in a small proportion of patients. Histologically, endochondral ossification and membranous bone formation are both present and resemble fracture callus. Risk factors include male gender, past history of having formed heterotopic bone, hypertrophic osteoarthritis, ankylosing spondylitis, and diffuse idiopathic skeletal hyperostosis. Heterotopic bone formation often is seen in head-injured and spinal cord–injured patients as well. Posttraumatic heterotopic ossification can follow muscle contusion and can be mistakenly diagnosed as an osteosarcoma.

Patients at risk for heterotopic ossification can be treated prophylactically with low-dose radiation (8 to 10 Gy) administered within the first 3 or 4 days after operation. This is thought to obliterate the cellular proliferative response. Indomethacin treatment for 6 weeks postoperatively also has been shown to significantly decrease the incidence of heterotopic bone formation. Ankylosis of joints secondary to heterotopic ossification can be treated by surgical excision after the ossification has matured, if postoperative radiation is used to prevent

recurrence.

Osteonecrosis of the Femoral Head (Avascular Necrosis)

Certain periarticular bony areas are prone to development of osteonecrosis because of their relatively precarious blood supply. The humeral head, femoral condyles, tibial plateau, talus, lunate, scaphoid, and femoral head are the most frequent areas of involvement. Osteonecrosis of the femoral head is the most common and serious of these.

Disruption of the blood supply can result from trauma (femoral neck fracture or hip dislocation), sickle cell anemia, abnormalities of fat metabolism (associated with alcoholism, lipid storage diseases, and corticosteroids), and decompression sickness ("the bends," secondary to nitrogen bubbles causing intraosseous vascular occlusion) and can also be idiopathic. The superior lateral quadrant of the femoral head is most often initially involved, and infarction causes marrow edema and venous outflow obstruction, progressively increasing local intramedullary pressure and widening the area of the infarction. The dead bone is gradually revascularized and replaced by creeping substitution, but during revascularization bone resorption can lead to mechanical failure, with subchondral fractures (crescent sign) indicative of impending joint surface collapse. The net result is flattening of the femoral head, with incongruity of the joint and, usually, rapid progression of secondary osteoarthritis.

Patients complain of hip pain, particularly with weight bearing or rotation of the hip; the infarction also may be clinically silent. When collapse occurs, however, patients almost always have pain and limitation of weight bearing and motion. A high proportion of cases with systemic causes (such as steroid use) have bilateral involvement. The progression may be rapid, over 1 to 2 months, or it may take years for flattening and collapse to cause symptoms. Steroid-induced osteonecrosis has been shown to progress to collapse in essentially all patients.

In patients at risk, the diagnosis should be suspected with a presentation of unexplained hip pain. Radiographs may be normal initially, but sclerotic changes soon develop in the femoral head, followed by subchondral lucency, flattening of the joint surface, and, later, secondary degenerative changes. Although bone scans can demonstrate decreased perfusion early in the evolution of osteonecrosis, MRI is the most sensitive method for detection of early necrosis when radiographs are normal.

Treatment

Treatment of osteonecrosis has been controversial and disappointing overall. Earlier literature supported the use of cortical strut bone grafting (Phemister, Enneking procedures) using tibial or fibular segments placed through the femoral neck to the subchondral bone to support the necrotic segment and prevent collapse. In steroid-induced avascular necrosis, however, a high proportion of patients progress despite this intervention. Ficat and Hungerford have popularized the procedure of core decompression, in which a channel is drilled through the femoral neck into the femoral head to decrease the elevated bone marrow pressure and facilitate healing. Prevention of progression in 90 percent of patients has been reported, but subsequent series have had failure rates as high as 80 percent, making this procedure very controversial. The procedure does provide good relief of rest pain, and, if done early, before there is any evidence of collapse or subchondral lucency, it appears to prevent progression in the majority of patients. Patients are kept on crutches with no weight bearing for 3 months after surgery. Urbaniak has proposed the use of a vascularized fibular graft with microvascular anastomosis to the femoral circumflex vessels to allow more rapid and reliable healing of the strut while mechanically supporting the necrotic segment. This procedure has given satisfactory results even in patients with subchondral lucency or slight collapse, although it is still too early to fully evaluate efficacy. Diagnosis and treatment of osteonecrosis are summarized in Fig. [41-110](#).

When secondary degenerative arthritis occurs, total hip replacement often is necessary. Arthrodesis has not proved very successful, with a high pseudarthrosis rate, presumably secondary to the bone necrosis. If collapse has occurred but no secondary degenerative changes are evident, bipolar endoprosthetic arthroplasty may be considered, although this is controversial. In general, results of total hip arthroplasty for osteonecrosis are not as good as those for primary osteoarthritis, with a higher loosening rate. In part this is a result of comorbidity caused by steroid use or underlying medical disease.

Knee Chondromalacia of the Patella

Patellar pain and mild degenerative changes are very common, particularly in young females. Chondromalacia refers to the early changes of degenerative arthritis, with softening and fibrillation of the articular cartilage. The medial facet is most often involved. Chondromalacia may be related to patellar subluxation, dislocation, or chondral contusion from a direct trauma. Patients complain of pain, especially with stair climbing and kneeling, which increase patellofemoral joint contact forces. Physical examination may reveal patellofemoral crepitus, pain with patellar compression, effusion, and tenderness of the patellar facets. Radiographs show patellar tilt or narrowing of the joint space, and articular cartilage degeneration and thinning is readily discernible on MRI. The evolution of chondromalacia to frank osteoarthritis is uncommon.

Treatment

Isometric quadriceps (straight-leg-raising) exercises and use of anti-inflammatory medications, along with avoidance of aggravating activities, often suffice to permit resolution of symptoms. A patellar sleeve-type brace also helps. In refractory cases with quadriceps malalignment, procedures to correct the alignment (as discussed earlier under Patellar Dislocation) can be helpful. Arthroscopic debridement of fibrillated cartilage is helpful in some cases, although the resultant degree of cartilage healing remains a subject of controversy. Arthroscopic debridement of either the patellofemoral or tibiofemoral joint does decrease inflammatory debris in the knee and usually improves symptoms, at least temporarily. In cases in which severe cartilage degeneration is present with exposed subchondral bone, abrasion arthroplasty of the bone surface to bleeding bone is recommended. As discussed previously (see Articular Cartilage Healing), growth of fibrocartilage in the defect then occurs, which allows resolution of symptoms for a number of years. Treatment of end-stage patellofemoral arthritis is with patellectomy.

Osteoarthritis of the Knee

Osteoarthritis of the knee is common in the aging population and can be associated with varus or valgus deformity, which accelerates the articular cartilage degeneration. Conservative therapy includes nonsteroidal anti-inflammatory medications; an elastic knee brace, which can decrease swelling and pain; and exercises to maintain range of motion and prevent contracture. Patients with varus and medial compartment degeneration (the more common pattern) may benefit from a lateral wedge in the shoe. In cases of unicompartmental (medial or lateral) involvement in the younger patient, osteotomy of the femur or tibia to effect realignment and decrease mechanical stress in the degenerated portion of the joint often is helpful. For varus deformities, valgus osteotomy of the proximal tibia is appropriate, although a significant number of patients progress later. With varus deformities, osteotomy of the distal femur is preferred, because osteotomy of the tibia can result in the joint line's not being parallel to the floor because of the normal valgus of the femur.

Total knee replacement arthroplasty is the treatment of choice for severely symptomatic older patients, with over 90 percent successful results at 5 to 10 years. The considerations are similar to those for total hip replacement arthroplasty, with the major problems being prosthetic loosening and infection. In patients with unicompartmental involvement, unicompartmental knee arthroplasty has been advocated because more bone stock is preserved compared to total knee replacement. Although results of unicompartmental replacement series vary, and loosening remains a significant problem, this is an option for the younger patient. Osteochondral allograft replacement for posttraumatic defects and arthritis about the knee also has been recommended by some, but these techniques remain controversial.

Ankle and Foot

Osteoarthritis in the ankle and foot most often results from trauma. Immobilization of the joint with an ankle lacer or brace and use of a rocker-bottom shoe sole is helpful in controlling pain. In severe cases, arthrodesis of the ankle may be successful. Similarly, subtalar arthritis can be handled by subtalar or triple arthrodesis if bracing is ineffective.

Hallux Valgus (Bunion)

Lateral deviation of the great toe (hallux valgus) is a disease of shoe-wearing populations. The incidence is higher in females. With progressive deviation of the toe caused by ill-fitting shoes, the lateral displacement of the flexor and extensor tendons contributes to progression of the deformity. Capsular traction on the MTP joint from the deformity and pressure medially from shoe wear cause exostosis, or osteophyte formation, on the medial metatarsal head, accentuating the deformity and causing pain. The articular cartilage responds to the mechanical incongruity of the joint by progressive degenerative change, and the stiffening of the joint aggravates pain on push-off. Initial treatment efforts are directed at more appropriate shoe wear with wide toe box shoes and insoles or metatarsal pads to alleviate metatarsalgia. In refractory cases, surgical intervention often is indicated.

Over 100 surgical procedures have been devised, suggesting the inadequacy of many of these operations in resolving the deformity and symptoms. One of the most frequent procedures is the Keller bunionectomy, which is a resection of the joint and osteophyte, resulting in a shortened but painless toe (Fig. [41-111 A,B](#)). Chevron osteotomy of the distal first metatarsal has been used in mild to moderate deformities in the absence of significant arthritic change, with satisfactory results (Fig. [41- 111 C](#)). If metatarsus primus varus exists, proximal osteotomy of the metatarsal base with or without a distal procedure may be helpful. In adolescents, soft-tissue procedures to release adduction contracture and tighten the medial capsule (Mitchell procedure) are favored.

Gout

Gout is a disease resulting from abnormalities in the metabolism of urate that cause deposition of urate crystals in joints, kidneys, and musculoskeletal soft tissues. Joint deposition is episodic and is associated with acute inflammation and pain. Deposits of urate crystals, called tophi, can occur in bone, cartilage, and synovium. The most common locations are the external ear, olecranon bursa, and around tendons of the distal extremities. Renal excretion can cause stones, and deposition in renal parenchyma can lead to renal damage. Secondary degenerative changes in joints can follow repeated inflammatory episodes.

Clinical Manifestations

The classic presentation is a middle-aged man with an acute, severe monarticular arthritis, usually involving the MTP joint of the great toe (podagra). Swelling, erythema, and severe pain with movement of the joint are typical. The diagnosis is confirmed by demonstration of urate crystals in the synovial fluid. The serum urate level also may be elevated. In chronic gouty arthritis, punched-out periarticular erosions due to tophi may be present radiographically, and later secondary degeneration, with joint space narrowing and subchondral sclerosis, is observed.

Treatment

Colchicine has been used for acute gout episodes since the days of Hippocrates. A dose of 0.65 mg is given every 1 to 2 h until acute symptoms subside; the effect is usually rapid and dramatic. For less acute symptoms, nonsteroidal anti-inflammatory medications such as indomethacin or naproxen are effective. Allopurinol, a xanthine oxidase inhibitor widely used prophylactically to lower the serum urate level, is helpful in preventing recurrent attacks and reducing tophaceous deposits. In severe chronic secondary degenerative arthritis, treatment as for osteoarthritis, including joint replacement arthroplasties, is appropriate.

Calcium Pyrophosphate Deposition Disease (Chondrocalcinosis)

Like gout, calcium pyrophosphate deposition disease occurs more commonly in males than in females, and usually in middle-aged patients. Acute inflammatory episodes, termed

pseudogout because of their resemblance to gout, are associated with this disease. The knee joint is most commonly involved, and radiographs will demonstrate the calcification of articular or meniscal cartilage. The diagnosis is confirmed by synovial fluid analysis under polarized light, which will demonstrate the weakly positively birefringent rhomboidal crystals of calcium pyrophosphate. The distal radioulnar joint, pubic symphysis, and acetabular or glenoid labrum can also be involved. Treatment is with aspiration and cortisone injection, or with systemic anti-inflammatory therapy in milder cases.

Hemophilic Arthritis

In patients with hemophilia acute hemorrhage in joints exposed to minor trauma leads to joint stiffness and articular cartilage degeneration, particularly with repeated episodes. Spontaneous bleeding occurs only when factor VIII levels are 0, and bleeding with minor injury occurs with levels of 1 to 5 percent. The synovium exhibits villous proliferation with marked brownish staining from hemosiderin. The most commonly involved joint is the knee, followed by the elbow, ankle, and shoulder. The joint assumes a position of maximal volume (knee—flexion of 20 degrees; hip—flexion, adduction, external rotation), and the patient will not move the joint because of pain and muscle spasm. With progressive episodes, joint contractures and eventual destruction of the articular cartilage occur. The characteristic radiographic findings include enlargement of epiphyses in children, and a squared-off appearance of the inferior patella and femoral condyles (Fig. 41-112). An additional concern in hemophiliac patients is HIV infection, which has become the leading cause of death in this group. Consequently, infectious processes such as tuberculosis or pyogenic arthritis may have to be considered in the differential diagnosis.

Treatment

Minimizing joint hemorrhage decreases the likelihood of chronic arthritis. Patients should be treated with factor VIII, splinting, and compression wrapping of the joint. Weight-bearing joints are protected on crutches until pain and effusion resolve and motion returns. Bracing may be needed to prevent deformity, and dynamic splints may help to improve contractures. In severe cases of joint destruction, total joint replacement arthroplasty is now feasible, with appropriate coagulation factor replacement to 100 percent levels in the perioperative period.

Synovial Lesions

Pigmented Villonodular Synovitis

Pigmented villonodular synovitis is an inflammatory synovial process of unknown cause that causes monarticular arthritis in children and young adults. Patients present with pain and intermittent, often chronic, swelling. Examination shows swelling and synovial thickening, and aspiration of the joint reveals bloody or brownish hemosiderin-stained fluid. Cytologic examination of the fluid may reveal the presence of hemosiderin-containing macrophages. Early in the course of the disease the radiographs are normal except for soft-tissue swelling or effusion. Later, periarticular erosions and cyst formation occur, and ultimately degenerative change of the articular surfaces with the radiographic features of osteoarthritis supervene.

Histopathologic inspection of the synovium reveals nodules and villous projections, brownish in color, containing fibrous tissue, giant cells, and monocytic cells with hemosiderin granules and hyperplasia of the synovial layer. Both nodular and diffuse forms have been described, with the diffuse form having a higher recurrence rate (approximately 50 percent) than the nodular form (25 percent) after synovectomy.

Treatment of the lesion involves synovectomy. Radiation synovectomy using dysprosium-165 has been reported to give satisfactory results, as has arthroscopic synovectomy of the knee. In joints such as the hip, or in the presence of large masses of the synovial tissue, open surgical synovectomy usually is preferable.

Synovial Chondromatosis

Synovial chondromatosis is a neoplastic-like condition of synovium in which the synovial tissue undergoes chondroid metaplasia, forming nodules of cartilage and bone that can detach and become loose bodies in the joint. The most commonly involved joints are the hip, knee, and shoulder. Three phases have been described, with an initial proliferative phase, a secondary phase of loose body formation, and an inactive phase in which mechanical damage to the joint leads to progressive secondary degenerative arthritis. Patients present with mild pain, effusion, and episodes of locking. Radiographs usually demonstrate the loose bodies, which often are calcified. MRI also is diagnostic and can demonstrate the lesions before calcification of the cartilage. Surgical synovectomy is the treatment of choice.

Ankylosing Spondylitis

Ankylosing spondylitis is one of the group of seronegative spondyloarthropathies with unknown causes. Other members of this group include psoriatic arthritis, Reiter's syndrome, and enteropathic arthritis. Characteristics include sacroiliac and spinal involvement and a higher than normal incidence of human leukocyte antigen (HLA) B27 positively. The incidence of ankylosing spondylitis is 1:1,000, and there is a male predominance. Progressive back pain, stiffness of the spine progressing to ankylosis (bamboo spine), morning stiffness, associated hip arthritis, and age of onset under 40 are common features.

Treatment consists of nonsteroidal anti-inflammatory agents and physical therapy to maintain range of motion. If severe spinal deformity (most commonly severe flexion deformity of the cervical spine) develops, corrective osteotomy occasionally is necessary. Total hip arthroplasty is helpful in cases with severe arthritis of the hips, and care must be taken to consider prophylactic treatment for heterotopic ossification in these patients using indomethacin or postoperative low-dose radiation treatment.

Transient Synovitis of the Hip in Children

Children 3 to 10 years of age may present with sudden onset of hip pain and an inability to walk that can mimic symptoms of septic arthritis. A limp may be noted for several days, and often there is a history of an antecedent viral illness. Radiographs usually are normal, and ultrasonographic examination will demonstrate effusion in the joint. Aspiration is mandatory for cultures and Gram stain to rule out septic arthritis. The white cell count in the aspirate is elevated, in some cases over 50,000/mm³. The erythrocyte sedimentation rate and peripheral white cell count also may be slightly elevated. If cultures are negative, a brief period of bed rest with light traction will allow spontaneous resolution. A small proportion of these children (2 percent) later develop Legg-Calvé-Perthes disease, and some develop recurrences of transient synovitis, but long-term problems generally do not occur.

Slipped Capital Femoral Epiphysis

Slipped capital femoral epiphysis, which occurs in children (girls aged 10 to 13 years; boys aged 12 to 15 years), consists of a displacement of the capital femoral epiphysis posteriorly and medially with respect to the femoral neck. Boys are affected more frequently (ratio 5:1). The slippage can occur acutely or gradually, and the incidence of bilaterality is 15 to 25 percent. Patients with endomorphic body habitus tend to be affected.

Pathology

Unlike in Salter type I and type II fractures, in which the epiphysis separates through the zone of provisional calcification of the growth plate, in this disorder the separation occurs through the hypertrophic zone above the calcified cartilage. The cause is unknown, and mechanical, endocrine, immunologic, and genetic causes have been proposed. Histologically, the hypertrophic zone of the growth plate is disorganized, and similar disorganization has been identified in preslipped growth plates. Much of the evidence suggests an abnormality in the collagen of the growth plate as the underlying problem.

Clinical Manifestations

The most common presentation is pain in the inguinal area or referred pain in the knee. The patient may have an antalgic gait, and physical examination reveals pain on motion of the joint with loss of internal rotation. With flexion of the hip the thigh externally rotates.

Diagnosis

Diagnosis is made radiographically with biplanar films. Both hips should be included because of the incidence of bilaterality. Preslip radiographic findings include widening of the growth plate. A minimal slip presents with only subtle radiographic findings, but a line projected along the superior femoral neck on the anteroposterior film usually will not intersect the epiphysis. With an acute slip, a tube lateral view is safer than a frog-leg view, which will displace the epiphysis further. Chronic slips exhibit some metaphyseal remodeling, with new bone formation inferior and posterior to the junction of the head and neck.

Treatment

Treatment is aimed at preventing further displacement of the epiphysis. Mild to moderate acute slips are treated by in situ pinning with threaded pins or screws across the physis and into the femoral head (Fig. 41-113). Great care must be taken to ensure that penetration into the joint does not occur, as this can lead to chondrolysis and a poor outcome. It has been demonstrated that a single screw is effective if placed centrally in the epiphysis. Weight bearing is protected until fusion of the growth plate occurs, usually within several months. Severely displaced acute slips (less than 1 to 2 weeks old) can be reduced by gentle traction and then pinned. Manipulative reduction is controversial, because it is associated with an increased incidence of chondrolysis and avascular necrosis. For severe chronic slips, osteotomy of the proximal femur and cuneiform osteotomy of the femoral neck have been recommended.

The complication of chondrolysis occurs more frequently in blacks and females with a slipped epiphysis and leads to rapidly degenerative arthritis. Patients with higher grades of displacement are prone to development of secondary osteoarthritis in the fifth to sixth decades.

Hypertrophic Pulmonary Osteoarthropathy

Hypertrophic pulmonary osteoarthropathy is associated with pulmonary diseases and consists of excessive bone formation adjacent to joints, clubbing of the distal digits, and periosteal new bone formation along shafts of the long bones. Arthritic changes resembling osteoarthritis can develop, and treatment is symptomatic. Occasionally resolution of the pulmonary pathology (as occurs after excision of a neoplasm) results in regression of the symptoms.

Neuropathic (Charcot) Arthropathy

A neuropathic, or Charcot, joint is a consequence of diseases that result in denervation or loss of proprioceptive sense in joints. Etiologic conditions include diabetes mellitus, syphilis (tabes dorsalis), lepromatous neuropathy, paraplegia, and syringomyelia.

Affected joints are characterized by progressive, severe, and total destruction, often following minor trauma. Fragmentation of joint surfaces and marked deformity can occur (Fig. 41-114). Massive swelling is present, and joint volumes are large, which is a helpful point in differentiating these conditions from chronic septic processes, in which the capsule is scarred and fibrotic and joint fluid volume small. These changes may be attended by gross ligamentous instability. The foot and ankle commonly are involved, although the knee, shoulder, and, rarely, the hip may be affected. Pathologically, synovial biopsy analysis shows moderate hypertrophy, minimal inflammatory infiltration, and fragments of articular cartilage or subchondral bone embedded in the synovial membrane, which are pathognomonic for Charcot arthropathy.

Bracing for control of instability and pain is the mainstay of treatment, and total joint arthroplasty is relatively contraindicated. Arthrodesis is helpful in symptomatic joints when conservative therapy fails, but it is difficult to achieve. With foot and ankle arthropathy in diabetics, below-knee amputation is a common long-term outcome.

TUMORS OF THE MUSCULOSKELETAL SYSTEM

General Considerations

Primary tumors of the musculoskeletal system are rare. The histogenetic type of the tumor is dependent on the tissue of origin. In addition, specific types of lesions tend to occur in particular bones or areas of bones, usually in areas of maximal growth or remodeling. Giant cell tumors (osteoclastomas) occur near the growth plate, where a high level of resorption takes place as part of remodeling; osteosarcomas occur in the metaphysis, where new bone formation is maximal; cartilage tumors involve the metaphysis near the growth plate, and round cell tumors occur in the metaphyseal/diaphyseal bone marrow (Fig. [41-115](#)). These neoplastic processes are thought to represent a derangement of normal growth and bone remodeling functions, which become uncontrolled.

Etiology

The cause of most bone and soft-tissue neoplasms is unknown, but recent molecular biology studies promise to elucidate these mechanisms, with identification of specific genetic mutations and chromosomal aberrations in some tumors, derangements of tumor suppressor gene function, and expression of oncogenes. Most histogenetic tumor types have variable levels of aggressiveness and occur in benign and malignant forms. Table [41-16](#) illustrates the range of incidence and histogenesis of musculoskeletal neoplasms. The biologic behavior of tumors can vary, as reflected by the pathologic grade of the tumor. While various grading systems exist for different tumors, in general the simplified system adopted by the Musculoskeletal Tumor Society reflects overall gross behavior differences, with benign, low-grade malignant, and high-grade malignant forms.

Characteristics

Musculoskeletal neoplasms are characterized by initial centrifugal growth from a single focus, pseudoencapsulation (formation of a zone of reactive tissue around the expanding lesion, which in malignant lesions can be focally invaded by the tumor), and a tendency to respect anatomic boundaries early in the evolution of the lesion. These tumors thus tend to spread along fascial planes and tend to remain contained in anatomic compartments, a crucial characteristic in strategies for staging and surgical treatment of these lesions. Anatomic compartments include bones, muscle compartments, joints, skin and subcutaneous tissue, and in some cases, major neurovascular sheaths.

Metastasis

Metastasis of malignant musculoskeletal neoplasms is associated with a poor prognosis. Metastases are most often pulmonary, although some tumors tend also to involve regional lymph nodes, and bony metastases also occur. Brain and visceral metastases are unusual, generally occurring only in terminal end-stage disseminated disease.

Staging

The most widely used staging system for musculoskeletal neoplasms, shown in Table [41-17](#), has been applied to both soft-tissue and bone lesions. Benign lesions are graded as latent, active, or aggressive. Malignant lesions are staged on the basis of whether they are high grade (stage II) or low grade (stage I), and intracompartmental (A) or extracompartmental (B). Metastatic tumors are all stage III regardless of local extent and have a dismal prognosis. This staging system has shown great value in predicting survival (Fig. [41-116](#)).

Clinical Manifestations

Patients typically present with a history of pain that is often worse at night and usually is not activity related. A mass or swelling may be present, but constitutional symptoms (weight loss, fevers, night sweats, malaise) usually are absent, except in cases with disseminated disease. Lesions adjacent to joints can cause effusion, contractures, and pain with motion. Soft-tissue tumors often are painless unless there is involvement of neurovascular structures. Compression of veins or lymphatics in a limb can cause distal edema, and larger masses exhibit a pattern of overlying venous distention. Malignant soft-tissue masses can be firm and fixed to subcutaneous tissue, muscle, or bone, and usually are nontender. Local warmth is evident because malignant lesions induce local angiogenesis. Patients may also present with a pathologic fracture as a manifestation of benign or malignant intraosseous lesions, with bone destruction and subsequent mechanical failure. Pain on weight bearing is an ominous clinical symptom that often indicates an impending fracture.

Evaluation should include a thorough history and physical examination of the affected region, with attention to joint, muscle, neurologic, and vascular structures. Examination of regional and distant lymph nodes is essential, as are pulmonary and abdominal examinations to assess the possibility of metastatic disease.

Radiographic Findings

The plain radiograph is the single most useful study in differential diagnosis of bone lesions. Considerations include the following:

- (1) Evidence of matrix production (bone formation, calcification)
- (2) Pattern of growth (permeative, geographic, moth-eaten, loculated, expansile, exophytic)
- (3) Presence of bony reaction to the lesion (periosteal reaction, sclerotic margination)
- (4) Zone of transition between the host bone and lesion (narrow or well- marginated versus wide or poorly defined)
- (5) Age of the patient
- (6) Bone involved (flat bone, long bone, skull, vertebrae, acral bone)
- (7) Location within the bone (epiphyseal, metaphyseal, diaphyseal)
- (8) Associated soft-tissue mass, clinical symptoms
- (9) Presence of solitary versus multiple lesions

Using these criteria, an accurate differential diagnosis can be formulated in most cases. Infection (osteomyelitis) must always be considered given its highly variable radiographic appearance. Metabolic, inflammatory, dysplastic, traumatic, congenital, and degenerative conditions also are always considered. Soft-tissue lesions are better evaluated by MRI than any other type of radiographic study.

Diagnostic Evaluation

Routine laboratory studies include complete blood count and differential; erythrocyte sedimentation rate; serum alkaline phosphatase, calcium, and phosphate levels, renal and liver function studies, and urinalysis. If multiple myeloma is within the differential diagnostic possibilities, determination of serum protein level or immunoelectrophoresis should also be performed. In most instances of primary tumors the majority of these studies are normal. The alkaline phosphatase level may be elevated in osteosarcoma, and the blood count and erythrocyte sedimentation rate are helpful in excluding infection. Further staging studies vary according to the location of the lesion, diagnostic possibilities, age of the patient, and likelihood of malignancy. A bone scan almost always is indicated to assess the activity and extent of the primary lesion as well as to exclude the presence of other lesions. With soft-tissue tumors bone scan is reserved for lesions close to bone or suspected of malignancy. For suspected malignant lesions, other recommended studies include chest radiograph (or preferably CT scan) and abdominal CT scan to exclude metastatic disease. A diagnostic staging algorithm is shown in Fig. [41-117](#).

Biopsy

For lesions with a radiographically benign appearance, imaging studies of the lesion usually are unnecessary, and the appropriate next step is tissue diagnosis by biopsy. For any potentially malignant lesion, three-dimensional imaging studies (CT or, preferably, MRI) before biopsy are recommended to fully assess the extent of the lesion and to plan the biopsy procedure, minimizing potential contamination of compartments, which could compromise subsequent definitive surgery. Depending on the experience of the surgeon and pathologist, needle or trocar biopsy is appropriate for the majority of soft-tissue and bone tumors. General principles of the biopsy procedure include the following:

- (1) Biopsy incisions should always be longitudinal on extremities.
- (2) Needle biopsy tracts and incisional biopsy should be placed so that they can be excised en bloc at the time of resection.
- (3) Radiographic localization should be done to ensure accuracy.
- (4) Frozen-section examination should be done to be sure that adequate tissue has been obtained.
- (5) Cultures and appropriate microbiologic studies should be performed.
- (6) The bone biopsy cortical window should be as small as possible and oval in shape to minimize the risk of pathologic fracture.
- (7) Central or necrotic areas should be avoided; biopsy at the periphery of the lesion is most helpful.
- (8) Exposure of any major neurovascular structures should be avoided.

(9) Hemostasis must be obtained to prevent hematoma, which could seed other compartments; for bone lesions suspected of malignancy, the biopsy site should be plugged with methacrylate cement to prevent hematoma.

(10) Tourniquet use is helpful for intraoperative accuracy of dissection.

(11) Use of a drain with its tract in line with the biopsy incision and near it will facilitate later en bloc resection.

(12) Contamination of any uninvolved compartment must be avoided.

(13) In general, the surgeon providing definitive treatment should also perform the biopsy whenever possible; this would usually involve a tertiary care referral center.

Treatment

In the treatment of benign and nonmetastatic malignant musculoskeletal tumors the primary goal is eradication of the disease; preservation of limb function is an important but secondary consideration. Long-term results have improved dramatically in the past two decades, and the treatment approach for malignant lesions has changed, with a shift away from amputations and toward limb salvage procedures. The specific treatment varies with the lesion but usually includes a combination of several modalities: surgery, chemotherapy, and radiotherapy. Benign lesions usually are treated surgically. For malignant tumors the primary treatment usually is surgery, with chemotherapy or radiotherapy as a secondary (adjuvant) treatment. Commonly used chemotherapeutic agents include doxorubicin, methotrexate, cyclophosphamide, ifosfamide, vincristine, and actinomycin D. Radiation treatment may be given preoperatively, postoperatively, or by implantation of catheters at operation followed by postoperative loading with short-range isotopes (brachytherapy). Effective doses for control of microscopic disease are generally in the range of 50 to 65 Gy.

Surgical Procedures

Surgical procedures used in treatment of tumors are defined as follows:

(1) Intralesional: leaves microscopic and macroscopic residual, as in curettage of a benign lesion.

(2) Marginal: removal through the reactive zone of the tumor; may leave microscopic residual in malignant tumors.

(3) Wide: removal with some normal tissue beyond reactive zone in all directions.

(4) Radical: complete removal of all compartments (bone, muscle, joint) involved with the tumor or its reactive zone.

These surgical procedure definitions are summarized in Fig. 41-118. For benign stage I or stage II lesions, intralesional or marginal excision is adequate, while stage III aggressive lesions require marginal to wide resection for cure. Low-grade (stage I) malignant tumors can be treated with wide surgical resection, with a high probability of local control. High-grade (stage II) malignant lesions can be treated by radical surgical excision, or wide excision plus adjuvant treatment, with comparable low recurrence rates (5 to 10 percent). In the treatment of tumors with similar margins the results of amputation versus resection are also comparable, forming the basis for the predominance of limb salvage surgery in recent years.

Technological advances also have contributed to this change in treatment approach, given the availability of custom computer-designed prosthetic implants, which can replace all or part of a bone or joint; osteochondral allografts; new limb lengthening techniques; and microvascular techniques for free tissue transfers of bone and soft tissue. Amputation may still be necessary for tumors in which involvement of major neurovascular structures or multiple compartments precludes resection with preservation of useful limb function. Specific treatment and adjuvant therapies vary according to the histogenetic tumor type and grade, and are summarized in Table 41-18.

Prognosis

Malignant musculoskeletal tumors remain serious and life-threatening diseases, although the prognosis has improved significantly over the past two decades. For stage II lesions 5-year survival rates range from 40 to 80 percent, while for stage I lesions 5-year survival rates are in the 70 to 90 percent range. Local control rates of 90 percent or better can be anticipated in the majority of tumor types. Local recurrence of benign lesions varies with stage and tumor type.

Given that the major problem in the treatment of skeletal malignancies remains late metastatic disease, further scientific advantages in this area are needed. Recent research has identified multiple drug resistance gene expression in patients treated with chemotherapy whose tumors become resistant to the drugs. These genes lead to the production of an ATPase (P-glycoprotein) that pumps a wide variety of drugs out of the tumor cells, maintaining sublethal intracellular levels. Other mechanisms of drug and radiation resistance also exist. Experimental pharmacologic approaches to the enhancement of chemotherapeutic effectiveness by inhibiting these resistance mechanisms are currently under study and show promise for improving outcomes of sarcoma treatment. Other experimental methods of metastatic sarcoma treatment under investigation include the use of immunotherapy, in which the patient's immune system is sensitized to tumor antigens, and bone marrow transplantation, in which high-intensity chemotherapy is followed by the reintroduction of autologous marrow obtained in advance.

Specific Musculoskeletal Tumors

Bone-Forming Tumors Osteoma

This small, sessile benign body tumor occurs most often in the skull and neither causes symptoms nor requires treatment. It consists of an abnormal excrescence of surface bone. Similar lesions occur posttraumatically on the femur in the area of the adductor magnus insertion (rider's bone), or in relation to the medial collateral ligament of the knee (Pellegrini-Stieda lesion).

Osteoid Osteoma

This benign bone-forming lesion primarily affects patients under 30 years of age and has a male preponderance. Patients present with local pain, which can be quite severe and is often relieved by aspirin. Radiographically, a small (less than 1 cm) lucent lesion (nidus) is seen, typically surrounded by marked reactive sclerosis (Fig. 41-119). Sometimes areas of radiodensity are seen within the lucent lesion, corresponding histologically to disorganized woven bone formation. The lesion gradually regresses over a period of 5 to 10 years, but most patients are unable to tolerate the symptoms and opt for surgical resection of the lesion, which usually is curative if the entire nidus is removed.

Osteoblastoma

Osteoblastoma is a benign bone-forming tumor affecting primarily children and young adults. Any bone may be involved, but the spine, particularly its posterior elements, is most often affected. The lesions are expansile and have a mixed lytic and blastic radiographic appearance. Patients usually present with pain, and treatment involves marginal resection or curettage and bone grafting if resection is not feasible without excessive morbidity. Histologically, vascular stroma, woven bone formation, giant cells, and osteoid may be present, and differentiation from osteoid osteoma is based on clinical and radiographic criteria (i.e., a lesion greater than 1 cm in diameter, without reactive cortical sclerosis) rather than histology. Cellular lesions can easily be confused with osteosarcoma, and careful evaluation by a qualified pathologist is essential to avoid misdiagnosis.

Osteosarcoma (Osteogenic Sarcoma)

Osteosarcoma is the most common primary bone malignancy apart from multiple myeloma, although it is nonetheless a rare disease (incidence 2.8/1,000,000). Patients 10 to 25 years of age are most often affected, and the most common sites are areas of maximal bone growth (distal femur—52 percent; proximal tibia—20 percent; proximal humerus—9 percent). Usually the lesions are metaphyseal. Although any bone can be involved, the disease seldom occurs in the small bones of the distal extremities and in the spine. This disease has a number of variants: (1) "classic" central or medullary high-grade osteosarcoma; (2) periosteal osteosarcoma; (3) parosteal osteosarcoma; (4) osteosarcoma secondary to malignant degeneration of Paget's disease, fibrous dysplasia, or radiation; and (5) telangiectatic osteosarcoma.

Osteosarcoma exhibits a blastic radiographic appearance in most cases because of the neoplastic woven bone formation. The periosteum may be raised off the bone by the tumor mass, causing a fusiform swelling with reactive periosteal bone at the periosteal margins (Codman's triangle). The malignant bone formation may have a sunburst appearance (Fig. 41-120), with invasion into adjacent compartments. Pathologic fractures can occur but are unusual.

Histologically, the tumor consists of small pleomorphic spindle cells, with osteoid and woven bone formation, and there is often cartilage formation as well. Cartilage formation is a prominent feature of periosteal and parosteal variants of osteosarcoma. Telangiectatic variants are lytic and expansile, resembling an aneurysmal bone cyst, and have prominent vascular spaces and relatively sparse bone formation.

Patients present with pain that often is nocturnal and a mass or swelling. Metastatic spread usually is pulmonary, and evaluation of the chest by CT is necessary. Serum alkaline phosphatase levels may be markedly elevated, but laboratory studies are otherwise usually negative. Evaluation also should include bone scan to rule out bone metastases, and CT or, preferably, MRI of the region for surgical planning. Most osteosarcomas present as stage IIB lesions.

Osteosarcoma is not particularly sensitive to radiation, but it does typically respond well to combination chemotherapy. Depending on the extent and location of the lesion, treatment typically involves wide surgical resection or amputation, usually after preoperative (neoadjuvant) chemotherapy. Bone resected in limb salvage operations can be reconstructed by custom prosthetic replacement, arthrodesis, or allografting (Fig. 41-121). Results of combination chemotherapy with resection are better than even radical surgical amputation without adjuvant therapy, with 50 to 70 percent 5-year survival rates and usually better than 90 percent local control (compared with 20 percent 5-year survival rates with radical surgery alone). Pathologic fracture, with contamination of all compartments, can preclude limb salvage surgery. Chemotherapy is continued after surgery for 1 year. Prosthetic designs that can be periodically lengthened by a minor surgical procedure allow limb salvage even in relatively young children with osteosarcoma, in whom progressive limb length discrepancy might otherwise be a severe problem. Preoperative intraarterial chemotherapy and radiotherapy have been used instead of neoadjuvant systemic chemotherapy, and results appear to be comparable.

Parosteal Osteosarcoma

These tumors occur in a slightly older age group and start adjacent to the periosteum rather than in the bone. The posterior aspect of the femur and proximal humerus and tibia are the most frequent sites. The tumor tends to be well circumscribed and slow-growing and metastasizes only late, usually to the lungs. Histologically, bland spindle cells with woven

bone formation, fibrous stroma, and focal cartilage formation are the typical characteristics. These lesions are not sensitive to adjuvant treatments and are treated by wide surgical resection and appropriate reconstruction, or by amputation. Prognosis is significantly better than for conventional high- grade osteosarcoma.

Secondary Osteosarcoma

In older patients, osteosarcoma can arise secondary to a chronic predisposing condition. The most common of these is Paget's disease (malignant degeneration reported in 1 to 10 percent of cases), but osteosarcoma has also been reported in fibrous dysplasia, and rarely with chronic osteomyelitis. As many as 5 to 10 percent of patients subjected to high-intensity radiation therapy for other cancers (sarcoma, lymphoma, etc.) may develop secondary sarcomas 10 to 20 years later, of which one type is osteosarcoma. Secondary osteosarcomas invariably are high-grade aggressive tumors, and prognosis is poor, with a tendency for early metastasis. Most patients in this age group are unable to tolerate the toxicity of intensive chemotherapy, and the usual treatment consists of surgical resection or amputation.

Cartilaginous Tumors Osteochondroma (Exostosis)

This lesion is a common exophytic benign lesion that occurs during childhood, usually in the metaphyses of the long bones (Fig. 41-122). It is thought to result from an aberrant fragment of the growth plate that is left behind and undergoes spontaneous growth. The lesions have a bony base with a cartilaginous cap, from which the growth occurs as it does in normal growth plates during childhood. A multiple hereditary form occurs and was discussed earlier (see Multiple Exostoses under Developmental Disorders). The lesions may cause pain from impingement on tendons, nerves, or muscle and frequently require surgical excision. Growth of the lesion, while not of concern in children, may indicate malignant transformation in adults. A cartilaginous cap thickness of more than 1 cm (assessed by CT or MRI) should arouse suspicion of malignancy. In solitary lesions the risk of malignant degeneration is less than 1 percent, while in multiple lesions it may be as high as 15 percent. Marginal to wide excision of benign or malignant lesions usually is curative if all the cartilage is removed.

Enchondroma

Enchondromas are intramedullary cartilage lesions, often exhibiting calcification and expansion of the bone. The small bones of the hands and feet are commonly involved, but long bone involvement also occurs (Fig. 41-123). The disease occurs in solitary and multiple forms. Patients may present with pain or pathologic fracture. The usual treatment is intralesional resection (curettage) and bone grafting. The most serious concern is the possibility of malignant degeneration, and careful sampling at the time of biopsy is necessary to exclude the possibility of chondrosarcoma.

Parosteal Chondroma

This is a rare benign cartilage lesion arising subperiosteally, often in the humerus or small bones of the hand or foot. The lesions are somewhat more aggressive than enchondromas and are prone to local recurrence. Accordingly, marginal or wide resection and bone grafting is indicated and is curative in the majority of cases.

Chondroblastoma

This is one of the few epiphyseal tumors and occurs most often in the first and second decades of life, when the growth plate is still open. Patients present with pain, joint effusions, or contractures, and radiographs show a lytic lesion with calcifications in the epiphysis (Fig. 41-124). The lesion is composed of chondroblasts, cartilage, giant cells, and vascular stroma. Treatment is by curettage and bone grafting and is often challenging because of the intraarticular location of the lesions. A rare malignant epiphyseal cartilage tumor in older adults, clear cell chondrosarcoma, probably represents the malignant degenerative counterpart of chondroblastoma.

Chondromyxoid Fibroma

This rare metaphyseal tumor affects children and young adults, typically arising in the femur or tibia. The lesion is benign, but it exhibits aggressive local behavior, with a high propensity for local recurrence and spread. The radiographic appearance is primarily geographic and lytic with occasional calcifications; an expansile or multilocular appearance with a relatively well- defined zone of transition between the tumor and host bone also is found. Histologic examination reveals a lobular configuration with three components: cellular fibroblastic areas, chondroid areas, and myxoid areas with typical stellate tumor cells. Treatment is by wide or marginal resection and bone graft reconstruction, although aggressive curettage with grafting also has been associated with satisfactory results.

Chondrosarcoma

Chondrosarcoma can be primary or secondary (as discussed above) and affects a broad age range (age 20 to 60 years). The pelvis, femur, tibia, and other long bones can be involved, and lesions closer to the axial skeleton are more likely to be malignant. Intramedullary calcifications are usually evident. Differential diagnosis includes bone infarction and enchondroma. Features of cortical destruction and pain are important indicators of possible malignancy (Fig. 41-125). The tumors are graded as low, intermediate, or high grade of malignancy on the basis of cytologic features and presence of matrix production. Lower-grade lesions can be treated by wide resection, but with high-grade lesions metastatic disease is frequent and the prognosis is poor. Limb salvage surgery often is feasible, but adjuvant treatments are not particularly helpful since these lesions tend to be resistant to chemotherapy and radiotherapy.

Fibrous Lesions Fibroma

Small intracortical fibrous lesions, referred to as fibrous cortical defects, are common incidental radiographic findings in the long bone metaphyses of children and tend to disappear spontaneously at skeletal maturity. Larger variants, which can progressively enlarge into the medullary cavity and occasionally cause pathologic fractures, are referred to as *nonossifying fibromas*. The tumor consists of bland fibroblastic and histiocytic cells, with osteoclasts and cholesterol clefts from lipid-laden macrophages. A variant of this tumor that ossifies occurs in the mandible (ossifying fibroma). In larger or symptomatic lesions, curettage and bone grafting is indicated, and recurrences after this treatment are uncommon.

Desmoid

This is a rare aggressive fibrous tumor of bone that is analogous to its soft-tissue counterpart, aggressive fibromatosis. Marginal to wide resection is indicated rather than curettage because of the tendency for local recurrence. With aggressive fibromatosis of soft tissues, local invasiveness causes frequent and progressively problematic recurrences after surgical treatment. The lesions do not metastasize and have been treated with wide surgical resection or radiation treatment, with local control rates of approximately 50 percent. Significantly better results have been obtained by marginal to wide local resection in conjunction with moderate-dose (45 to 55 Gy) radiation therapy. Systemic therapy with methotrexate also has been reported to control or cause regression of aggressive fibromatosis.

Fibrosarcoma

Primary fibrosarcoma of bone is rare and is characterized by a geographic lytic radiographic appearance with cortical destruction and associated soft- tissue mass (Fig. 41-126). Some of these lesions are better classified as malignant fibrous histiocytomas, with a mixed cell population. The tumors are moderately radiosensitive, and adjuvant chemotherapy can be effective in improving survival rates. Surgery consists of wide or radical resection and reconstruction rather than amputation, often in conjunction with adjuvant radiation treatment. These tumors also arise as secondary lesions in fibrous dysplasia and Paget's disease and after radiation treatment for other cancers.

Cystic Lesions Unicameral (Solitary) Bone Cyst

This lesion occurs in children in the metaphysis of the long bones adjacent to the growth plate, most often the humerus or femur, although the radius, calcaneus, and tibia also can be affected. Usually the lesions are painless and may present with a pathologic fracture as the initial manifestation of the disease. The lesions are lytic, expansile, and well marginated (Fig. 41-127), and may be found in the diaphysis in older children as a result of continued growth of the growth plate away from the lesion. In young children fractures heal, but the lesions usually recur, causing recurrent fractures during childhood. The cyst fluid contains high levels of bone resorptive cytokines, presumably produced by the living tissue and accounting for the aggressive bone resorption in these lesions. At skeletal maturity the cysts tend gradually to disappear. In older children and young adults, the lesions become latent (stage I) and do not progress. Recurrence rates in active (stage II) lesions in younger children after surgical treatment (curettage and bone grafting) average 50 percent. Partial or complete healing of the majority of these lesions has been obtained after intraosseous injection of methylprednisolone, currently the preferred treatment (70 to 90 percent effective with up to three sequential injections). In older children or adults with latent cysts, curettage and bone grafting is effective, and steroid injections appear to have little effect.

Aneurysmal Bone Cyst

This tumor, found most often in children or young adults, consists of a cystic lesion with large vascular spaces, characterized by aggressive, expansile lysis of bone. The tumor is composed of fibrous tissue, vascular spaces with a lining resembling endothelium, giant cells, and reactive bone formation at the periphery. Aneurysmal cysts can arise as a secondary degenerative vascular lesion within another primary benign or malignant bone tumor, such as giant cell tumor or chondroblastoma; however, about half are thought to represent primary lesions. Because recurrence is relatively frequent with simple curettage, local resection with bone grafting is preferable. Embolization has been used successfully in unresectable spinal or pelvic lesions, as has intermediate-dose radiation treatment. Preoperative embolization of large lesions is helpful in decreasing the risk of hemorrhage.

Round Cell Tumors Ewing's Sarcoma

Ewing's sarcoma is a highly malignant primary bone tumor of children (age range 5 to 15 years) that tends to arise in the diaphyses of long bones. The spine and pelvis also may be primary sites. The radiographic appearance usually is that of an aggressive lesion, with a permeative pattern of bone lysis and periosteal reaction (Fig. 41-128). Often there is an associated large soft-tissue mass, and patients have systemic symptoms (fever, weight loss) in addition to local pain, which tends to be worse at night. A soft-tissue variant of Ewing's sarcoma, primitive neuroectodermal tumor (PNET), occurs as well, usually exhibiting evidence of neural differentiation immunohistochemically. Differential diagnosis includes osteomyelitis, lymphoma, and eosinophilic granuloma.

Diagnostic evaluation includes chest and abdominal CT scans and bone scan to rule out metastases. Treatment consists of a combination of local radiation therapy and systemic chemotherapy. Five-year survival rates with this approach are around 50 percent. A multimodality treatment that uses adjuvant surgery (wide or marginal resection) has resulted in 5-year survival rates of 75 percent. In young children amputation may be necessary because of the severe compromise of bone growth that can result from the effect of the required levels of radiation on the growth plates.

Histiocytic Lymphoma (Reticulum Cell Sarcoma)

This tumor occurs in patients 20 to 40 years of age, usually affecting the diaphyses of long bones. Its radiographic appearance is similar to that of Ewing's sarcoma, with permeation, periosteal reaction, and frequently a large associated soft-issue mass. Pathologic fracture may occur. A significant proportion of patients present with or develop regional or distant lymph node involvement. Treatment consists of radiation to the local lesion in conjunction with systemic chemotherapy. If feasible, resection of the primary tumor improves survival and decreases the risk of local recurrence.

Other Tumors Giant Cell Tumor (Osteoclastoma)

These tumors arise in the epiphyses of young adults, most commonly in the proximal tibia, distal femur, proximal femur, and distal radius. Characteristically the lesion is radiographically purely lytic, well circumscribed, and occasionally expansile with cortical destruction. The lesion often extends to the subchondral surface and can even invade the joint (Fig. [41-129](#)). Although usually benign, a malignant variant occurs in a small proportion of cases, and even the benign lesions are stage III tumors, with local aggressive behavior and a high tendency to recur after surgical treatment. Patients usually present with pain, and pathologic fracture may occur. The tumor consists of monocytic stromal cells, vascular tissue, and sheets of large, multinucleated osteoclast-like cells. The key feature in differentiating these tumors from other tumors that can contain large numbers of giant cells (eosinophilic granuloma, brown tumor of hyperparathyroidism, aneurysmal bone cyst, chondroblastoma, osteoblastoma, nonossifying fibroma) is that the oval nuclei of the monocytic stroma resemble those of the giant cells, suggesting a common origin. The most common cause of malignant giant cell tumors is prior radiation therapy for a benign giant cell tumor, which was a former mode of treatment and can be associated with malignant recurrence in up to 10 percent of cases. Because of this radiation is no longer used in the treatment of giant cell tumor except in dire circumstances (such as unresectable lesions in the spine with threat of neurologic deficit).

The most common treatment of giant cell tumor, curettage of the lesion, is associated with recurrences in 25 to 50 percent of cases. Alternative treatments therefore have included wide resection (usually reserved for recurrent cases) and adjuvant local treatments such as cryotherapy with liquid nitrogen or phenol and, most recently, filling the defect with methyl methacrylate. The lowest recurrence rates have been with cryotherapy and methyl methacrylate cementation. Cementation causes a thermal kill of tissue within several millimeters of the margin in bone as a result of the exothermic reaction that occurs during polymerization of the cement. If local recurrence occurs after cementation, it is readily detectable radiographically as a lucency next to the cement. With bone grafting, remodeling changes in the graft can obscure signs of recurrence. Because these are epiphyseal lesions, the presence of cement next to the articular cartilage may predispose to cartilage degeneration, and in young patients, some advise removal of the cement and bone grafting after 2 years if the patient remains free of recurrence. Given an incidence of joint degeneration of only 15 to 20 percent in long-term follow-up studies, the indications for cement removal are controversial. Control of the lesion with this treatment approach has been successful in 90 percent of cases.

Vascular Tumors Hemangioma

Hemangiomas of bone often are noted in the spine as an incidental finding. These benign lesions are characterized by endothelial vascular spaces, and because they typically do not cause symptoms management usually is simply observation. A more aggressive lesion is the hemangioendothelioma, which can occur in bone or in the soft tissues and generally is characterized as a low-grade malignancy. In bone the lesions appear cystic and well marginated, with increased local perfusion on bone scan or angiography. Occasionally lesions occur in several bones. Treatment is with wide local resection, although curettage with radiotherapy has been successful in some cases.

Angiosarcoma

Angiosarcoma is a highly malignant sarcoma of the bone or soft tissues. The prognosis of this lesion is poor, with early hematogenous spread to the lungs the rule. Amputation or radical resection in nonmetastatic cases is appropriate.

Tumors Arising from Included Tissues Adamantinoma

Adamantinoma is a rare epithelial tumor occurring in the jaw and occasionally in the tibia or fibula of young adults. The tumor, although malignant, is slow growing and presents with pain and a lytic, multiloculated or bubbly radiographic appearance. The diaphyseal portion of the bone tends to be involved. Treatment is with wide resection or amputation, and adjuvant therapies have not been shown to be effective. Metastasis to the lungs occurs in about 50 percent of cases.

Chordoma

This rare, low-grade malignant neoplasm arises in the sacrococcygeal or occipitocervical area and is thought to develop from embryonic remnants of the notochord. Sixty percent of cases occur in the sacrum or coccyx (Fig. [41-130](#)). Patients present with a mass, neurologic symptoms, or pain. The lesions are slow growing and occur usually in older adults. Differential diagnosis includes plasmacytoma, giant cell tumor, and metastatic carcinoma. The tumor is composed of cords and nests of cells resembling chondrocytes, with typical highly vacuolated “basket” or physaliferous cells. The stroma consists of a basophilic, mucoid, or myxoid ground substance. The location makes wide resection difficult and causes significant morbidity, but without treatment the lesion is uniformly fatal, with late pulmonary metastases. The lesions are not responsive to radiotherapy or chemotherapy, and surgical resection is the treatment of choice. Some recent evidence suggests that this tumor may be somewhat responsive to proton beam irradiation.

Soft-Tissue Sarcoma

Soft-tissue sarcomas are more than twice as common as malignant primary bone tumors. Malignant fibrous histiocytoma is the most common type, but a wide variety of other histogenetic types exist, including fibrosarcoma, liposarcoma, malignant nerve sheath tumors (neurofibrosarcoma or malignant schwannoma), rhabdomyosarcoma, synovial sarcoma, lymphoma, primitive neuroectodermal tumor (PNET), and extraskeletal chondrosarcoma. In general these lesions occur in patients over 50 years of age, and the treatment is similar for all tumors despite the differences in histogenesis. Soft-tissue sarcomas are as a rule somewhat sensitive to radiation. While chemotherapy has proved benefit in controlling disease in patients with metastasis and prolonging their survival, its role as an adjuvant therapy is controversial, with the majority of recent data indicating only minimal efficacy in improving outcome. Exceptions to this include rhabdomyosarcoma, PNET, and lymphoma.

Treatment usually involves appropriate staging followed by a combination of surgery and radiation therapy. Achieving wide to radical surgical margins is necessary, and in most cases can be accomplished by a limb salvage operation. MRI is essential in treatment planning and in assessment of local compartment involvement (Fig. [41-131](#)). Radiation therapy may be administered preoperatively or postoperatively by brachytherapy or external beam irradiation. With this approach 90 to 95 percent local control can be anticipated, but a significant proportion of patients (about one-third) succumb to later metastatic disease. In selected cases of soft-tissue and other sarcomas, resection of pulmonary metastases has led to cures in approximately 30 percent of those treated.

Metastatic Bone Tumors

Carcinomas often metastasize to the skeleton, and metastatic lesions are much more common than primary bone lesions in general orthopaedic practice. The five primary cancers with a strong propensity to metastasize to bone are those originating in the breast, prostate, lung, kidney, and thyroid. Multiple myeloma, although technically a primary bone tumor, also must be considered in this group because of its similar age distribution (patients over age 50 years), radiographic presentation, and orthopaedic problems and treatment (pathologic fractures). Over 90 percent of patients with metastatic breast or prostate carcinoma have at least microscopic bone involvement.

The axial skeleton, including the skull, thoracic spine, ribs, lumbar spine, and pelvis, is most commonly involved. The proximal long bones, particularly the humerus and femur, also are affected frequently. Acral (distal) metastases are uncommon and are almost always secondary to lung carcinoma when they do occur. The predilection of particular tumors for bone, and for particular regions of specific bones, is thought to be caused by cytokines, local growth factors, or matrix components that attract and support growth of these lesions in specific areas. Lesions can be blastic (breast, prostate), lytic (breast, lung, myeloma, kidney, thyroid), or mixed (breast, lung) in radiographic appearance. Blastic or sclerotic lesions are less prone to pathologic fractures.

Patients with multiple lesions may have elevated alkaline phosphatase levels and occasionally are hypercalcemic, a result of secretion of PTH-like protein by some tumors, or more frequently, secondary to massive osteolysis by the tumor cells. The presenting complaint usually is pain in the affected area. Patients with spinal lesions may present with neurologic deficit or back pain. The major orthopaedic problem is that of fracture or impending fracture, with resulting functional disability and pain.

The mainstay of treatment of metastatic disease is radiation therapy, which often controls symptomatic lesions with relatively moderate doses (35 Gy). Larger lesions (larger than 3 cm in a weight-bearing bone), lesions that progress despite radiation, lesions that involve more than one-third of the cortex, and lesions that present with pain on weight bearing (impending fractures) should be internally fixed prophylactically. Fractures are treated surgically if the patient is able to tolerate the procedure medically, since aggressive mobilization significantly improves quality of life. Newer prosthetic implants for joint reconstruction and fracture fixation allow stabilization in the majority of cases (Figs. [41-132](#) and [41-133](#)). Bracing or casting is rarely successful for pathologic fractures, since pain control remains a persistent problem, and fractures usually will not heal by closed means if irradiated because of the suppression of callus formation by radiation therapy. Exceptions include spinal fractures, which respond to bracing and radiation treatment, but if neurologic deficit occurs they require surgical decompression and internal fixation either anteriorly or posteriorly. If large areas of bone are destroyed, stabilization often necessitates filling the defect with methyl methacrylate cement to supplement hardware fixation. The goals of treatment are maintenance or restoration of function and pain relief, since carcinoma metastatic to the bones is essentially always incurable. Rarely, a solitary metastasis is amenable to curative resection if the primary tumor has been removed; this situation can occur with renal cell carcinoma. Resection or amputation is also considered for pain relief or control of bulky, fungating lesions. Experimental treatments under investigation for metastatic disease, including immunotherapy and bone marrow transplantation, may offer future alternatives to current palliative treatment approaches.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 42 - Surgery of the Hand

Clayton A. Peimer

Philosophically and anatomically, the hand and brain uniquely identify *Homo sapiens*. Throughout history humankind's progress is measured through the evolution of a strong and mobile upper limb with an independently opposable thumb and the cognitive powers to use them. The balance, precision, and specialization of the hand give it a central role. The goal of surgical treatment of the hand is to retain maximum useful part length, independent, stable motion, and unimpaired mobility of sensate parts.

GENERAL CONSIDERATIONS

Examination

Prior records and diagnostic images may precisely define the extent, limitations, and duration of the patient's disorder, and the clinical course. The history should include information of relevant systemic disease such as diabetes, atherosclerosis, neurologic and psychiatric disorders, and other serious diseases or chronic diseases.

The examiner should use the patient's normal anatomy—the contralateral, uninvolved limb—to observe for differences in alignment, contour, and symmetry. Observing the hand and forearm at rest in pronation and in supination, should reveal any swelling, masses, erythema, ulceration, atrophy, anhidrosis, or excoriation (Fig. 42-1). The reproducibility of the patient's active participation in the examination process is important. Responses should be consistent; repeated efforts, such as in grip testing, should produce similar values. Accurate recording of information by the examiner is important (Fig. 42-2).

Light palpation provides information concerning excessive or absent sweating associated with anxiety or insensibility in particular zones, nerve distributions, dermatomes, or body parts. Variations in skin contour, texture, color, temperature, capillary refill, and hair characteristics offer information regarding circulation, nerve supply, masses, and joint swelling. Abnormal, injured, and scarred soft tissues can restrain joint motion, produce skin blanching with attempted active function, or cause visible “dimpling” of adherent deep structures, such as injured, repaired, or adherent tendons.

The nails and eponychial and paronychial cuticular tissues often mirror systemic disease as well as acute and chronic injury. Nails have a limited range of biologic responses. Splitting and fissuring, onycholysis, and onychorrhexis may reflect loss of nail adherence to the bed matrix after trauma, aging, or malnutrition. The transverse posttraumatic nail crease that parallels the proximal nail fold and advances with growth (Beau's line) represents a single alteration of nail metabolism at the time of trauma. It is common after injury but does not offer a poor prognosis. Multiple transverse grooves (Mee's lines) can occur with diseases such as Hodgkin's disease, malaria, and psoriasis and are normal in the latter part of pregnancy. Pigmented longitudinal bands may occur in melanoma, glomus tumor, and carpal tunnel syndrome. Nail bed pigmentation can be found with systemic sepsis, subungual infection, and benign and malignant tumors (Fig. 42-3).

Motion should be recorded with a small goniometer and strength with a dynamometer. Simple line sketches record sites of injury, swelling, part loss, or dysfunction and can precisely record and communicate findings (Fig. 42-4).

Imaging Studies

Diagnostic imaging includes traditional roentgenography, single- and multiple-phase technetium bone scans, computed tomography (CT), and magnetic resonance imaging (MRI). Most patients should receive plain radiographs in posteroanterior lateral and one or both oblique projections. Radiographs provide information with relatively intermediate sensitivity, high specificity, and reasonable cost. Diagnoses can be missed if only specialized and expensive evaluations such as trispiral or CT, MRI, or bone scans are used.

The x-ray beam must be centered on the part in question. Requesting an x-ray of the hand may be too general for diagnosing a problem in a specific finger. The physician evaluating the imaging studies should receive the history, physical findings, and a working differential diagnosis (Fig. 42-5).

TRAUMA

It is estimated that one-third of musculoskeletal trauma occurs to the upper limb. The annual economic impact from diminished and lost hand and upper extremity function runs into the billions of dollars in the United States. The best care is delivered initially, when tissues are fresh and potentially can be salvaged, revascularized, and directly repaired, without the burdens of secondary scar tissue after delayed healing, osteoarticular degeneration, or infection. It is in the acute situation that success is greatest in achieving a potentially functional and aesthetically satisfactory result.

Skeletal Trauma

Forearm and Wrist

The forearm is a two-bone musculotendinous unit with complex biomechanical interactions. The osteoarticular and ligamentous connections of radius, ulna, interosseous membrane, and proximal and distal radioulnar joints allow the variety of hand placement and force transfers. The disruption of any one part of this anatomic and mechanical construct should cause the physician to search also for less obvious injuries elsewhere. Most suboptimal outcomes after forearm fractures can be related to failure to recognize injury to a proximal or distal region, often at the radioulnar joint, or failure to appreciate progressive loss of initial reduction after treatment. Severe neurovascular problems are uncommon, but, when present, they may be related to acute vascular compromise from trauma, immobilization, or compartment syndrome.

The distal radius forms a biconcave articular surface to seat the proximal row of carpal bones, scaphoid, lunate, and triquetrum. The radius articulates distally not only with the carpals but also with the ulna. The length and relationships of these several bones must be considered after trauma. The traumatized limb usually can be compared to a preinjury state by radiographs of the opposite side (Figs. 42-6 and 42-7).

Distal radius fractures have been variously classified, but a guide that integrates assessment of fracture patterns and treatment is useful (Table 42-1). Because there are many variations of fracture patterns in the distal radius, this system simplifies common characteristics and suggests treatment for each generic type of fracture. The treatment for a specific patient should be based not only on the pattern but also on age, hand dominance, occupation, social needs, cognitive and psychosocial factors, and living arrangements.

Radiographically, radius fractures are stable or unstable, displaced or nondisplaced, and with or without involvement of the radiocarpal or radioulnar articular surfaces. The condition of surrounding soft tissues and the presence of associated injury in other regions of the hand, forearm, and elbow are part of the decision-making process. Because these fractures most commonly occur from a fall on an outstretched hand, they result in a bending moment with the axial load through the radial metaphysis, and transient or complete associated neurovascular injuries, particularly to the median nerve at the carpal canal, are common. There may be an associated avulsion fracture of the ulnar styloid process.

The radius is more commonly displaced dorsally, as in Colles' fracture, with impaction, proximal displacement, and dorsal angulation of its distal articular surface. In such cases, the dorsal cortex of the distal fragment is comminuted and has significant prognostic implications for postreduction bone stability when managed entirely by external manipulation and cast or splint fixation (Fig. 42-8). Less frequently, a flexion or bending occurs in a fall, creating palmar displacement (Smith's fracture). When intraarticular, either of these fracture patterns may have associated radiocarpal subluxation, but this problem is more often associated with the volar lip variant, Smith Type 3 or Barton's fracture pattern.

Even nondisplaced fractures can potentially displace during the healing process as resorption and bone remodeling occur at the fracture site. Maintaining a satisfactory degree of length and alignment during healing is important, and patients should be observed every 10 to 15 days with examination and x-rays. This injury may be associated with significant swelling, and circular cast immobilization should not be used initially; cast immobilization should be delayed until initial swelling has subsided. A sugar-tongs forearm and wrist splint may be applied that maintains bone length and alignment while simultaneously controlling forearm rotation without a rigid circumferential shell (Fig. 42-9).

The displaced dorsal fragment can be reduced comfortably under fracture block anesthesia, with or without addition of intravenous or intramuscular sedation. After sterile skin preparation a 22-gauge needle is introduced obliquely into the fracture site via overlying skin. Aspiration will reveal when the needle is within the fracture hematoma; 6 to 8 mL 1% mepivacaine hydrochloride (Carbocaine) or lidocaine hydrochloride (Xylocaine) without epinephrine can be injected to produce anesthesia in 10 to 15 min. The typical fracture, with dorsal displacement and angulation from an extension vector, can be reduced by distraction applied by allowing the arm to hang from finger traps placed on the index and middle fingers, with the arm itself serving as a counterweight. A dorsal-to-palmar force is applied over the area of displacement after the fragments have been disimpacted. The wrist is immobilized in neutral forearm rotation with only slight wrist flexion or ulnar deviation to avoid causing secondary compression of the median nerve. The sugar-tongs splint is applied. The width of the plaster should allow adequate space for tissue swelling. The dorsal and palmar edges of the splint should not touch each other. For most adults, a 3-inch plaster width is adequate, but in large individuals 4-inch plaster may be needed, which is applied over generous soft-tissue padding. After reduction and splint application, x-rays are obtained to record the reduction. If fracture realignment is incomplete, or if significant intraarticular displacement (1 mm or more) osteoarticular incongruity remains, repeated manipulative reduction or an alternate method of treatment should be considered.

Patients should have a neurovascular examination before manipulative reduction, and their median nerve sensory status should be assessed again after the application of the plaster splint. Fracture position and neurovascular status must be followed carefully.

After 7 to 10 days, the patient has repeat x-rays. Elevation and digital motion will have diminished swelling significantly in many cases, and the splint may be replaced with a circular cast. Some circumstances may dictate snugging the splint with a replacement circular gauze overwrap and delaying cast application an additional week or two. In young patients, a long arm cast may be preferable; in the older individual, the risk of elbow stiffness is significant, and prolonged immobilization of that joint is not advisable. Elevation of the hand plus finger mobilization and a therapy program that is directed to the entire upper limb, including the shoulder, is started at this time. When the splint is changed into a cast, x-rays are obtained after cast application. Radiographs are repeated every 2 weeks to monitor healing and observe for collapse, angulation, and displacement (Fig. 42-10). Most casts can be removed 6 weeks after trauma, when radiographs demonstrate obvious new bone formation and the fracture region is relatively nontender. If no motion or significant pain is elicited at the fracture site, cast immobilization may be discontinued. Therapists can fabricate a custom thermoplastic resting/protective splint and active motion exercises can begin (Fig. 42-11). The custom splint can generally be discontinued after another 2 to 4 weeks, depending on patient comfort and progress in rehabilitation. Motion exercises for the hand and wrist are followed by progressive strengthening, increasing activities of daily living, and return to function.

Unstable fractures of the distal radius are treated with percutaneous or open fracture pinning under fluoroscopic monitoring. Many of these fractures require an external fixation device to prevent progressive collapse and loss of alignment at the comminuted fracture line. When fractures are significantly impacted, areas of obvious bone loss within the submetaphyseal region after reduction may require bone graft or bone substitute to prevent delayed union or nonunion. Management by open reduction, with the combination of external and internal fixation with plates and screws, is appropriate for the more severe subgroup, most often those of younger age whose injuries are the result of high-impact trauma, such as those sustained in vehicular accidents and mishaps with heavy machinery (Figs. 42-12, 42-13, and 42-14).

Carpal Bone Fractures

The eight carpal bones have a large proportion of their surfaces covered with articular cartilage, a fact that has two clinical implications. First, the limited periosteal attachment offers a tenuous blood supply; after fracture one of the fragments is potentially at risk for avascular necrosis. Second, most carpal fractures are intraarticular injuries. The displaced fracture often needs surgical repair to avoid secondary arthritis from joint surface incongruity. The pattern of carpal fracture or fracture dissociation may not be clearly discernible on standard posteroanterior and lateral radiographs, and oblique views, carpal tunnel projection, and other views may be necessary. If results are still equivocal, trispiral tomography or CT should demonstrate the fracture patterns and fragment positions. When the question of whether or not a fracture is present, a frequent problem with injuries about the radial side of the carpus, especially the scaphoid, the use of technetium bone scan 72 h after trauma is diagnostic.

Scaphoid Fracture

Nearly two-thirds of all carpal fractures are of the scaphoid. This injury occurs most often in males aged 15 to 30 years. Scaphoid fractures occur most commonly through the middle third of the waist or at the juncture of the middle and proximal poles. Diagnosis requires clinical and imaging information. After a fall on the outstretched hand, the patient's wrist is tender at the anatomic snuff box, the hollow between the thumb extensor tendons on the radial aspect of the wrist, just dorsal and distal to the styloid process of the radius. Pain is elicited and symptoms reproduced with direct pressure over the tuberosity of the scaphoid at the base of the thenar eminence and with passive wrist motion. Routine radiographs in posteroanterior, lateral, and oblique views along with a posteroanterior projection in ulnar deviation to elongate the scaphoid helps to visualize the fracture. If initial radiographs are normal but the history and physical examination suggest the possibility of scaphoid fracture, continuous immobilization in a thumb spica splint or cast is advised. Repeat radiographs in 2 to 3 weeks or technetium bone scan after 72 h will make the diagnosis.

Fracture configuration (Fig. 42-15) and location affect stability and lability of the blood supply. The proximal pole of the scaphoid is supplied from vessels entering the distal two-thirds. Fracture of the proximal third or a smaller fragment risks avascularity in the small proximal fragment, resulting in nonunion and secondary arthrosis. Nondisplaced scaphoid fractures treated with adequate immobilization have a union rate of 90 to 95 percent. Displaced fractures, defined as displacement of 1.0 mm or more, are associated with avascular necrosis in one-half and nonunion in one-half of patients if not reduced and stabilized operatively. Scaphoid fracture fragment displacement of 2.0 mm or more should raise suspicion of an associated intercarpal ligament injury, such as transscaphoid perilunate instability or subluxation.

In the treatment of the nondisplaced, stable scaphoid fracture, one study demonstrated decreased time to union and decreased incidence of delayed and nonunion when long arm, above-elbow thumb spica casts were used for 6 weeks, followed by short arm spica cast for approximately 6 weeks. Immobilization of the wrist in slight flexion and radial deviation relaxes the volar radioscaphoid ligament. The thumb metaphalangeal joint should be included in the cast, at least during the initial 6 weeks. Open reduction and internal fixation can be done effectively with an interfragmentary lag compression screw or with Kirschner wires for all displaced fractures. The screw technique is more stable and allows earlier mobilization.

The use of percutaneous compression screws for immediate internal stabilization of the acute but nondisplaced fracture is increasingly popular outside the United States. Limited experience suggests a more rapid course with decreased acute and long-term disability without significant increase in complications. Interfragmentary screw technique for this bone is technically demanding.

When immediate postinjury imaging does not clearly demonstrate the presence of fracture, immobilization and additional imaging information by standard radiographs in 2 weeks or bone scan after 3 days are necessary to make the diagnosis. Fracture displacement is unacceptable, and 1 to 2 mm of malposition, angulation, or any intercarpal collapse should prompt open repair. Less than 5 percent of nondisplaced fractures result in nonunion, which is defined as absent x-ray evidence of healing 4 to 6 months after injury, while a 50 percent nonunion rate for displaced fractures occurs.

Other Carpal Fractures

The *lunate* most commonly fractures secondary to idiopathic avascular necrosis or lunatomalacia (Kienböck's disease) without a history of acute trauma. The lunate is more radiodense on posteroanterior projection radiographs. Mild radiodensity is not uncommon after carpal injuries and should not be confused with Kienböck's disease. Simple, acute lunate fractures are not common. They should be treated with immobilization when nondisplaced; open treatment rarely is required.

Fractures of the *capitate* are uncommon. The proximal pole, like that of the scaphoid, receives its blood supply from vessels that enter distally. Capitate neck fractures are at potential risk for avascular necrosis. When avascular necrosis occurs, usually it is incomplete, that is, temporary. Capitate head collapse is uncommon.

Isolated fractures of the *trapezium* and *trapezoid* are uncommon. Injuries to these bones are often associated with intraarticular fractures of the base of the first metacarpal. Open reduction is often necessary, with internal fixation by Kirschner wires or compression screws.

Fractures of the *pisiform* usually are secondary to a direct blow to the hypothenar eminence. The fractured pisiform is best seen in carpal tunnel projection radiographs or CT imaging. Excision may be required for displaced fractures, nonunion, malunion, and secondary arthritis. There are two types of fracture of the *hamate*: those involving the body and those of the hamulus (hook). Fractures of the body of the hamate are difficult to diagnose on plain radiographs; imaging may require several oblique projections or CT scan. The patient may have pain referred to the dorsal wrist with fractures of the body or hook. Fractures of the body of the hamate heal with immobilization for 4 to 6 weeks. Fractures of the hamulus are more common and usually are the result of a direct force transmitted into the base of the palm from a grasped object. The palm is tender to direct pressure over the hamate; sometimes the discomfort is reported as dorsal. Secondary ulnar neuropathy at Guyon's canal also may be present. Late flexor tendon ruptures have been reported, especially in those whose undiagnosed palmar pain syndrome is treated with repeated steroid injections. Routine radiographs and carpal tunnel views may be negative; CT scan can make the diagnosis. The acute hook fracture should heal in a short arm cast; displaced fractures and symptomatic nonunions are most efficiently treated with excision of the hamulus and smoothing of the fracture base.

Carpal Dislocations and Instabilities

The radiocarpal and intercarpal articulations are not inherently stable on the basis of their osseous anatomy; it is the integration of osteoligamentous anatomy that secures the complex kinematics of wrist function (Fig. 42-16). Most carpal dislocations are caused by an acute axial load with wrist hyperextension. The primary dislocation occurs at the midcarpal joint with dorsal displacement of the capitate. When the capitate displaces, the scaphoid must fracture or its ligaments will tear, allowing it to rotate from a relatively horizontal position to one of vertical malalignment with the proximal pole rotating dorsally. This configuration is called *dorsal perilunate dislocation* (Figs. 42-17, 42-18, and 42-19). These serious and unstable intraarticular injuries, with or without scaphoid fracture or triquetral break, require careful reduction and internal fixation. The majority require open reduction. Direct trauma to the median nerve from impact, by secondary stretching resulting from dorsal displacement of the carpus, or from acute bleeding and swelling within the carpal tunnel, should be elucidated by neurovascular examination. Carpal instabilities of all types should be treated aggressively to prevent chronic instability and dysfunction (Fig. 42-20).

Metacarpal Fractures

Because of their subcutaneous location and relatively rigid proximal articulations, the metacarpals represent one-third of hand and wrist fractures. Failure to reconstitute the metacarpals may lead to permanent functional deficit. Complication rates after extensive exposure for plate fixation can be high, and the risk of additional injury must be weighed against outcomes expected with conservative measures.

The goal is early restoration of hand function to prevent stiffness. Whether internal or external immobilization is used is immaterial, as long as bone length and articular relationships are preserved and soft-tissue management and therapy techniques can be instituted rapidly.

The metacarpals form a rigid longitudinal arch because of their convexity dorsally. There also is a dynamic transverse arch based on the stable and mobile carpometacarpal articulations of the thumb and those of the ring and little fingers. The thenar and hypothenar muscles mobilize these arches, allowing precision and strength in hand use. Because of stable proximal and distal ligamentous support, isolated fractures of the central third and fourth metacarpals—the middle and ring fingers—are less likely to shorten, rotate, and angulate. Spiral and oblique fractures that displace do so with shortening and rotation. Metacarpal shortening also may occur by direct bone loss or angular deformity. Midshaft angulation produces a more serious deformity (Fig. 42-21).

Pain and swelling are the hallmarks of metacarpal fractures, as the loose dorsal tissues allow large amounts of edema fluid and fracture hematoma to accumulate. The bony prominence of an angulated fracture apex is always located dorsally because of the pull of the interosseous intrinsic muscles. A skin laceration often connotes an open fracture and mandates surgical treatment. This is important in metacarpal fractures caused by a tooth impact, as in a fight; this results in a contaminated puncture wound at the fracture site or at the metacarpophalangeal joint. Patients with human or animal bites require surgical irrigation of the fracture site or joint plus high-dose antibiotics.

Rotational alignment of a metacarpal fracture is best assessed with the fingers flexed at the metacarpophalangeal joint. With an uncooperative juvenile patient or an unconscious patient, the wrist can be passively flexed and extended, with the resulting extrinsic flexor and extensor effect on digital alignment observed. Malrotation or active flexion produces a degree of visible digital overlap. Malrotation and radial-ulnar angulation interferes with hand function and should be corrected.

Fractures of the metacarpal heads are less common, usually the result of direct trauma. The second and fifth metacarpals are most commonly traumatized, with a 3:1 male predominance. As these are intraarticular injuries, a step-off of 1.0 mm or more is significant. Unstable fractures are fixed with pins, screws, or plates.

The metacarpal neck is the most common fracture site. As with displaced and angulated fractures, the cortex on the angulated side usually is comminuted. The normal pull of the intrinsic muscles further flexes the head fragment, making it difficult to maintain reduction. The degree of angulation and the metacarpal involved determines the best treatment for the specific fracture. Because the second and third carpometacarpal joints are rigid, no more than 10 to 15 degrees of palmar angulation of the distal fragment is acceptable. Considerably more angulation (30 to 50 degrees) may be acceptable in the neck region of the fourth and fifth metacarpals. Closed reduction may be achieved through the combination of direct and counterpressure applied with the finger flexed (Fig. 42-22). The hand should not be immobilized in the position depicted for manipulation. Open reduction usually is not necessary, but when the fracture is unstable and residual or recurring angulation is not acceptable, then internal fixation is required. Metacarpal shaft fractures should be protected when position and angulation are acceptable, but repaired when they are not. Spiral and oblique fractures undergo malrotation and displacement because of the normal forces of the flexor and extensor tendons and hand intrinsic muscles. Those that are not initially displaced or rotated must be carefully observed. Internal fixation allows more rapid soft-tissue mobilization and often can be treated with percutaneous intermetacarpal pin technique, local block anesthesia, and fluoroscopic monitoring. Some of these fractures require open reduction for fixation (Figs. 42-23, 42-24, and 42-25).

Intraarticular carpometacarpal fracture-dislocations are of functional significance in the mobile fourth and fifth metacarpal-hamate saddle joints. These mobile joints have tendon

insertions around the bases of the metacarpal and periarticular tissue that allow displacement (Figs. [42-26](#) and [42-27](#)). Reduction is best maintained with internal fixation. Some require open reduction and fixation to restore accurate osteoarticular alignment.

Delayed union and nonunion of closed metacarpal injuries are uncommon. Bone consolidation should be discernible by radiograph within 12 to 16 weeks. Immobilization of the hand for several months seriously risks compromising function. Nonunion may be associated with inadequate immobilization, loss of bone substance, infection, or disruption of blood supply. Digital impairment may result from tendon adhesions directly over a fracture site, secondary small-joint contracture from prolonged immobilization, or scarring of traumatized intrinsic muscles. Simultaneous skeletal stabilization and small-joint mobilization should be achieved.

Phalangeal Fractures

The goal of phalangeal fracture treatment is restoration of anatomy, bone healing, and full functional recovery. Dysfunctional angulation and rotation are not acceptable. Stabilized fracture anatomy must allow rapid mobilization. Each method of fracture care has relative advantages and risks. Less invasive methods may offer less stability, but they inflict less soft-tissue damage. An algorithm for care is outlined in Fig. [42-28](#). When operation is required, the least traumatic method should be used to avoid violation of gliding structures when possible. The patient's active participation in a rehabilitation program encompassing supervised therapy, custom splinting, and home exercises is critical for recovery of function. Proximal interphalangeal joint motion, particularly extension, can be difficult to regain if an injured, swollen finger is immobilized in flexion. Scar can tether the extensor tendons or prevent the flexors from gliding, impairing grasp and manipulation and preventing return to preinjury employment.

When Kirschner wires are used, they may be buried, and they may then be retrieved in the outpatient setting under local anesthesia after 4 weeks. Sufficient fracture healing usually has occurred by then despite the delayed appearance of significant interfragmentary callus on radiographs. When Kirschner wires are left external to the skin, as in juveniles, pins must be capped and cared for meticulously. Screws and plates usually are not removed until at least 6 to 12 months after fracture healing. Small bone plates and screws need not be removed except to treat symptoms from the hardware.

Finger Ligament Injuries

Metacarpophalangeal (MP) Joint

MP joint dislocations can be managed by closed means through gentle reduction and splinting under local anesthesia. If significant residual collateral ligament instability in a particular finger is present, surgical repair is necessary. The small subgroup of irreducible fractures requires operative repair. Patients with acute collateral injuries may have a malrotated finger (Fig. [42-29](#)) because of rotation about the intact ligament. The ruptured ligament region is swollen and tender. Evaluation by gentle passive stress should be done with the MP joint in flexion, a position in which the collateral ligaments are normally tight. Some perform simultaneous radiographic evaluation during this passive stress. Patients with particular discomfort who cannot tolerate soft-tissue stress in order to evaluate joint stability can be examined after 1.0 mL of local anesthetic agent is injected into the joint. Dorsal dislocations that are irreducible are characterized by dimpling of the palmar skin over a prominent metacarpal head. Interposed soft tissues can prevent joint reduction. In these cases surgical treatment is required (Fig. [42-30](#)).

Thumb MP joint injuries result from axial load and angular displacement. These injuries often occur when the patient jams the thumb into an object while falling. Disruption of the ulnar collateral ligament of the thumb is called *gamekeeper's thumb*, although the term was originally applied only to chronic ulnar collateral instability. A larger percentage of the injuries are caused by jamming the thumb into snow in a fall while skiing. Collateral laxity at the thumb MP joint is dysfunctional and painful and may lead to late arthritis. After plain radiographs fail to detect the presence of intraarticular fractures, the thumb is carefully examined in about 30 degrees of MP flexion, gently and progressively stressing the suspect collateral ligament (Fig. [42-31](#)). Radiographs may be obtained simultaneously (Fig. [42-32](#)); the stress radiograph is best performed by the examining physician. Treatment of incomplete collateral ligament injuries without associated instability is best done closed, with cast immobilization for approximately 4 weeks, followed by custom-splint immobilization. Soreness may persist for several months. Complete disruption of the ulnar or radial collateral ligament of the thumb MP joint should be repaired and protected by temporary pin fixation of the joint, which is most likely to give a better result and shorter period of disability than secondary reconstruction.

Proximal Interphalangeal (PIP) Joint

The tightly congruent osteoarticular contours of the proximal interphalangeal joint make restoration of stable alignment of disrupted or displaced structures essential, allowing safe institution of early mobilization. Stiffness, rather than instability, is the outcome that must be avoided after trauma in the region of the PIP joint. Most dorsal and lateral PIP dislocations can be treated by closed reduction and should be stable. Immobilization for 10 to 15 days allows the patient to recover from the acute posttraumatic effects before a protected mobilization program is started, with buddy tapes to an adjacent finger. Joints without an actual history of displacement, deformity, or reduction by patient, coach, trainer, or physician may have considerable swelling and stiffness if not mobilized early. Dislocations with fractures are more likely unstable (Figs. [42-33](#) and [42-34](#)). Postoperative immobilization that inadvertently stresses an osteoarticular fragment results in posttraumatic instability (Fig. [42-35](#)). The combination of joint surface impaction and ligament disruption have the worst prognosis (Fig. [42-36](#)). These fracture-dislocations have an ultimate outcome that is often unsatisfactory.

Palmar (Volar) PIP Dislocations

In volar PIP dislocations the middle phalanx is displaced palmarward, sometimes resulting in serious instability. This PIP dislocation results from the combination of axial load and palmar vector force, most often during sports activities (Fig. [42-37](#)). Often unrecognized is that this trauma has an associated disruption of the central slip of the extensor tendon and one collateral ligament. Closed reduction and pinning, or open reduction for the irreducible variant, with prolonged postoperative therapy is the rule.

Distal Interphalangeal (DIP) Joint

The DIP joint must be comfortable and stable for precision pinch and power grip. Ideally, rehabilitation after DIP joint trauma restores a pain-free and stable arc of motion, but this is not always possible. Some compromise between stability, motion, and symptoms may be necessary; functional position, near extension, and DIP joint stability are the most critical to regaining a useful hand. Collateral ligament injuries and dorsal or palmar dislocations may occur at this level. Stable joints for which closed reduction is possible need not be pinned. Percutaneous fixation under fluoroscopic guidance, with maintenance of pin fixation for 4 to 5 weeks, is a useful adjunct, because it allows the rest of the hand to be rapidly mobilized. It is preferable to bury pins in adults and to remove them in an outpatient setting using local anesthesia.

Swelling and discomfort persist after DIP and PIP injury for 3 to 6 months in most patients. Functional recovery of mobility and power occurs slowly. Protection during sports and similar activities may be needed for 6 months or more. It should be explained early to patients that the joint is likely to be sore or swollen for some time; the sooner patients understand this, the more likely they are to accept their role in recovery.

Fingertip Injuries

Conservative treatment, such as healing by secondary intention of fingertip amputations, often results in painful scarring and deformity. There are several requirements for a satisfactory outcome after fingertip amputation: (1) Optimum functional finger length must be maintained, and additional shortening during or as a complication of treatment must be avoided. (2) The residual tip/pulp requires a resistant and resilient character like normal skin. (3) Excellent fingertip sensibility should be maintained to avoid "blinding" the finger. (4) Finally, bone support for the nail is needed to minimize beaking deformity. Achieving all of these targets simultaneously may be impossible, and choices may be necessary. Anatomy and function in conjunction with the type and level of injury in each patient should be considered (Figs. [42-38](#) and [42-39](#)).

Which finger is injured and how it was injured influence treatment. For the thumb, every reasonable effort must be made to restore a sensate and durable pulp. Requirements for sensibility are more critical in the index and middle fingers, but they are also significant in the ulnar pulp of the small finger. Amputations can be clean and sharp, but the common avulsion may have a component of avulsion, crush, blast, and burn, as in explosions. Explosions cause extensive trauma to surrounding skin, soft tissue, and neurovascular tissue that requires debridement and, in some cases, staging of the closure. Treatment of partial amputations, crush injuries, and partial devascularizing injuries should be directed toward preserving soft tissues.

Distal phalangeal fractures, including bursting or tuft fractures, are frequently associated with crush trauma and nail bed disruption or lacerations. Nail bed injuries are not always obvious, and subungual hematoma may be the only sign of nail bed injury. Nail bed injuries should be repaired to prevent permanent late nail deformity. Nail bed repairs usually are done with fine 6-0 absorbable suture. After repair, the nail that was removed is replaced beneath the cuticle to splint the bed (Figs. [42-40](#) and [42-41](#)).

Surgical treatments used to treat fingertip amputations include the following:

Bone Shortening and Primary Closure

This is performed under local or regional anesthesia and consists of debriding enough bone so that the skin can be closed with a few 5-0 sutures, without tension. Thorough debridement of contaminated or devitalized hard and soft tissues is required. This method affords coverage with soft tissues of normal sensibility, and this well-padded fingertip is not painful, but the cost is some length and at least a portion of the fingernail. Inadequate bone resection produces a fingertip with unpadded bone, resulting in pain during grasping.

Composite Pulp Reattachment

Reapplication of the "composite" of skin and pulp, or skin, pulp, and bone, can be done when the mass of the amputated part is very small. This choice should be reserved for young children. It is best to debride any residual bone. Superficial necrosis of the reappplied part should be expected. In most situations, this is a temporary biologic dressing.

Skin Grafting

Grafts are a means of coverage for skin defects; they are not time consuming and theoretically could be applied to a wide range of fingertip amputations. The major drawbacks are sensory loss in the graft area and the inadequate padding if graft is applied over unpadded bone and periosteum on prehensile surfaces.

The aesthetics of the graft are affected by the donor site. The best cosmetic result for the pulp surface is achieved in all races with split- thickness or full-thickness skin graft taken from the glabrous skin at the hypothenar eminence under local anesthesia. The defect covered with skin graft should be a skin defect, and the recipient bed must have adequate native padding. Skin graft will not satisfactorily protect bone, no matter what the skin donor source. Skin graft to the palm from any area other than glabrous skin produces a result that is relatively hyperpigmented. Split grafts usually are inadequate for pressure and friction surfaces. Toe- to-finger and foot instep-to-hand pulp skin grafting can be performed, but the short-term disadvantages are obvious as compared to full-thickness hypothenar skin as the donor.

Local Flaps

Local tissue transfer from more proximally on the injured finger affords vascularized, padded, and most often sensate tissue. Advancement is by transposition or rotation (Figs.

[42-42](#) and [42-43](#)).

Regional and Distant Flaps

Cross-finger, thenar, and other heterodigital flaps have been used since the early part of the twentieth century, generally for more extensive pulp loss and otherwise uncoverable bone and tendon (Fig. [42-44](#)). These flaps have the advantage of retaining finger length but carry the risk of posttraumatic deformity or dysfunction in an adjacent donor finger. Care must be taken to avoid dysfunction from immobilization of the injured or the donor part because of nonphysiologic positioning during flap healing and before pedicle detachment. Such flaps usually are not sensate.

Replantation and Microvascular Neurosensory Flaps

Microsurgical advances have made finger- and hand-part reattachment possible and allowed reconstruction by composite neurovascular pulp tissue from toes, with or without joints and tendons. The isolated single digit amputation in the adult usually is not suitable for replantation, especially if proximal to the PIP joint, because the functional and aesthetic recovery usually does not justify the morbidity and costs of the replantation procedure. Multiple digit amputations, subtotal hand amputations, amputations throughout the upper limb proximal to the hand, and most pediatric amputations should be evaluated for replantation or primary composite microvascular reconstruction (Fig. [42-45](#)).

Soft-Tissue Trauma

Tendon Injuries

Flexor Tendons

Flexor mechanism injuries in the hand and fingers are no longer treated by late reconstruction in most patients, because direct primary and delayed primary repairs offer good to excellent results, even when done in the middle of the digits. Satisfactory results are reported in 75 to 98 percent of patients in various series. Flexor tendon repair and functional rehabilitation is a challenge.

Flexor tendons are not difficult to repair, but achieving good function of repaired tendons is difficult, particularly in zones in which multiple tendons of different excursion are in juxtaposition. Tendons must glide, and simultaneously they are restrained by ligaments, such as within the digital sheath and within the carpal canal. Getting flexor tendons to glide after repair remains the problem. The critical operative principle in tendon repair is to achieve near-perfect anatomic alignment of the tendon ends. There should be no gaps at the repair site, and “bunching” of the repair zone should be avoided to permit the repaired tendon(s) to glide within a sheath or pulley system. The zones of flexor tendons are defined by the number of tendons, restraints, and pulleys and the presence or absence of synovial membrane at that specific anatomic level (Figs. [42-46](#) and [42-47](#)).

In the diagnosis of tendon disruption, the patient often presents with an open wound and loss of active motion. Observing the part at rest (Fig. [42- 48](#)) along with active, separate evaluation of the flexor digitorum profundus and flexor digitorum superficialis tendons (Fig. [42-49](#)) makes the diagnosis. A high level of suspicion should be maintained with injuries that have loss of active flexion or extension when x-rays do not show skeletal disruption. Tendon avulsions may occur without attached bone and can be diagnosed only if the examiner is suspicious. Closed, isolated flexor profundus avulsion is most common in the ring finger; the DIP joint will not flex, but the PIP joint does, however painful. For primary or delayed primary repair to be effective, early diagnosis is essential (Figs. [42-50](#) and [42-51](#)).

Partial tendon lacerations, approximately up to one-third of the tendon's cross-sectional area, do not present serious risk of rupture, but the lacerated edge may catch on a nearby pulley, producing posttraumatic triggering. Lacerations involving 30 to 50 percent of the tendon's cross- sectional area may be treated by epitendinous suture alone. Division of 50 percent or more of cross-sectional area should be treated surgically as though division were complete. Flexor mechanism salvage by graft and staged reconstruction, or by posttraumatic tenolysis or grafting, is beyond the scope of this text. Secondary tenolysis should be reserved for those patients whose fingers have achieved a stable biologic state more than 4 months after trauma. Soft tissues are no longer edematous, maximum active and passive joint mobilization has been achieved before secondary operation, and the patient must be willing to undergo a second operation for additional recovery.

Extensor Tendons

The superficial location of the extensor tendons on the dorsum of the fingers and hand make them vulnerable to injury, especially when the fingers are flexed. Trauma comes from lacerations, crush impacts, abrasions, and bites. Extensor tendon injuries are more common than those of flexor injuries and are often treated casually in the emergency department.

Extensor dysfunction may result in loss of active flexion from scar tattering and in diminished active extension. The extensor system is more intricate and complex than the flexor system. The interconnections of the *extrinsic* digital extensor tendons from the muscles in the forearm and tendons in the hand, and the *intrinsic* tendons in which muscles and tendons are in the hand, are complex. The two sets of tendons collaborate to flex the metacarpophalangeal joints and extend the interphalangeal joints. Because excursion of the extensor mechanism is limited over the finger joints, preservation of tendon length is more critical to maintain and restore tendon balance than with flexor tendon injury.

The flexor tendons are thick, round, cordlike structures with spiraling fibers. The extensor tendons are thin and flat, and the longitudinal fibers of the extensors do not hold sutures well. The limited amount of soft tissues about these tendons also makes repairs prone to adherence and scarring.

The extrinsic extensors of the forearm, i.e., brachioradialis and extensor carpi radialis longus, are innervated by the radial nerve, and the extensor carpi radialis brevis muscle by the deep branch of the radial nerve. The posterior interosseous branch of the radial nerve innervates the extensor carpi ulnaris muscle and all proper and common thumb and digital extensors. The tendons cross the wrist through six pulley compartments, serving to extend the wrist and the MP and *interphalangeal (IP)* joints. These six tendon tunnels are defined by reflections of the extensor retinaculum into the dorsal cortex of the radius and wrist capsule and serve to limit the tendon vector effect of the digital extensors at the wrist joint by maintaining their proximity to the center of axis of wrist motion. Extrinsic digital extensor tendons elevate the proximal phalanx. That is, they extend the MP joint via the aponeurotic sagittal fibers that reach around the lateral sides of the phalanx to insert on the palmar margin of the bone and volar plate, thereby lifting the metacarpal from a broad palmar attachment rather than a single point dorsally. The extrinsic extensor tendon is the only MP joint extensor. Distally, the function of the intrinsic and extrinsic tendons together form the dorsal tendon apparatus in the fingers (Fig. [42-52](#)). The intrinsic tendons arise from the four dorsal and three palmar interosseous muscles. There also are three thenar, three hypothenar, and four lumbrical muscles. The group serves as the primary independent metacarpophalangeal joint flexors and as interphalangeal joint extensors. The MP joint flexor fibers course distally from a palmar position, sending up the transverse aponeurosis dorsally to insert on the lateral edges of the extrinsic extensor tendon in the proximal third of the proximal phalanx (see Fig. [42-52](#)). The direct distal continuation of the intrinsic tendon is the lateral band that continues distally to reach a position dorsal to the center of axis of PIP motion before crossing over the distal third of the proximal phalanx, thereby making it an extensor tendon for both interphalangeal joints. At the metaphysis of the proximal phalanx and the base of the middle phalanx, the extrinsic and intrinsic tendons converge to become conjoined extensors. The central slip inserts into the dorsal lip of the middle phalanx as its direct extensor, but the conjoined lateral bands run along the dorsal lateral edge of the PIP joint and converge distally over the middle phalanx to become the terminal tendon that inserts into the dorsal lip of the distal phalanx, functioning as this last joint's only extensor.

Because of the normal dorsolateral position of the lateral bands, in certain direct injuries to the dorsum of the finger at the PIP joint the lateral bands may subluxate volarly, hyperextending the terminal joint. This is called the *boutonni ère deformity*. When the terminal tendon insertion at the distal interphalangeal joint is avulsed or transected, the distal joint droops and the secondary proximal and dorsal retraction of the lateral bands produces gradual hyperextension at proximal interphalangeal level. This deformity is known as *mallet* or *baseball finger*; it progresses to the *swan-neck deformity* when the PIP hyperextension is added.

The type of injury and the results of surgery vary because of the structural and functional differences in the extensor system from fingertip to forearm. Extensor tendon characteristics have been categorized by eight anatomical zones; the four with odd numbers overlie the joints, and the four with even numbers are the tendon segments between the joints (Fig. [42-53](#)). Generic recommendations for repair methods are illustrated in Fig. [42-54](#).

Zone 1 (Mallet Finger and Secondary Swan-Neck)

Terminal tendon injury may occur by avulsion with or without attached bone fragment, and by transection from laceration or crush (Figs. [42-55](#) and [42-56](#)). Closed injuries can be successfully treated by closed means; open injuries also may be treated by extension splinting. Open or closed injuries that include fracture of the joint surface with secondary palmar subluxation of the distal phalanx require reduction and internal fixation in neutral extension to restore DIP joint congruence and proper tendon relationships. This permits almost immediate joint plus hand rehabilitation while protecting the repaired tendon and joint and making wound care easier. Otherwise, the terminal joint is rarely pinned. Splints that immobilize the DIP articulation in extension (6 to 8 weeks) but leave the PIP joint unrestrained usually are preferred (Fig. [42-57](#)).

Zone 2

Zone 2 is the area over the middle phalanx where the lateral bands fuse to form the terminal tendon. The lateral bands are connected proximally by the thin triangular ligament. Injuries in this area usually are from laceration with resultant mallet deformity. Direct repair and terminal joint pinning for 6 to 8 weeks is appropriate for open cases; closed splinting for closed injuries without significant fracture is effective.

Zone 3 (Boutonnière)

Zone 3 is the area over the PIP joint where the central slip and lateral bands interconnect. Injury may be closed or open and may include avulsion of the central slip, with or without a dorsal bone fragment. The latter injuries require accurate reduction of the joint, bone, and contiguous tendon mechanism. Pinning the PIP joint in extension allows early rehabilitation of the other joints. Untreated, this boutonnière deformity progresses to a fixed PIP joint flexion contracture with secondary hyperextension of the terminal joint. Closed splint or percutaneous pin management may be equally effective for the pure tendon injuries (Fig. [42-58](#)).

Zone 4

The dorsum of the proximal phalanx is covered entirely by the confluent extrinsic extensor tendon and by the two lateral bands arising from tendons of intrinsic muscles. Because of this local anatomy, most tendon injuries in zone 4 are partial and the cut tendon ends do not retract significantly. Only direct inspection can confirm this diagnosis. If interphalangeal joint extension is normal, a partial tendon injury need not be repaired. Splinting the PIP joint for 3 to 4 weeks usually is adequate.

Zone 5

Lacerations and bite wounds are common at the metacarpophalangeal joint. With open injuries in this region, a human or animal bite should be suspected until otherwise disproved. Bite wounds are serious contaminated injuries requiring primary surgical debridement, irrigation of the wound and joint, and aggressive intravenous antibiotics for 24 to 48 h. Patients may be reluctant to seek treatment early, and many deny the mechanism of injury. The incidence of complications is directly related to the delay in treatment. Radiographs are taken to rule out the presence of a foreign body such as a piece of tooth, an intraarticular fracture, or air in the joint, proving contamination. When caused by a bite, the wound is left open and tendon repair is performed secondarily after healing by secondary intention. In a simple laceration, the tendon can be repaired directly. The more dorsal extrinsic extensor tendon and the sagittal band mechanism should be repaired to prevent subluxation of the extensor into the intermetacarpal valley. Closed extensor tendon dislocation is

almost always to the ulnar side of the joint and requires tendon and sagittal fiber reconstruction.

Where direct repair is possible, primary and delayed primary repairs are preferable, using suture techniques described in Fig. 42-54. The wrist is immobilized in 30 degrees or less of extension, and the MP joint is splinted at 60 to 70 degrees or treated with a custom dynamic MP extension splint for early passive motion. The PIP and DIP joints are left free.

Zone 6

Zone 6 covers the dorsum of the hand, which includes the metacarpals and distal carpals, where there are four common extensor tendons to the fingers and two proper tendons, one to the index and one to the little finger. The three wrist extensors insert in this region: extensor carpi radialis longus into the dorsoradial base of the second metacarpal, extensor carpi radialis brevis into the dorsoradial base of the third metacarpal, and extensor carpi ulnaris into the ulnar dorsal edge of the fifth metacarpal. The extensor pollicis longus tendon crosses from proximal to distal in a radial to ulnar line. Single or partial tendon laceration may not produce metacarpophalangeal extension loss, because forces are transmitted through the tendinous interconnections extending from adjacent extensors, such as the juncturae tendinum. In most cases, however, the involved finger lies slightly more flexed or is less able to extend than the others. Paradoxically, the affected finger may flex at the MP and extend at the IP joints when the patient attempts active extension because of loss of the extrinsic extensor tendon without disruption of the intrinsics. The tendons are oval in cross-section and thicker here than distally. Core sutures of the type used in flexor repairs are recommended (see Fig. 42-54). Conventional postrepair treatment was immobilization with the wrist in 45 degrees or more of extension and the fingers in mild MP flexion for about 4 to 6 weeks. This is discouraged, however, because patients can become significantly stiff with this protocol. Instead, the wrist should be kept in about 30 degrees of extension and the metacarpophalangeal joints fitted with proximal phalangeal extensor cuffs with the interphalangeal joints free, allowing active flexion and passive extension. Motivated patients may begin this program on the seventh to tenth postoperative day; others are immobilized for double that time. At least one adjacent finger must be included in the splint. Multiple tendon injuries may necessitate including all fingers in splints throughout the rehabilitation protocol.

Zone 7

Zone 7 is the proximal wrist region under the extensor retinaculum in which the extensors traverse the fibro-osseous tunnel in six synovial compartments. Unlike for the flexor pulleys, repair of the extensor retinaculum is relatively easy to perform, but retinacular release should be avoided when possible. To prevent extensor tendon bowstringing, preservation or reconstruction of some portion (40 percent or more) of the proximal or distal retinaculum is critical. Step-cut retinaculum release more easily allows closure without producing a space too tight to allow gliding of repaired tendon(s). Tendon lacerations in this area often are associated with injuries to the nearby radial or ulnar sensory nerve branches, which should be considered at examination and surgery.

Closed rupture of extensor tendons, usually the extensor pollicis longus tendon, is not rare, especially after Colles' fracture and other injuries to the distal radius. Many extensor pollicis longus ruptures occur several weeks or months after nondisplaced Colles' fractures. It is postulated that fracture hematoma reduces the limited blood supply in the extensor pollicis longus tendon, leading to the attrition rupture. Systemic connective tissue diseases, such as rheumatoid arthritis, and pathologic conditions that produce a sharp bone edge, chronic tenosynovitis, or both, also may contribute to this problem. Attrition ruptures should be repaired by tendon transfer or grafting. Direct repair is almost never possible because of the wide zone of tendon trauma that occurs before breakage.

Zone 8

Trauma to the radial sensory or posterior interosseous nerve branches may occur concomitantly with extensor injuries in the forearm in zone 8. Multiple adjacent muscles or tendons make it difficult to identify individual tendons. The priority of repair is to restore independent wrist and thumb extension and group extension of the fingers. With lacerations at the musculotendinous junctions, the tendons may be seen distally, but proximally their fibrous septa retract into the muscle bellies. For repair at this level, the suture line must include fascia or the intramuscular tendinous septa to prevent pullout and failure of operation. With injuries in the proximal forearm, division of the posterior interosseous nerve alone may produce loss of extensor function by denervation, or it may occur in combination with injury to some or all of the muscles and tendons. After repair, elbow flexion and wrist extension may be needed to reduce tension at the suture line. Thumb extensor injuries are dealt with in manner similar to injuries to the finger.

Tendon Transfers

Tendon transfer is a reconstructive procedure that antedates the twentieth century. Transfers in the upper limb are designed to restore motion in a nonfunctioning part. Tendon transfers are used in isolated peripheral nerve paralysis, for irreparable tendon damage after extensive segmental loss from devastating trauma or in destructive connective tissue diseases such as rheumatoid arthritis, and to rebalance the hand or provide movement to a spastic or paralyzed limb after central nervous system injury or disease. Upper extremity reconstruction by tendon transfer requires careful patient selection and extended therapy supervision, often with preoperative and postoperative rehabilitation protocols.

Nerve Injury

The upper extremity is innervated by the brachial plexus and several sensory branches arising from the plexus and intercostal nerves (Fig. 42-59). Innervation patterns vary, with intercommunications and fiber exchanges within and between nerves. Nerves are composed of axons and associated Schwann cells enclosed in a basement membrane. Thin collagen fibers called endoneurium are immediately outside this basement membrane; the term *endoneurial tube* refers to the axon and Schwann cell composite. Endoneurial tubes are grouped together to form a variable number of fascicles. Perineurium surrounds each fascicle, composed of concentric layers of flattened cells and collagen fibers. The perineurium creates a diffusion barrier against the surrounding environment, providing the peripheral nerve the equivalent of a blood-brain barrier (Fig. 42-60). Surrounding the layers of perineurium is epineurium. Epineurium that fills the space between fascicles is called *internal* and that surrounding the nerve itself is termed *external* or *outer*. The outer layer is composed of collagen and some elastin fibers. The fascicular organization and internal anatomy of the nerve is not constant over its length. The topographic fascicular organization of the nerve rearranges as branches come off these fascicles. The fascicles move to the perimeter within the nerve as the nerve courses toward the periphery. The internal organization of the nerve reflects the location and position of those final branches.

For normal nerve function to be maintained, the peripheral nerve must glide. Focused tension or compression induces local injury, nerve dysfunction, and symptoms. Compression and traction neuropathies result when a nerve is tethered, especially at or near a joint. The ulnar nerve, for example, has a longitudinal excursion of 10 mm or more proximal to the elbow; the median and ulnar nerves glide nearly 15 mm at the wrist. After injury to a nerve, the nerve is at risk for local scar formation adhesions and secondary traction neuritis from tension focused at the site of adherence.

Classification of nerve injuries, as described by Seddon, is as follows:

Neurapraxia describes paralysis/dysfunction in the absence of nerve degeneration. This dysfunction is often of some duration, though recovery is always achieved in a shorter time than would be required after complete transection and nerve degeneration and regeneration. Recovery by definition is invariably complete.

Axonotmesis includes damage to the nerve fibers of a severity that causes complete nerve degeneration. The epineurium and other supporting structures of the nerve are not disrupted, so the internal architecture is relatively well preserved. Spontaneous recovery is the rule, and generally it is of very good quality because the regenerating fascicles are guided into their paths via the intact sheaths. Recovery takes longer than for neurapraxia.

Neurotmesis is when all nerve structures have been divided. Laceration produces neurotmesis, but physical gaps in the nerve may occur even though the epineurial sheath appears in continuity, such as after traction or crush. At the site of damage the nerve will be completely replaced by fibrous tissue, and there is complete loss of anatomic continuity.

Recovery after nerve injury depends on successful reinnervation of sensory or motor end-organs. After denervation, muscles begin to lose their bulk; a loss of cross-sectional area without any loss in muscle fiber count begins within 1 week of denervation. Connective tissue surrounding the muscle undergoes degeneration and thickening. Interstitial fibrosis predominates over time, but passive exercises may delay or prevent this phenomenon. For function to be resumed, motor end-plates must be reinnervated within 18 months of trauma. Sensory end-organs may be usefully reinnervated long after initial injury, but the quality of recovery diminishes with the passage of time.

The result after repair depends on numerous factors: injury level and mechanism, associated bone and soft-tissue loss, residual function, patient compliance and motivation, timing of repair, and supervised rehabilitation. Quantitative postoperative assessment of motor and sensory function should be documented.

Repair should be done with microsutures with the aid of magnification to produce a spatially correct, tension-free suture line. Nerve grafts are used when direct repair after segmental loss or fibrosis would require tension at the repair site. Joint posturing into extreme flexion or extension to decrease tension at the nerve repair site should be avoided; nerve graft is substituted for such destructive splinting maneuvers. Primary or delayed primary repair should be done whenever appropriate conditions allow. The combination of group fascicular and epineurial nonreactive microsutures after identification of the internal topography should produce the best anatomic result. Repairs are protected by relaxed joint posturing for about 3 weeks, and the results of repair are maximized by beginning sensory and motor reeducation after reinnervation (Figs. 42-61 and 42-62).

Vascular Trauma and Replantation

The majority of upper extremity replantation surgery is based on microsurgical technique. Ocular loupes are useful for lower magnification and wide-field dissections and are particularly helpful in preparing the ends of nerves and vessels for repair. However, the operating microscope offers a steady field magnification range of 25× or more. Miniaturized needle holders and instruments, bipolar coagulation, and other modifications are necessary for proper microvessel handling. The typical repair in vessels of 0.5 to 3.0 mm diameter is performed with nonreactive monofilament interrupted sutures to prevent luminal impingement collapse, although microvessel grafts and end-to-side connections also are used (Figs. 42-63 and 42-64). A contaminated soft and bony tissue injury is healed by primary debridement and stabilization of open fractures, repair of extensor and flexor mechanisms, and repair of nerves and vessels (Fig. 42-65).

Each patient should be considered individually on the basis of the location and extent of the injury and the patient's age and condition. Relevant questions include: (1) Can function be obtained from the replanted part? (2) Will that function exceed what can be achieved through amputation closure and possible prosthetic fitting? (3) Will long-term function be improved or compromised by part replantation? (4) Does the potential benefit to the patient outweigh the surgical risks, costs, and loss of productivity?

Segmental, extensive, or multiple-level injuries require repair and reconstruction over an extended area. Neither complete nor near-complete part amputation makes any patient an automatic candidate for revascularization or reattachment. Single finger amputation in the adult, especially at a level proximal to the PIP joint, including both superficialis and profundus tendons and digital nerves is not suitable for replantation in the vast majority of cases. Consideration should be given to replantation for thumb amputations at and proximal to the interphalangeal joint, for single- finger amputations in children, and for partial hand and more proximal wrist, forearm, or arm amputations. In adults over the age of 40 years, repair of the ulnar nerve proximal to the elbow rarely produces a functional result. Crush and avulsion injuries often make it impossible to achieve successful reattachment and revascularization. Reperfusion before tissues are nonviable is essential. Muscle is the most oxygen-sensitive upper-extremity tissue and must be revascularized within 6 h of amputation.

Handling of Amputated Parts

The amputated part should be cleansed under saline solution, wrapped in a saline-moistened gauze, and placed in a plastic bag. The plastic bag containing the part should then be placed on, not packed in, a bed of ice in a suitable container. The part should not be immersed in nonphysiologic solution such as antiseptics or alcohols. The amputated part is never put in dry ice, it is not perfused, and it should not be allowed to freeze.

Preparing the Patient

The patient is stabilized, and a compression dressing is applied to the stump before transport to the replantation center. Intravenous access lines should be started and blood samples drawn while awaiting transportation. If time permits, x-rays of the stump and also of the amputated part can be obtained.

Most replantation centers request that the patient be given intravenous antibiotics, an aspirin suppository (325 mg), and 25 to 50 mL per hour intravenous supplement of low-molecular-weight dextran in dextrose, the latter for antiaggregation platelet effects. Before the patient is transported to the replantation center, someone responsible at the receiving institution must knowingly accept the patient as a candidate for part reattachment. Even if the center has an active replant team, its members may be occupied with another case. Accurate communication by telephone, physician-to-physician, is an essential part of replantation triage. Even if the referring physician is unsure of the patient's status, preparations can still be made for possible transport. When the patient is transported, the patient should be told that referral is for evaluation for possible replantation.

Amputations

(See Chap. 20)

The value of traditional methods to manage amputations is worth emphasis. Digital amputation affects precision pinch and power grip, the latter more significantly if the hand is painful. The treatment principles for fingertip amputation were discussed earlier. It is important that stable soft-tissue and skin coverage over an amputation stump at any level be obtained (Fig. 42-66). If a residual stump is stiff or painful, amputation or amputation revision through the metacarpal can be a functional enhancement. Where a central ray is excised, second-ray transfer or third–fifth intermetacarpal ligament closure preserves and restores metacarpal alignment and the functional contour of the hand (Figs. 42-67 and 42-68).

Complications of Trauma

Compartment Syndrome/Volkmann's Contracture

In acute compartment syndrome, increased fluid pressure in the tissues contained within a fascial space or subcompartment increases to a level that reduces capillary blood flow below that necessary for continued tissue viability. When untreated, continued pressure elevation produces irreversible muscle and nerve damage because of ischemia, with secondary necrosis, fibrosis, contractures, and sensibility deficits or chronic pain. Acute compartment syndrome results from an increase in the volume of fluid within a compartment or limitations on the dimensions of an anatomic compartment. Posttraumatic edema or hemorrhage, hematoma, swelling from infection, or burns increase compartment fluid, as does revascularization. Other causes include venous obstruction and transiently strenuous exercise. Constrictive dressings and casts, excessively tight surgical closure, and prolonged direct limb pressure during unconsciousness with alcohol and drug stupor or during extended surgical procedures add to the limited dimensions of the anatomic compartment.

Acute compartment syndrome is diagnosed clinically but can be confirmed by measurement of intracompartmental tissue pressure. Clinical findings include a swollen, tense, and tender compartment with pain out of proportion to that expected from the originating injury, peripheral sensibility deficits and, finally, motor weakness or paralysis. Pain is accentuated by passive stretch of the affected muscle. Peripheral pulses usually remain intact because systolic arterial pressure usually is well in excess of the dangerously elevated intracompartmental pressure. While blood flow through the major arteries is not impeded, capillary perfusion is compromised by the elevated pressure (30 to 60 mmHg) within the compartment. Pressure measurement devices are confirmatory but not infallible, and in treatment decisions clinical concerns should outweigh specific pressure measurements. Threshold pressure measurements of 30 mmHg or more are consistent with compartment syndrome, and surgical decompression should be prompt. Because tissue perfusion is affected by systemic blood pressure, a lower threshold pressure for fasciotomy should be used in hypotensive patients. While MRI, CT, or ultrasonography may delineate areas of muscle edema or necrosis, these studies do not help with the diagnosis of acute compartment syndrome. Treatment should not be delayed to order and obtain imaging beyond plain radiographs.

Treatment includes removal of all occlusive dressings, wraps, layers, and splints and splitting tight casts and cast padding down to the skin. If symptoms are not rapidly relieved, fasciotomy of the affected areas is required (Figs. 42-69 and 42-70). After surgical decompression, the wounds are left open but dressed to prevent desiccation. Skin closure by direct means or with skin grafting is delayed for 48 to 96 h at a minimum, but may be performed after 5 to 10 days as swelling permits. Hand therapy is started at 48 h.

Neuromas

Neuromas represent a normal physiologic response after nerve injury. All badly injured and severed nerves form neuromas, but only those neuromas that are exposed, superficial, and likely to be impacted become symptomatic. Only sensory fibers develop painful neuromas (Figs. 42-71, 42-72, 42-73, and 42-74). Medical and surgical management of symptomatic neuromas may be difficult, but prevention is more important. Inadvertent injuries to nerves can be avoided during resection, but with amputation end-nerve divisions are at risk for neuroma formation. Divided nerve stumps should be transposed to deep locations, preferably between or within muscle, or into bone when padded tissue is scant. In the fingers, where there is often limited soft-tissue padding, the practice of dividing the nerve under traction and allowing it to retract proximally is not as certain a method as leaving the nerve end long and transposing it to a site where it is less likely to be struck but more likely to be protected.

A symptomatic neuroma is a therapeutic challenge. More than a hundred methods of surgical treatment have been described, but no method is universally successful. The symptomatic neuroma should be identified, isolated, and dissected intact. The scar bulb is kept in continuity with the nerve. The symptomatic nerve and its continuous neuroma are transposed to a deeper, more padded, and often more proximal location, beneath muscle if possible, but within bone when needed. The neuroma bulb is not excised from the nerve because its excision stimulates germination of another neuroma whose contents may not be contained with this secondary procedure.

Reflex Sympathetic Dystrophy

The first clinical description of abnormally exaggerated and prolonged pain after injury is attributed to the Civil War surgeon S. W. Mitchell, who coined the term *causalgia* from the Greek meaning burning pain. Synonyms include inflammatory bone atrophy and Sudeck's atrophy. Reflex sympathetic dystrophy emphasizes the importance of the sympathetic nervous system in posttraumatic pain pathophysiology. Prompt diagnosis and early therapeutic intervention are the most important factors in optimizing clinical and functional outcome. Reflex sympathetic dystrophy is not a disease. It is a complex interaction of physiologic responses initiated by trauma and exacerbated by posttraumatic events. This process is staged by time and inflammatory phase with characteristic changes (Table 42-2) and by descriptive terminology (Table 42-3). The presumptive diagnosis is based on pain, which is often diffuse, burning, and *hyperpathic*, including *allodynia* (pain to light touch), *hyperalgesia* (painful response to nonpainful stimuli), *dysesthesia* (pins and needles following minor stimulus), and *hyperesthesia* (increased sensitivity or pain with nonpainful stimuli). In addition, the clinical diagnosis of reflex sympathetic dystrophy requires at least three of the following: (1) diminished hand function, (2) joint stiffness, (3) atrophic changes (edema, atrophy, or fibrosis), and (4) vasomotor instability or vasomotor disturbance.

Reflex sympathetic dystrophy presents acutely as a hot, swollen, painful, or dysesthetic extremity. A specific precipitating injury, such as a neuroma-in-continuity or an entrapped peripheral nerve, may not become apparent until the acute manifestations are treated and resolve. In the case of nerve entrapment, careful consideration should be given to surgical decompression. Prolonged discomfort or pain-limited motion does *not* automatically mean that reflex sympathetic dystrophy or a reflex sympathetic dystrophy–like syndrome is present. Patients may have secondary soft-tissue and periarticular fibrosis after trauma or surgery and a focally tender scar, but with pain isolated only to that area, not generalized. Disuse from any cause can result in osteopenia.

Patients with reflex sympathetic dystrophy often require chronic treatments, psychologic support, including counseling and medication, and an extended, intensive, and closely monitored therapy program (Table 42-4). Early recognition and treatment prevents secondary stiffness from joint and tendon adhesions.

BURNS

Approximately 2 million people sustain burns that require medical attention annually in the United States, resulting in almost 500,000 emergency department visits. Isolated hand burns can result in severe functional and aesthetic impairment. The direct effect of thermal injury to the skin and the consequences of this trauma to hand function include: (1) heat, (2) edema, (3) decreased circulation, and (4) infection. Systemic problems include burn shock, the requirement for fluid resuscitation, secondary immunologic deficits, proteolysis, and the accumulation of secondary toxins.

Management of the burned hand depends on the depth of burn, its surface area, patient age and reliability, and coexisting systemic conditions. Treatment must be individualized. Patients who are reliable and have an adequate support system may be appropriately treated outside of the hospital. Patients with burns that prevent self-care and individuals who have no personal assistance require a more controlled setting.

Primary treatment should focus on preserving viability of soft tissues by preventing secondary wound dehydration via application of a moist or biologic dressing. Blisters can be aspirated to preserve overlying epithelium as a biologic dressing, but if they are leaking and require debridement, the wound should be covered with an occlusive and nondesiccating composite dressing.

Circulation maintenance requires prevention of hypovolemia and avoidance of mechanical obstruction to circulation from enveloping eschar. Along with clinical examination, the use of Doppler ultrasonography or a wick catheter for compartment pressure measurement is helpful in assessing vascular sufficiency.

Escharotomy or fasciotomy in a burned extremity may be required to prevent secondary ischemic necrosis. Midlateral incisions are used, whether in the forearm or the fingers. The incision must be deep enough to release all burn eschar and to identify normal tissues at the wound depth. If a single midaxis release does not restore circulation to the periphery, both midaxes must be incised. The neurovascular bundles themselves should be avoided to prevent iatrogenic injury or desiccation after unneeded exposure. The edematous hand may develop an acute carpal or ulnar tunnel syndrome.

Infection is prevented by prophylactic systemic antibiotics in the first 2 days to avoid selection of resistant organisms, but the mainstay of antimicrobial prophylaxis treatment is topical. The most frequently used antimicrobial is silver sulfadiazine (Silvadene), which does not penetrate eschar, is not painful, has broad coverage, prevents desiccation, and can be removed with water, saline solution, or hydrotherapy cleansing. Temporary reversible bone marrow suppression and neutropenia may result from extensive, extended use of silver sulfadiazine. In burn wound sepsis, full-thickness wound biopsy cultures allow diagnosis of sepsis and adjustment of appropriate antibiotics. Surgical debridement of infected burns is necessary, but in widespread burn wound sepsis, mortality may result.

Functional restoration is the most important goal. Most hand burns are on the dorsum because of its exposed position. Deep burns to the palm are more rare except with electrical, chemical, and occasional direct contact thermal burns. Bilateral palmar burns or glovelike burns in children should be considered as evidence of possible child abuse and carefully investigated.

Nonoperative treatment allowing spontaneous healing, and early excision with grafting yield the best functional results in appropriately selected patients. Prolonged inflammation diminishes the chances of recovering hand motion. If initial assessments indicate that the burned hand requires more than 2 weeks before skin healing, early tangential excision and skin grafting should be undertaken promptly. If spontaneous epithelialization and wound closure are anticipated within 2 weeks, nonoperative treatment with continuous exercises and splinting is appropriate.

Splinting should be given high priority in an antideformity position. Splinting may be done in positions other than the traditional position of function, that is, wrist extension, MP flexion, and IP extension. In dorsal hand burns, the splint position is wrist extension of 30 degrees or less with maximum MP flexion and full IP extension (Fig. 42-75). To preserve the first web, the thumb should be widely abducted palmarward and flexed slightly at the carpometacarpal joint to a position where the radial border of the hand is flat, i.e., the first metacarpal is positioned almost palmar to the second metacarpal. The goal of splinting is to stretch the healing wound and prevent anatomic distortion by scar tissue that can prevent restoration of hand function. Neglected dorsal burns develop metacarpophalangeal hyperextension and interphalangeal flexion deformities, thumb adduction, and wrist flexion contractures (Fig. 42-76). Hand rehabilitation requires a coordinated approach between surgeon and therapist to assure a maximal and timely recovery. At first, hand therapy is directed toward minimizing edema and preventing deformity. The exercise program starts between the first and fifth posttrauma day to encourage gliding of flexor and extensor tendons and movement of small joints. Therapy eventually progresses to activities of daily living and reintegration into normal life.

Referral of patients requiring postburn reconstruction is made to a specialized center where coverage and therapy techniques are available for hypertrophic scars and contiguous tendon and joint injuries for selected cases.

INFECTION

Bacterial Infection

Skin infections most commonly derive from direct bacterial inoculation. Secondary spread from contiguous sites and hematogenous seeding are less likely. The most common infecting organisms are staphylococcus and streptococcus species; gram-negative, anaerobic, and mixed infections are seen, depending on the inoculation method, e.g., a tooth. Serious, deep infections require hospital admission and extended use of high-dosage intravenous antibiotics. Wound and blood cultures are obtained before antibiotic therapy is started, and adjustments are made as indicated.

Paronychial infections are common. These involve the nail and nail bed, and constitute about 15 percent of hand infections. Occurrence is associated with hangnails, nail biting, finger sucking, and occupations requiring the hands to be damp frequently. Acute infection is always bacterial, creating a localized abscess, but chronic inflammation is most often yeast or fungal, requiring different therapeutic approach (Fig. 42-77).

Herpetic whitlow is an infection of the soft tissues of the distal phalanx or paronychial area by the herpes simplex virus. It is characterized by intense pain and cutaneous vesicles or blisters. The vesicle fluid is clear at first but may become cloudy over a few days. It is important to distinguish this from bacterial infection. Surgical intervention may spread the herpes virus systemically or dispose to local secondary bacterial infection. Only a bacterial abscess needs surgical drainage. Herpetic whitlow is self-limited, generally resolving within 3 to 4 weeks.

Felon is an expanding abscess within the finger pulp, and represents up to one-quarter of hand infections. Felons can also be extremely painful, often reported as throbbing pulp pain. The expanding abscess produces a localized compartment syndrome as a result of the presence of the fibrous septa that normally anchor the pulp skin and subcutaneous tissues to the distal phalanx. Felons usually are caused by penetrating direct trauma producing bacterial inoculation. Untreated felons, like other compartment syndromes, compromise local circulation and produce secondary tissue ischemia and necrosis in addition to septic destruction. In surgical drainage additional injury to the finger pulp should be avoided, but an abscess may already point to a superficial location (Fig. 42-78).

Deep-space palmar infections may occur more often in the immunocompromised, drug abusers, elderly, and neglected populations. These are extremely serious infections with secondary systemic symptoms requiring the combination of extended medical and staged surgical therapy (Fig. 42-79).

Tenosynovitis

Acute pyogenic digital tenosynovitis is most frequently a result of direct penetrating trauma. Kanavel's cardinal signs of tenosynovitis include: (1) fusiform digital swelling, (2) semiflexed digital posture, (3) significant pain from passive extension of the finger, and (4) tenderness along the entire flexor sheath. Proper management for this closed-space tenosynovial abscess is surgical drainage and intravenous antibiotics. A high index of clinical suspicion is required for diagnosis. Aspiration of the sheath will confirm the diagnosis. In early cases, systemic antibiotics alone may be considered, but there must be profound resolution within 12 to 24 h; otherwise, prompt operative drainage is necessary (Fig. 42-80).

Arthritis and Osteomyelitis

Septic arthritis and osteomyelitis result from neglected soft-tissue infection and may occur in the undertreated or unhealthy population. These problems require extended surgical and medical therapies and often multiple, staged salvage or reconstructive procedures.

High-Pressure Injection Injuries

These injuries occur from paint and grease guns, hydraulic lines, and diesel injectors that propel material under pressures of up to 7,000 pounds per square inch. Penetration through skin and along extended tissue planes is the rule rather than the exception. These injuries almost always involve the hands. The severity of injury is related most directly to the nature of the injected material. Paint and solvents that are cytotoxic produce intense inflammation in addition to the trauma from the high-force injury. This prolonged inflammatory phase is sometimes mistaken for infection. Such injuries require immediate mechanical and pulse-lavage debridement in order to prevent extended tissue loss (Figs. 42-81 and 42-82).

Nonbacterial Infection

Nonbacterial infections include tuberculous, mycotic, and similar diseases. Granuloma, a collection of macrophages and histiocytes characteristic of the systemic response to these agents, is diagnostic. A high index of suspicion and a careful history are necessary for accurate diagnosis. Patients present with relatively painless, chronic, indolent soft-tissue problems. The patient may recall a local traumatic event. The correct diagnosis often is appreciated only after months or years. For example, the history of a penetrating fish hook injury weeks or months before a nodular or boggy tenosynovitis in a fisherman could suggest an atypical mycobacterial infection to the suspicious clinician. Tissue biopsy and cultures confirm the diagnosis of specific granulomatous disease. The distinction between superficial and deep tissue involvement is important in tubercular and fungal infections, because superficial infections are treated medically and deep infections require surgical debridement. Mycobacterial infection, including tuberculosis and the atypical mycobacteria originally thought to be just saprophytes, occurs primarily as soft-tissue infection with secondary bone and joint penetration. The pathologic atypical mycobacteria include *Mycobacterium marinum*, *kansasii*, *avium-intracellulare*, *fortuitum*, and *chelonae*. Outside of the United States, the organism most frequently producing deformity and destruction of upper extremity function is *Mycobacterium leprae*, Hansen's bacillus. Hansen's disease should be considered in patients from other nations.

CHRONIC SYNDROMES

Tendinitis

De Quervain's Tenosynovitis

Inflammation of the tendons in the first dorsal compartment and the abductor pollicis longus and extensor pollicis brevis became associated with de Quervain after his 1895 report of five cases. This tendon inflammation is one of the most common causes of pain along the radial side of the wrist and the proximal dorsoradial thumb. The problem is more common in women, and particularly in new mothers because of frequent positioning of the wrist into flexion and thumb in extension while carrying or manipulating their infant. As both tendons pass across the radial styloid process in their tendon sheath, they are subjected to angulation and shear.

Patients present with tenderness over the midaxis of the radial styloid process. Active thumb and wrist motion, particularly active thumb extension and passive flexion with wrist deviation, are painful. There may be crepitus over the tendon sheath, occasional locking, or a secondary ganglionic mass. There also may be secondary irritation of the overlying superficial branches of the radial nerve with paresthesias along the thumb and first web (Fig. 42-83). The tendon inflammation causes restriction of activities and motion. In Finkelstein's test, designed to reproduce symptoms by passive traction on these tendons, the patient grasps the flexed thumb under the fingers of the same hand, and then the examiner actively deviates the wrist ulnarly. The test is positive when symptoms are reproduced by this maneuver.

The initial treatment for de Quervain's tenosynovitis includes rest and custom splint support of the wrist and thumb. Oral nonsteroidal anti-inflammatory agents may sometimes be helpful, but injection with a steroid and local anesthetic combination is best. The injection should not be made into subcutaneous tissues, because it can produce fat atrophy or depigmentation in darker-skinned individuals. Once acute symptoms resolve, splinting may be continued for symptom-inducing activities, such as when mothers care for their infants. As with other tendon inflammations, injection should relieve symptoms adequately and for a minimum of 4 to 6 months to merit repeating.

When symptoms persist or recur, surgery is carried out through a transverse 2-cm incision centered over the thickened radial styloid process. All deeper dissection is done by gentle longitudinal spreading. The branches of the radial sensory nerve are retracted gently and the thickened sheath incised longitudinally. Multiple tendon slips are evident when the abductor pollicis longus is exposed; this tendon is commonly multistranded. The extensor pollicis brevis tendon is in another subcompartment that also must be released separately in up to 90 percent of cases. Many surgeons excise a portion of the sheath; others close it deep to the tendons. Postoperatively, the thumb and wrist are immobilized for 1 week before active motion with decreasing splint protection is started. Resistive exercises begin when they can be tolerated.

Other Wrist Tendinitis

Flexor and extensor sheath tendinitis may occur in any of the dorsal and volar subgroups, including the second through sixth extensor compartments, the flexor carpi radialis tendon tunnel, and at the flexor carpi ulnaris–pisiform and pisotriquetral joint. Chronic inflammation of any tendon compartment is associated with local and radiating symptoms. Passive stretch and resisted active function of the affected tendons reproduces symptoms and confirms the diagnosis in the absence of other clinical or x-ray findings. Flexor tenosynovitis

is not rare after prolonged gripping activities, such as in stringed-instrument players and typists. Extensor compartment inflammation may be found with intense repetitive wrist and finger motion from any cause.

More proximally, intersection syndromes can occur where the extensor carpi radialis longus and extensor carpi radialis brevis tendons cross the first compartment tendons at their musculotendinous junctions. These crossover tendon inflammations may sometimes be accompanied by audible and palpable crepitus. Similar presentations may occur with the extensor pollicis longus and extensor indicis proprius tendons.

Trigger Finger

Chronic flexor tenosynovitis or tenovaginitis occurs most commonly in the middle and ring fingers and in the thumb, and most often in postmenopausal females. Patients may not recall an injury that predates symptoms, but many can describe an episode of prolonged use or forceful trauma with impact or hyperextension that immediately preceded the symptom onset. The snapping phenomenon occurs as the flexor digitorum superficialis and profundus tendons, or the flexor pollicis longus in the case of the thumb, pull through a tight A1 flexor pulley at the proximal edge of the sheath. There is debate as to whether tenosynovial inflammation or pulley thickening is the cause. There is a relative lack of volume in the sheath for the tendon, inducing the crepitus, catching or locking, and pain. Many patients have other associated systemic diseases, such as diabetes, connective tissue disease, or additional sites of tenosynovitis or carpal tunnel syndrome, but the process may be isolated.

Patients presenting with complaints of finger pain, distal palmar pain on grasping, finger catching and locking should be evaluated for systemic and local problems. Thickening of the flexor tendon at the base of the finger, in the distal palm, and palpation of tendon nodule that glides with active finger motion will make the diagnosis. In the locking finger, an audible and palpable snap is noted (Fig. 42-84). Any condition that causes finger stiffness, diminished flexion, or a flexion contracture especially at the PIP joint, snapping or locking can be potentially *misdiagnosed* as trigger finger. Diagnosis is made by careful clinical examination and findings of a localized nodularity and tenderness about the flexor sheath in the distal palm.

Nonsurgical treatment should be offered except for patients with a fixed PIP joint flexion contracture that will not unlock after local anesthetic and steroid injection. Approximately 1 to 2 mL of the mix is injected into the flexor sheath and pulley. At introduction, the needle pierces the skin and tendons at the base of the proximal phalanx/metacarpal head and is then withdrawn slightly. The sheath is aspirated and injection proceeds with gentle pressure. Firm resistance indicates that the needle is in a tendon and needs to be withdrawn or inserted further. Palpable and visible introduction of fluid into the flexor sheath is a successful injection. Most patients improve, but those with continuing and recurring symptoms are candidates for surgery.

Surgery routinely cures the problem, providing long-term relief unless incomplete A1 pulley release, digital nerve injury, or division of the next more distal A2 pulley occurs. With the patient under local anesthesia and with a forearm or elbow tourniquet, a longitudinal incision is made directly over the flexor tendon between the proximal metacarpophalangeal joint flexion crease and the distal palmar crease (Fig. 42-85). The thickened pulley lies directly beneath this skin incision, just dorsal to the fascia and subcutaneous fat layers. The neurovascular bundles are both lateral to the sheath. Longitudinal incision in the sheath is made under direct vision, carefully releasing the proximal and distal extent of the pulley. To assure restoration of unimpeded tendon gliding via active flexion and passive traction, a blunt hook is used. In the thumb, special care is taken to avoid inadvertent division of the radial proper digital nerve, which usually crosses over and is palmar to the flexor tendon proximal to the pulley. Tendon release under direct vision protects this structure. For tendon release in the thumb, an oblique longitudinal or Bruner-style incision is performed (Fig. 42-86).

Direct trauma producing a partial tendon laceration distally may cause a triggering phenomenon in the distal flexor sheath. Systemic diseases that produce tendon nodularity, such as rheumatoid arthritis and gout, may cause similar, but more distal, locking or catching of the tendon. These less common causes require more extended flexor sheath exploration, excision of tendon nodules, and rehabilitation.

Neuropathies

Median Nerve

The median nerve may be compressed anywhere along its course from the cervical roots to the fingertips, but the most common site is within the carpal tunnel, where it is dorsal to the transverse carpal ligament. All anatomic sites of compression must be considered and evaluated in the differential diagnosis of this increasingly common peripheral neuropathy.

Carpal tunnel syndrome results from increased pressure within the rigid carpal canal, producing median nerve ischemia and physiologic dysfunction. Symptoms include paresthesias and numbness in the radial 3½ fingers, burning digital dysesthesias, and, later in the course, hand weakness or awkwardness. Focal wrist and hand pain are not a part of the syndrome, while nocturnal presence of symptoms is a hallmark of this diagnosis. All treatments are designed to reduce pressure within the canal and relieve nerve compression.

The carpal canal serves as a mechanical conduit for the digital flexor tendons. The carpal bones form the walls and floor or dorsal surface of the canal, and the palmar aspect is roofed by the transverse carpal ligament. Tunnel cross-sectional area is dynamic, with the smallest area probably obtaining at the extremes of wrist flexion and extension. There is debate as to the cause of the increased pressure. Some have postulated tenosynovitis, while other studies have shown collagen, amyloid deposits, and edema as causes.

Eighty percent of carpal tunnel patients are over the age of 40 years. The female/male ratio varies from 4:1 for idiopathic cases, but be as low as 1.5:1 with occupational presentation. A direct connection between carpal tunnel syndrome and forceful or repetitive use of the hands has not been conclusively demonstrated. There are reports of low and high instances among workers in manual industries. Such studies have used clinical criteria alone for the diagnosis and have not eliminated avocational or systemic disease-induced causes. The causation is probably multifactorial in most patients. Carpal tunnel syndrome has been associated with endocrine disorders, including diabetes, myxedema, hyperthyroidism, acromegaly, pregnancy, and the postpartum state. Chronic infections and hematologic and autoimmune disorders also are associated with carpal tunnel syndrome. Space-occupying lesions such as lipomas, bone abnormalities of the radius or carpals, posttraumatic edema, and hematomas may induce increased pressure within the canal and compromise median nerve function.

The diagnosis of carpal tunnel syndrome is clinical. Classic symptoms include paresthesias, with a predominance of nocturnal or early morning onset, burning or numbness in the median sensory distribution, and awkwardness in use of the hand. On physical examination, direct digital pressure over the median nerve at the carpal tunnel reproduces symptoms within 30 seconds. In Phalen's maneuver, gravity-induced wrist flexion reproduces symptoms within a minute. When direct percussion of the nerve elicits and reproduces paresthesias in the median distribution, it is a positive Tinel's sign. The application of a pneumatic tourniquet to the upper limb to reproduce digital symptoms is of no value in making this diagnosis. Examination should include objective documentation of sensory and motor loss, the former by threshold testing, including vibration and Semms- Weinstein monofilaments rather than innervation density or two-point discrimination. Examination of motor function includes observation for thenar loss and assessing abductor muscle resistance against force (Fig. 42-87).

Electrophysiologic studies provide important confirmatory and differential diagnostic information. Electrophysiologic studies alone do not form the basis for this diagnosis, but surgery should not be done without electrophysiologic evaluation. Electrophysiologic tests are useful when the diagnosis is difficult or when surgical release is contemplated. Underlying peripheral neuropathies and multifocal compressions that are otherwise unsuspected may be uncovered. Electrical studies provide a baseline for later comparison if the response to surgery is disappointing. Evaluation should include studies of the median nerve as well as of a second nerve in the more symptomatic extremity. Comparison of median and ulnar or of median and radial sensory stimulation values at the wrist is useful in confirming the diagnosis. Studies are not necessarily of prognostic value for the response to surgery.

Routine radiographs including the carpal tunnel view are recommended by the American Academy of Orthopaedic Surgery for evaluation and treatment of carpal tunnel syndrome. Radiographs are evaluated for carpal fractures, arthritis, Kienböck's disease, or other problems that could alter treatment. CT and MRI scans are seldom needed, but basic laboratory studies to screen for endocrine and hematologic disorders are helpful. Predisposing medical diseases, such as thyroid dysfunction or rheumatoid arthritis, should be treated and may frequently improve or resolve the neuropathy without surgery. In pregnancy, carpal tunnel syndrome is treated by salt restriction, wrist splinting, analgesics, and, occasionally, diuretics. Local injection may be needed in the third trimester. Most patients recover within about 6 months of delivery.

For acute posttraumatic carpal tunnel syndrome associated with swelling or hemorrhage, loosening of constrictive bandages and moving the wrist from a position at the extreme of flexion or extension may suffice to reverse or significantly improve symptoms. Pressure studies and early surgery may be appropriate for those who do not respond.

Splints and nonsteroidal anti-inflammatory medications are widely used. Splints should fit comfortably and position the wrist in neutral to minimal extension. Splints are worn at night if nocturnal symptoms are a major complaint. Night splinting may be all that is required. With activity-induced symptoms, daytime splint use during provoking tasks may be needed. Oral nonsteroidal anti-inflammatory agents are helpful, with monitoring for possible gastrointestinal and systemic side effects. Although subclinical vitamin B₆ deficiency is a possible cause of carpal tunnel syndrome, no prospective study has demonstrated the efficacy of pyridoxine, but it is nontoxic. Local steroid injection results in improvement in 80 to 90 percent of patients, but there is gradual deterioration over the next 12 to 24 months. As with other sites, injections should not be repeated more than 2 or 3 times annually. Inadvertent injection directly into the median nerve will worsen symptoms (Fig. 42-88).

Surgical treatment requires complete division of the transverse carpal ligament for the entire length of the carpal tunnel under direct vision. Surgical failure is most often associated with incorrect diagnosis or incomplete ligament division. Internal neurolysis, flexor tenosynovectomy, concomitant ulnar nerve decompression in Guyon's canal, or carpal ligament reconstruction are not indicated with primary release and may be harmful. Open and endoscopic release can effectively divide the transverse carpal ligament and increase canal volume (Figs. 42-89, 42-90, and 42-91). Open release is performed with the patient supine and under tourniquet control. After limb exsanguination and tourniquet inflation, the field is infiltrated with local anesthetic. Intravenous sedation may be used as a supplement. Incision for open release parallels the thenar crease from the distal end of the transverse carpal ligament to almost the distal wrist crease and is made in line with the third web. The skin should be marked before incision. Small twigs of palmar cutaneous nerve branches are identified during subcutaneous dissection and preserved where found crossing the incision. The palmar fascia is split longitudinally, using a small curved hemostat, from the proximal end of the carpal canal moving distally in the most palmar and ulnar quadrant of the canal. The nerve and flexor tendons beneath and radial to the clamp are continuously identified and protected. The motor nerve to the thenar muscles is identified and protected. The ligament incision roughly parallels the ulnar border of the median nerve and leaves a small tissue flap attached to the hook of the hamate, more ulnarly. The wound is inspected to assure complete division of the ligament, release of the contents, and the absence of soft-tissue and bone anomalies. The median nerve is found in the palmaradial portion of the canal, generally adherent to the underside of the ligament. The nerve may have a central-narrowing, hourglass constriction at the site of maximum compression. The surgeon should avoid manipulating the nerve because this induces more intraneural scarring and interferes both with postoperative nerve gliding and ultimate recovery. After tourniquet release and hemostasis, palmar fascia may be closed, but most surgeons close only the skin, using a fine monofilament nonabsorbable suture. The wrist is splinted in slight extension for about 2 weeks before therapy is started.

Endoscopic release was introduced to avoid the morbidity of a palmar scar. Studies suggest a more rapid recovery with equivalent increase in canal volume by this approach. The

single-portal endoscopic decompression is associated with less perioperative discomfort and shorter immobilization and recovery. After endoscopic release, a canvas wrist splint is offered but not required. Therapy usually can be started within the first postoperative week.

The incidence of inadvertent tissue trauma from open and endoscopic methods has not been determined. The best method for avoiding complications is to operate carefully, in a bloodless field, cutting only what can be seen clearly and identified precisely before it is incised. If the patient continues to have symptoms after surgery, appropriate clinical and adjunctive diagnostic investigation should be undertaken. Incomplete ligament division or inaccurate preoperative diagnosis are the most frequent problems, but patients with hidden agendas of a nonanatomic nature may experience prolonged wound discomfort and limited recovery. In such situations, the real value of preoperative electrodiagnostics becomes evident.

Ulnar Nerve

Ulnar nerve compression at the elbow, the cubital tunnel syndrome, has been known for more than a century. It has been called posttraumatic ulnar neuritis and tardy ulnar nerve palsy to emphasize the traumatic causation. Distal compression in the canal of Guyon (the ulnar tunnel) at the wrist is a less common problem and more often caused by a space-occupying lesion or direct trauma. The possible sites of ulnar nerve compression in the fibromuscular groove posterior to the medial epicondyle are summarized in Fig. [42-92](#). All sites of nerve compression must be considered. The differential diagnosis must include medial epicondylitis and its coexistence with ulnar nerve irritation. Some patients have the mechanical problem of a hypermobile or subluxating ulnar nerve. Ulnar traction neuritis with elbow flexion and anterior nerve subluxation reproduce radiating paresthesias in the ulnar two fingers. Patients who have actual motor weakness, and especially the subgroup with intrinsic muscle atrophy and electrophysiologic changes, have a guarded prognosis after delayed decompression.

Patients presenting with medial epicondylitis should be treated for that problem, but the presence of secondary coexistent nerve irritation must be addressed. Those who do not respond to conservative measures as outlined for the carpal tunnel should be treated surgically. At operation, after nerve decompression, the resultant problem of iatrogenic subluxation must be dealt with. Some prefer simultaneous subperiosteal medial epicondylectomy, excising enough of the bony prominence to flatten the skeletal contour of the medial side of the elbow so that flexion/extension does not produce snapping of the nerve over a tissue prominence. Alternatively, submuscular transposition can be performed. There is no clear advantage of one technique over the other in most patients (Fig. [42-93](#)).

Ulnar tunnel decompression at the wrist must include the management of space-occupying lesions which complicate this diagnosis. Ulnar tunnel syndrome is far less frequent than compromise of the median nerve in the carpal tunnel syndrome (Fig. [42-94](#)). Pathologic conditions predisposing to ulnar nerve and artery compression in the ulnar tunnel should be attended to simultaneously, including excision of a hamulus nonunion, ulnar artery revascularization, or removal of ganglia.

Radial Nerve

Radial nerve entrapment may cause sensory symptoms of paresthesias and dysesthesias in the nerve's afferent muscular and cutaneous distributions in the dorsal forearm, wrist, and hand. Symptoms depend on the primary site of nerve irritation or compression, which is most frequently in the proximal forearm. The term *radial tunnel syndrome* describes a compression neuropathy involving the posterior interosseous nerve branch of the radial nerve. The common presentation is aching discomfort in the dorsal and dorsal-lateral forearm. Patients with this problem may have preceding or coexistent lateral epicondylitis, and the distinction between epicondylitis and posterior interosseous nerve compression must be included in the differential workup of all cases of resistant tennis elbow. Unlike epicondylitis, the site of focal tenderness is distal to the epicondyle, over the extensor muscles, at the site where the posterior interosseous nerve passes into the fibromuscular tunnel bounded by the fibrous proximal edge of the superficial heads of the supinator, approximately at the neck of the radius (Fig. [42-95](#)).

Nonsurgical treatment includes rest, activity modification, splint protection, and nonsteroidal anti-inflammatory agents. Injectable steroids do not have a useful role. Electrodiagnostic studies are not helpful because of the deep and variable location of the nerve. Only patients with denervation of the forearm muscles will dependably have electrical changes—and this group should be easy to diagnose clinically before operative decompression.

The radial and posterior interosseous nerves can be decompressed in the proximal forearm and anterolateral elbow region in patients with resistant symptoms. The brachioradialis splitting incision is most efficient (Fig. [42-96](#)).

Radial sensory entrapment distally, Wartenberg's disease or cheiralgia paresthetica, may occur, but usually only after direct trauma to the radial wrist, e.g., after application of handcuffs. Operation is rarely required for this typically transient problem. Radial sensory symptoms of local paresthesias more often coexist with or are secondary to the more common problem of de Quervain's tenosynovitis in the first dorsal compartment. This tendinitis should be excluded in the differential diagnosis and is a more probable reason for nerve irritation than primary entrapment (Fig. [42-97](#)).

ACQUIRED DYSFUNCTION

Dupuytren's Contracture

Although this disorder is associated with the nineteenth-century French surgeon Baron Guillaume Dupuytren, he was not the first to describe it. John Hunter, in 1777, and Sir Astley Cooper, in 1822, described the disease, and Cooper recommended subcutaneous fasciotomy. The pathologic proliferation is primarily of the longitudinal portion of the palmar fascia and its digital expansions palmar and dorsal to the neurovascular bundles (Fig. [42-98](#)). The diagnosis in advanced cases is not difficult, but in early stages the disorder may be confusing. Knuckle pads consisting of fasciotendinous proliferations over the dorsum of the PIP joints, of palmar fascial nodules, or dimpling or pitting of the palmar skin are characteristic findings. While palmar fascial nodules are believed pathognomonic, other masses including, retinacular cysts, tendon nodules, foreign bodies, and trigger finger, occasionally cause confusion. The Dupuytren's fascial nodule does not move with active flexor tendon excursion and is located just deep to the skin and subcutaneous fat. Fascial skin tethering may result in fat bulging on either side of the diseased pretendinous band and cause the skin pits that are characteristic of this pathologic process. Proliferating nodules precede cords, but patients may not present until cords and contractures are present. Joint deformity and contracture is the eventual result of coalescence of nodules with the development of a shortened, pathologic fascial mass. There is no proven relationship to trauma, occupation, handedness, or repetitive use in work or sports.

Dupuytren's contracture is most commonly seen in Caucasian males of Northern European descent who are in their sixth decade or older (Figs. [42-99](#), [42-100](#), and [42-101](#)). Hand dominance or trauma are not causes; the male-to-female ratio varies from 2:1 to 10:1. Dupuytren's contracture is familial and is inherited as an autosomal dominant but with variable penetrance. There are significant associations with a number of diseases and conditions, the most prominent of which are diabetes and alcoholism. HIV may be a risk factor.

There is no effective nonsurgical treatment for Dupuytren's contracture. Operation should be reserved for those whose disease is complicated by contracture. Tender palmar nodules are a transient phenomenon, caused by the coexistence of active cellular proliferation and repetitive daily contact, impact, or load. In patients with such complaints, often truck drivers and others who must grip or lift continuously, padded gloves such as those worn by bicyclists and weight lifters may help. Symptoms resolve within a few weeks or months. Splinting does not prevent later contracture. When the patient can no longer place the hand flat on the table, the Huston tabletop test, operation is indicated. Contracture correction at the MP joint is easier than at the PIP joint, where surgery should proceed when the fixed PIP contracture approaches 45 degrees.

Operating on a patient with Dupuytren's contracture requires a detailed knowledge of normal hand anatomy, palmar fascial structure, and location of the pathology as it applies to the deformity, including pathologic displacement of the neurovascular bundles, which puts them at risk of injury during even the most careful of surgical release procedures (Fig. [42-102](#)). Most patients have the ulnar palm affected first and most significantly. In decreasing order, the fourth, fifth, third, second, and first rays are most frequently involved. The most effective technique is digital fasciectomy as opposed to fasciotomy. The results deteriorate with time, however; surgery does not cure the disease but treats joint deformities, contracture, and dysfunction. Fasciectomy is performed under loupe magnification and exsanguinated pneumatic tourniquet control. Skin flaps are dissected at the level of palmar fascia, preserving the maximum thickness of contiguous skin and subcutaneous tissues to avoid devascularized, over-thinned flaps (Fig. [42-103](#)). Closed fasciotomy is not performed because of the risk of neurovascular injury in this blind release technique.

Postoperative management varies only to a minor degree whether Bruner- style, Z-plasty, or the McCash open-palm incision technique is used. With the open technique, transverse areas at the skin creases minimally affect the recovery protocol. A drain is not necessary. The lessened risk of hematoma and diminished short-term pain afforded by leaving some incisions open and connecting the transverse crease incision via oblique longitudinal incisions avoids the need to dissect under an awning of palmar skin (see Fig. [42-103 D](#)). A voluminous and moderately compressive dressing is applied, supplemented by a palmar, or palmar and dorsal, plaster splint(s) that maintains the wrist in 35 degrees or more of extension and the MP and IP joints in full corrected extension. Therapy is started under close supervision by the end of the first postoperative week. Rehabilitation includes active and passive motion and custom extension splinting of released joints at night. Sutures are removed after 2 weeks, depending on wound healing. Soaking and washing, especially when the McCash technique is used, is more an individual choice than required. In addition to wound infection and skin slough, secondary swelling is a serious but uncommon complication. Prolonged pain leading to reflex sympathetic dystrophy is a difficult problem for patient, therapist, and physician. Digital nerves can be injured during operation no matter how expertly the procedure is performed, but such injury must be recognized and repaired.

Arthritis

Inflammatory Arthropathies

The hand is a mirror of many inflammatory arthropathies, not just gout or rheumatoid arthritis (Figs. [42-104](#), [42-105](#), and [42-106](#) and Table [42-5](#)).

Adult and Juvenile Rheumatoid Arthritis

Primary consideration should be given to arthritis because of worldwide prevalence and the severe disability if untreated or, occasionally, when treated aggressively. Rheumatoid arthritis is a chronic systemic disorder of unknown cause whose major manifestation is inflammatory synovitis with secondary bone and tendon invasion and destruction. There may be late tendon dysfunction through nodularity and locking or scarring, joint subluxation, and pain. In most cases, synovitis deformities are symmetrical. Rheumatoid arthritis affects the elbows, wrists, and metacarpophalangeal joints. Proximal interphalangeal involvement is less common but may be significant in a given patient. On radiographs, the hands and feet show some of the earliest signs of periarticular osteopenia, demineralization. The earliest erosions occur along the radial-palmar aspects of the metacarpal heads, at the proximal phalanges, and in the prestyloid recess of the ulna.

Operative intervention is best limited to patients who, despite medical management, have persistent dysfunction because of pain, stiffness, or instability, or those who have progressively worsened function and increased deformities.

Scleroderma or Systemic Sclerosis

This is a generalized vasculitis affecting the skin, gastrointestinal tract, kidneys, and hands, resulting in thickened, dense, and inelastic skin and connective tissues. Pathologic joint

involvement occurs in up to 80 percent of the patients. Vasculitis and secondary small-joint deformity may combine to produce unstable skin, chronic ulcerations that cannot heal, and secondary infections and painful loss of use (Fig. [42-107](#)).

Psoriasis

This should always be a consideration in the patient with inflammatory arthritis of the hands, particularly with nail deformities and oligoarticular arthritis. Psoriatic arthritis usually affects the distal interphalangeal joints.

Crystal Arthropathies

Crystal arthropathies include gout and pseudogout, which are diagnosed definitively after examination of joint aspiration fluid or biopsy specimen. The serum uric acid levels may be normal even in an acute attack. Most patients with hyperuricemia never have acute gouty arthropathy. Calcium pyrophosphate crystalline inflammation, or pseudogout, often affects the wrist, with chondrocalcinosis classically seen on the posteroanterior wrist radiograph at the prestyloid recess (Fig. [42-108](#)).

Noninflammatory Arthropathies

Noninflammatory arthropathies include osteoarthritis, heritable abnormalities of cartilage production, primary and secondary osteonecrosis or osteomalacia, endocrine-associated articular changes from thyroid, parathyroid, pituitary glands and pancreas, hematologic diseases such as hemophilia and hemoglobinopathies, the collagen storage diseases, and miscellaneous bone, nerve, and other connective tissue pathologies, including amyloid. Sarcoidosis is inflammatory.

Osteoarthritis

Osteoarthritis is the most common upper extremity arthropathy. Although classically defined as noninflammatory, osteoarthritis is a cartilage disease with at least intermittent low-to-moderate levels of inflammation. Its incidence increases with age. There is a significant hereditary component, especially for women. Patients may demonstrate progressive loss of articular cartilage, seen on radiographs first as diminished joint space, with secondary subchondral sclerosis and marginal bone spurs or lipping. Joint enlargement as a result of lipping usually occurs. The prevalence of distal interphalangeal joint nodularity, Heberden's nodes (Fig. [42-109](#)), is up to ten times greater in women, especially for those with a family history. Secondary, posttraumatic, mechanical osteoarthritis is more common in individuals whose occupations expose them to injuries or repetitive load, motion, and impact. The inflammatory variant often affects the hands, particularly the interphalangeal joints, and can be clinically and radiographically aggressive. The interphalangeal joints (particularly the terminal interphalangeal joints of index and thumb), the trapeziometacarpal, thumb basilar joint, the pantrapezial and radioscapoid articulations are most frequently affected. With inflammatory problems due to chronic or progressive synovitis, tendon involvement, secondary joint locking and tendon rupture may contribute to symptoms.

Extensor or flexor tenosynovectomies in the lower forearm, wrist, palm, or digits may be necessary and should be combined with a supervised postoperative therapy program to recover motion (Fig. [42-110](#)). When tendon ruptures occur, the attritional defect in tendon substance and segmental tendon loss prevents direct repair and requires tendon graft or transfer (Fig. [42-111](#)). Tendon subluxation may occur as a result of tendon disease or secondarily from joint involvement deep to that tendon.

Focal small-joint deformities are best treated with arthroplasty, especially in MP joints and for the less active, older patient, or with arthrodesis at selected limited intercarpal and interphalangeal joints (Figs. [42-112](#), [42-113](#), and [42-114](#)). For successful arthrodesis, selection of operative method is not as important as meticulous, precise technique. Stabilized continuous bone contact over the entire surface to be fused, in the presence of good bone stock with durable soft-tissue coverage, produces a positive outcome. Living bone provides the most durable arthrodesis. With removal of all the unsightly, painful, prominent osteophytes about the dorsal, palmar, radial, and ulnar joint margins, the results are excellent.

Thumb stiffness, pain, and malalignment produce marked hand impairment; the problems are far out of proportion to the lesion because of the critical importance of comfortable thumb mobility and stability in precision and power hand use (Figs. [42-115](#) and [42-116](#)). Thumb basilar arthroplasty yields functional, aesthetic results.

CONGENITAL DEFORMITIES

Failures of development, separation, and segmentation and intrauterine injury such as amniotic bands or congenital constriction ring syndrome affect mobility, facility, and self-image. Abnormalities of the shoulder and humerus, elbow, forearm, wrist, and hand produce important but different impairments, and all diminish hand facility to different degrees (Fig. [42-117](#)). Among the most common congenital afflictions in the hand are syndactyly and polydactyly (Fig. [42-118](#)). Consideration of repair should begin when the patient is 3 to 6 months of age.

Congenital trigger thumb may present to the primary pediatric caregiver as a snapping that may or may not be painful, but it often presents as a fixed flexion of the terminal thumb joint. Trigger thumbs are rarely locked in extension. Pathologic findings are localized to the flexor pollicis longus tendon and the proximal annular pulley of the thumb. It is not clear whether the tendon enlargement, known as Notta's node, or thickening of the pulley with relative lessening of the internal diameter of the sheath is the primary pathology. Only 10 to 20 percent are bilateral, at times sequential rather than simultaneous. Other trigger fingers may occur in the infant or young child, but only rarely.

Surgery is conservative management. There is no justification for steroid injection in treating congenital trigger thumb. Children who are diagnosed at 12 months of age may be observed for 6 to 12 months for possible spontaneous correction, because waiting does not compromise outcome. At any age, when fixed flexion deformity of the thumb interphalangeal joint produces secondary metacarpophalangeal joint hyperextension, or when a child over the age of 2 years initially presents with symptomatic locking, surgery to release the proximal flexor pulley is in order. The thickened tendon nodule is not debulked or debrided. None of these children has permanent loss of interphalangeal extension. If an interphalangeal joint cannot be fully extended at surgery, it means that the pulley has not been released completely until proven otherwise.

TUMORS

Principles

Localized masses are common in the hand and upper limb, but most are benign. Most have characteristics that assist in making the diagnosis. The relative rarity of malignant tumors of the musculoskeletal system distal to the elbow can lead to misdiagnosis and undermanagement.

Every mass, particularly those that are atypical in appearance or location, should be diagnosed with staging and imaging procedures leading to careful incisional biopsy. Hand masses tend to present earlier, when smaller, because of their superficial location. Enlarging, symptomatic masses are evaluated with history, laboratory studies, imaging by plain films, ultrasonography, scintigraphy, CT scans, or MRI. Biopsy is the last step in diagnosis, and only very small lesions or lesions that are typical should be excised initially.

Benign Neoplasms

Benign tumors can be subdivided into three categories:

- (1) *Latent Benign.* Tumors arising during childhood may heal spontaneously. Most are well encapsulated, with a clearly defined plane between the tumor capsule and normal surrounding tissue. In bone, the growth process is slow, allowing a margin of mature cortical bone to develop and contain the lesion.
- (2) *Active Benign.* Lesions continue to grow, albeit slowly, and are not self-limited in size or by patient age. The tumor is well encapsulated, but the reactive zone is thicker and less mature than in the preceding category. Within bone, the tumor has an irregular shape that alters the internal or external bone architecture. Surgical management is dictated by determining the grade of the lesion and adequacy of local resection. Operative method is determined by the anatomical setting and the implications for altered musculoskeletal part function.
- (3) *Aggressive Benign.* Lesions do not metastasize but are more difficult to control locally. These lesions do not have clear zones of capsular containment. Nodules or extensions of the tumor may grow out into nearby normal tissue, such as in Dupuytren's contracture. Excision through the reactive zone exposes these tumor projections at the surgical margins, allowing microscopic contamination into unaffected tissue. Failure to fully remove the tumor ensures local recurrence.

Malignant Neoplasms

Surgical staging and treatment for true malignant tumors is outlined in Tables [42-6](#) and [42-7](#).

Specific Tumors

Ganglion

Joint and tendon ganglions are among the most common benign soft-tissue tumor masses in the upper extremity, representing up to 50 to 75 percent of reported tumors. Although potentially located anywhere, the majority of ganglions are in specific sites: the middorsal wrist; the volar radial wrist; the flexor sheath at the metacarpal flexion crease as a seed or pea ganglion that is extremely small but hard and tender; and at the dorsum of the distal interphalangeal joint and nail base. The latter is associated with secondary nail deformity, such as in mucous cyst, particularly in the older population (see Fig. [42-3](#)).

Treatment options include closed rupture impact, hypodermic needle aspiration, and operative excision. Rupture by digital pressure or with a swift blow is unnecessarily traumatic and has little chance of succeeding. Aspiration and steroid instillation may be of value, particularly when the expanding lesion has not been diagnosed or is associated with discomfort. At the dorsal wrist, the most common site of origin is from the scapholunate interosseous ligament, and the occult ganglion may account for a significant amount of dorsal wrist pain, particularly in the female teenage population. Volar ganglions are most commonly situated between the flexor carpi radialis tendon and the radial artery, at or just proximal to the wrist at the radioscapoid joint. Most arise from a radiocarpal or intercarpal capsule. Aspiration is helpful and may be entirely curative for the flexor sheath ganglion that appears as a 3- to 10-mm hard mass at or just distal to the metacarpophalangeal joint flexion crease. Aspiration and injection of the mucous cyst distal interphalangeal joint ganglion is less likely to be curative. Repeated drainage increases the risk of joint contamination.

Surgical excision must include the capsular base origin, sometimes referred to as the stalk or root. Deflating the ganglion during operation by incising it before dissecting is much easier than trying to protect nearby cutaneous nerves or vessels while still avoiding an excessively large skin incision around an inflated cyst.

Giant Cell Tumor of Tendon Sheath

Also known as nodular tenosynovitis, fibroxanthoma, giant cell tumor of synovium, and pigmented villonodular synovitis, it is the most common solid soft-tissue tumor in the hand. It is more frequent in females, and patients are generally between the ages of 30 and 60 years. It presents as a firm, lobular, nontender, slowly enlarging mass in the palm, finger, or

thumb. It is more frequently seen on the palmar surface, given that synovium is present in the fingers only about the flexor tendons and in the joints. Secondary tendon, joint, and skeletal invasion is well known (Fig. 42- 119). Effective treatment requires meticulously complete excision, with care being taken not to injure the neurovascular bundles or the critical flexor sheath pulleys. Recurrences occur in up to 10 percent of patients within 2 years and may occur up to 10 years later, though late recurrences may be new lesions entirely.

Lipoma

Lipomas are benign tumors that contain mature fat cells. They are rare in people under the age of 20 years. Some are multifocal. Tumors usually are asymptomatic but gradually enlarging soft to moderately firm masses. When they arise near a nerve or in a nerve tunnel, they may cause secondary symptoms. Lipomas can be superficial, subcutaneous, or intramuscular, and in the hand they also may be large and deep. Surgical treatment is for diagnosis of the unknown enlarging mass or for improving functional impairment. Recurrences are rare (Fig. 42-120).

Enchondroma

Enchondromas, the most common cartilage lesion of bone, are most frequently found at the small tubular bones in the hand. They can present at any age, but most are found in young adults. Virtually all enchondromas present as pathologic fractures, although a small number may be found as an asymptomatic enlargement of a bone. Radiographs usually are diagnostic (Fig. 42-121). Surgical treatment may be for diagnosis or therapy. Pathologic fractures may heal, but the tumor is unlikely to regress spontaneously with fracture union. It is best to treat the fracture and the tumor together, but some patients prefer to allow the fracture to heal while waiting for the tumor to resolve spontaneously. Treating the lesion and the fracture simultaneously limits the disability to just one interval. Thorough lesional curettage is required, and autogenous bone grafting is appropriate; some bones heal without added graft. Incidental discovery of small asymptomatic enchondromas does not mandate treatment. Aggressive and malignant tumors distal to the elbow are uncommon but are more likely to be of soft-tissue origin.

SURGICAL PRINCIPLES

Skin Preparation

Detergents and solutions assist in mechanically debriding skin and also mechanically decrease the microflora population; however, all can irritate skin, and each has limitations.

Alcohols work primarily through the denaturation of proteins and produce the most rapid reduction in microbial counts. They work against most gram-positive and gram-negative organisms. Alcohols are not sporicidal but are active against many fungi and tuberculosis; they also act against some viruses, including HIV and cytomegalovirus. Alcohols do not have persistent effects.

Hexachlorophene is bactericidal through cell wall destruction and is especially active against gram-positive cocci. It is minimally effective against gram-negative, viral, and fungal organisms. This agent is potentially toxic systemically when absorbed and is not recommended for open wounds.

Iodophors, iodine complexes that irritate the skin less than the iodine/alcohol tincture solutions, are effective against a broad spectrum of gram-positive and gram-negative bacteria, fungi, viruses, and mycobacteria by cell wall penetration and oxidation. Iodophors have almost immediate onset of action, and residual activity declines quickly. Because the iodophors can cause skin irritation, tissue damage, and allergic reactions in some patients, these solutions are not recommended for chronic use on open wounds, because the additional tissue irritation may make the wound more susceptible to late infection.

Chlorhexidine gluconate is a broad-spectrum antibacterial that is effective against many viruses, with limited activity against fungi and the tubercle bacillus. The time of onset of action is intermediate, and residual bactericidal action continues for several hours. Chlorhexidine gluconate alcohol-based solution may provide the added benefit of rapid onset and residual activity with minimal toxicity.

Chloroxylonol or parachlorometaxylonol (PCMX) is active through destruction of microbial cell walls. It has intermediate onset and good activity against gram-positive organisms but only fair activity against gram- negatives, fungi, mycobacterial species, and viruses. It is not toxic and rarely causes skin irritation. It may be a good choice in many situations.

Hair Removal

While the conventional wisdom is that routine hair removal eliminates a potential wound contaminant, most studies reveal that hair removal is not a benign procedure and that close skin shaving increases the risk of postoperative wound infection, with risk rising as the time between the shave and the surgical procedure increases. In a case in which the presence of hair would interfere with wound closure or tissue and skin manipulations, the use of electric clippers or depilatories is preferable, and hair removal should be done at the start of the procedure rather than the day before.

Anesthesia

Regional anesthesia for upper limb surgery offers effective pain control and the avoidance of mental confusion or other side effects from sedatives and general anesthesia (Fig. 42-122). Regional anesthesia is not risk-free or always fully satisfactory; systemic and local reactions may be serious. Appropriate monitoring is mandatory. Forearm or axillary tourniquet is used for most hand surgery, but patients often are not able to tolerate continuous pneumatic tourniquet applications for more than 30 min. Isolated peripheral blocks have more limited usefulness.

Distal peripheral blocks in the upper extremity should always be done without epinephrine added to the anesthetic solution. The injection technique is based on infiltration of anesthetic around the nerve and not directly into nerve substance. Although inadvertent needle entry into nerves is common, without epinephrine in the injection solution and with the use of a fine gauge needle it should present no problem. Should a patient complain of paresthesias, the needle is withdrawn and redirected. Intraneural injection with epinephrine-containing solutions may result in extended intraneural ischemia and secondary fibrosis as well as peripheral vascular compromise, particularly in the digital end-arterial circulation.

Ulnar Nerve Block

Proximal block is among the more useful peripheral techniques (Fig. 42- 123). The ulnar nerve is palpated just posterior to the medial epicondyle and injected with 5 to 8 mL 1% mepivacaine hydrochloride without epinephrine via a 23- to 26-gauge needle. The nerve should not be pinned to the epicondyle with the needle; intense paresthesias elicited from neural perforation warrant immediate withdrawal and redirection.

The ulnar nerve at the wrist is in the volar flexor compartment located dorsal to the flexor carpi ulnaris tendon, and just ulnar to the ulnar artery; both nerve and artery are dorsal to the tendon. The dorsal cutaneous branch of the ulnar nerve has already branched 4 to 8 cm proximal to the ulnar styloid process. Initially deep to the flexor carpi ulnaris tendon, it courses dorsally to exit on its dorsal edge distal to the ulnar styloid process, where it can be blocked separately (Fig. 42-124). A fine-gauge needle is inserted into the skin just dorsal and ulnar to the flexor carpi ulnaris tendon; the needle is aimed palmarward and distally, toward the ring finger into Guyon's canal. The skin concavity for needle entry is dorsal to the flexor carpi ulnaris tendon, and easily palpated and visualized during active wrist flexion and ulnar deviation. After needle entry, paresthesias may be elicited, and 5 mL of anesthetic is injected. Aspiration before injection avoids intraarterial injection. The dorsal branch of the ulnar nerve is blocked with an additional subcutaneous infiltration of 2 to 3 mL after first pulling the needle proximally and then redirecting it dorsal and distal to the ulnar styloid.

Radial Nerve Block

The radial nerve is located between the lateral edge of the biceps and the anterior border of the triceps muscles; motor and sensory components can be anesthetized with injection approximately 4 cm proximal to the lateral epicondyle, where the nerve lies on the humerus in this intermuscular space (Fig. 42-125). The needle is aimed distally and inserted; paresthesias confirm needle and nerve location. The needle is then withdrawn slightly and 7 to 10 mL of anesthetic is injected, with another 2 to 3 mL of anesthetic infiltrated in the subcutaneous plane, which also blocks the lateral antebrachial cutaneous nerve.

The purely sensory superficial radial nerve emerges at the dorsal edge of brachioradialis tendon 4 to 6 cm proximal to the radial styloid process. As it courses distally, the nerve divides into multiple terminal branches. Subcutaneous anesthetic infiltration at the styloid process of the radius with a total of 6 to 8 mL of anesthetic effectively blocks the superficial radial nerve in this region (Fig. 42-126).

Median Nerve Block

Blocking the median nerve is more difficult in the region of the elbow than in the region of the wrist. At the elbow the median nerve is posterior and ulnar to the brachial artery. A fine hypodermic needle is introduced from a location medial to the palpated brachial artery, aiming distally. After aspiration, 8 to 10 mL of anesthetic is injected. If intense paresthesias are elicited, the needle is withdrawn slightly before injection (Fig. 42-127).

The median nerve becomes progressively more superficial as it approaches the wrist, where it lies in the most palmar and radial quadrant of the carpal canal. Proximal to the canal, the nerve is located dorsal and slightly radial to the palmaris longus tendon, and ulnar to the flexor carpi radialis tendon—i.e., between the two tendons. Proper technique avoids the canal contents and the nerve. The most consistent and comfortable skin portal is about 3 cm proximal to the wrist crease. The needle is aimed 30 to 45 degrees dorsal and distal, toward the third web space (Fig. 42-128). Injecting 6 to 8 mL of anesthetic consistently blocks the median nerve. The palmar cutaneous branch of the median and lateral antebrachial cutaneous nerves can be blocked with another 2 to 3 mL of anesthetic injected subcutaneously, aiming more superficially and 45 degrees radially from the same entry point.

Digital Nerve Block

The fingers receive their sensory supply from the common digital nerve branches of the median and ulnar nerves. Digital anesthesia can be achieved by injecting the anesthetic into the looser web tissues about the common digital nerves, which is preferable to a ring block in the base of the finger. The so-called ring block technique risks vascular compromise from volume compression when a solution is injected circumferentially about the base of the finger. Digital anesthetic solution should not include epinephrine, because any resulting digital vessel spasm may compromise finger circulation.

Anesthetic is injected retrograde from the web, advancing about 1 cm proximally into the palm, where 2 mL of anesthetic is injected after aspiration. The needle can be withdrawn and turned into the dorsal subcutaneous tissues of the web to ensure anesthesia of the dorsal branch of the digital nerve with another 1 to 2 mL of anesthesia. The technique is repeated on the opposite side of the finger or sequentially in several digits as needed. No more than 5 to 7 mL total of anesthetic solution should be injected for any one finger with this technique (Fig. 42- 129).

Flexor Sheath Block

Single-digit anesthesia can also be achieved with injection of 2 mL of anesthetic directly into the flexor sheath. A fine hypodermic needle is introduced into the flexor tendon from the palmar side at the level of the distal palm or MP flexion crease. Rapid onset of anesthesia can be achieved. This method has the advantage of a single injection but the disadvantage of sometimes failing to completely anesthetize the dorsal divisions of the proper digital nerves.

Tourniquet

The use of tourniquets dates to Roman times, but the device acquired its name from surgical application in eighteenth-century France, from *tourner*, meaning to turn. Hand surgery is performed using an axillary or forearm pneumatic tourniquet. Fingertip procedures can be done using a digital tourniquet made from a ¼-inch rubber drain hose or with the finger sleeve cut from a sterile surgical glove; the tip of the finger sleeve is pierced, and the sleeve is placed over the patient's finger and rolled proximally, simultaneously exsanguinating and achieving a tourniquet effect. In the absence of proximal anesthetic blockade, the maximum tourniquet time a patient will tolerate is 30 to 60 min.

Exception in the presence of infections and suspected aggressive and malignant tumors, the arm should be exsanguinated before tourniquet inflation; limb elevation may be used for partial exsanguination. Covering the arm with a fabric stockinette before elastic bandage exsanguination reduces skin shear; this is important in patients with delicate skin, those with connective tissue diseases, and those who are on steroids. Axillary and forearm tourniquet cuffs are best applied over cast padding. Nonsterile pneumatic tourniquets should be draped away from the operative field with an occlusive plastic tape or drape distal to the cuff to prevent wicking of antiseptic solution during extremity preparation and risking chemical burn (Fig. [42-130](#)).

Pneumatic tourniquet pressures of 225 to 250 mmHg for adults and 200 mmHg in children are adequate. Patients with very large or obese arms require higher pressures and larger cuffs, as do hypertensive patients, in whom tourniquet pressure should be at least 100 mmHg over systolic blood pressure. Tourniquet time is limited by the most oxygen-sensitive tissue—muscle—and the most oxygen-sensitive organelle—mitochondria. Continuous tourniquet application should not exceed 3 h. In cases in which longer tourniquet times are required, the tourniquet should be deflated after the wound has been dressed temporarily and left deflated for at least 10 min per hour of prior inflation. Tourniquet complications involve not only ischemia in labile distal tissues but also ischemia and direct injury to skin, nerves, and muscles located immediately beneath the tourniquet. Assuming operative tourniquet time of 30 min or more, at tourniquet deflation tissues show relative, reactive hyperemia driven by the tourniquet-induced hypoxia and directly proportional to the time of tourniquet use. This hyperemia may complicate hemostasis.

The tourniquet is deflated before wound closure; deflation should be immediately followed by 7 to 10 min of direct, moderate wound pressure before electrocautery is used. To avoid hematomas, the wound is closed when an acceptably dry field has been achieved. If oozing persists, a suction drain is used to keep dead spaces empty.

Incisions and Exposures

Skin incisions can be linear, curved, or angled. They may be oriented in longitudinal or transverse directions relative to the limb. Ideally, elective wounds are placed to lie in and about the soft-tissue skin creases. Hand incisions are not made perpendicular to joint creases, so that iatrogenic contracture and unsightly scars are prevented (Fig. [42-131](#)).

A sterile skin-marking pen is used to draw out the incisions. Cross-hatching the incision at regular intervals assists in realigning the skin edges for closure (Fig. [42-132](#)). Angles, pedicles, and turns in incisions should not be so narrow as to risk vascular compromise by creating a narrow skin peninsula.

Dressings and Splints

The hand dressing is an intrinsic part of the surgical procedure. The dressing and splint are as important to the outcome as the operation. Application of dressings and splints cannot be delegated without supervision by the responsible surgeon. A poorly applied dressing may destroy or disrupt the intended effect of the operation.

The bottom layer of the dressing should be conforming, nonocclusive, and preferably nonadherent, such as Xeroform or Adaptic. Applied dressing sponges may be dry or moistened for contour. When interdigital dressings are appropriate, a single gauze pad is folded, not twisted, between fingers. The involved fingers or the entire hand is then overwrapped loosely with a Kerlix type of bulky rolled gauze. Padded dorsal or volar splints are applied to maintain the desired position of the operated part.

The generic position for hand immobilization includes splinting the wrist at about 30 degrees of extension, the metacarpophalangeal joints at 70 degrees of flexion, and the interphalangeal joints at 0 to 5 degrees of flexion. The splint is extended to the fingertips, and care is taken to avoid compressing the dressing and splint too tightly and risking circulatory compromise. Fingertips should be exposed for circulation checks in hospital and at home. Hand and arm elevation is encouraged for comfort and for minimizing edema during the first several postinjury and postoperative days. With or without a sling, when the patient is supine, sitting, or walking, the hand is kept at or above the level of the heart.

Postoperative Hand Therapy

Hand therapy is begun early and depends on the specific diagnosis, procedure, and patient. Operative goals include minimizing the time of immobilization, enhancing internal stabilization, preferably with minimal invasion, and allowing early mobilization of skin, joints, and tendons.

Exercises appropriate for the condition and surgery performed are prescribed, and a therapist instructs the patient in these exercises. Exercises should be gentle, not painful, and should take the patient to the limit of potential motion at that time. The therapy program should emphasize soft-tissue mobilization and a decrease of edema. When doing therapy for the hand, mobility in the forearm, elbow, and shoulder should be included, especially in older patients. The use of whirlpools is limited to patients with special needs, such as those with burns and those whose wounds require periodic debridement. Heating the tissues is rarely, if ever, done acutely; ice is often more appropriate for posttraumatic conditions. Use of warm-water or paraffin baths is reserved for chronic conditions of systemic inflammation and periarticular stiffness. After injury, tissue swelling often increases proportionally to heat, worsening the prospects of rehabilitation in those swollen parts.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 43 - Plastic and Reconstructive Surgery

Robert J. Wood
M.J. Jurkiewicz

INTRODUCTION

Plastic surgery is that aspect of the discipline of surgery in which concentration and interest are focused on the restoration of form as well as function. The disorders that the plastic surgeon encounters might be the result of trauma, aging, congenital defects, cancer, or prior surgical procedures. Deformity and its consequences are protean, and hence a plastic surgeon must have versatility, a sure spatial sense, and an appreciation of the profound effects that deformity can have on the human spirit. The foundation for the study of plastic surgery is a thorough education in surgical principles and a detailed knowledge of anatomy, surgical technique, and the biology of wound repair.

The word *plastic* derives from the Greek *plastikos*, meaning to mold. The term *plastic surgery* appeared sporadically in a number of surgical texts through the eighteenth and nineteenth centuries. It was firmly established with the publication of *Plastic Surgery—Its Principles and Practice* in 1919 by John Staige Davis.

Many modern plastic surgical procedures have their roots in ancient times with early practitioners. In the *Sushruta Samhita*, an early text written in the sixth or seventh century b.c. by the practitioner Sushruta, reconstruction of an amputated nose with a pedicled forehead flap and reconstruction of auricular defects with cheek flaps are described. Amputation of the nose remained a common form of punishment in India through the eighteenth century. When this technique was noted by English physicians of the seventeenth and eighteenth centuries it became known as the Indian method of nasal reconstruction. The Roman physician Celsus described operations for the repair of traumatic injuries to the nose, eyelids, ears, and lips in the first century a.d. Another Roman physician, Paulus Aegineta, also described operations for the treatment of facial injuries in the first century a.d.

The Middle Ages produced few known practitioners of reconstructive procedures. The period of the Renaissance marked a resurgence in attempts at reconstructive surgery. During the fifteenth century members of the Branca family in Italy practiced techniques of nasal reconstruction probably derived from Sushruta. Gaspara Tagliacozzi (1545–1599) published *De Curtorum Chirurgia per Insitionem* in 1597. This is believed by many to be the first modern plastic surgery text; it describes the technique of nasal reconstruction using an arm flap.

Baronio described the successful grafting of skin in sheep in 1804. Techniques for human skin grafting were described by Revendin in France, as well as by Ollier and Thiersh in a series of publications during the latter part of the nineteenth century. As with other surgical specialties, great advances in plastic surgery took place during the First and Second World Wars. During the period of the First World War the discipline of plastic surgery became well established by melding techniques of dental surgery, otolaryngology, ophthalmology, and general surgery. In England Sir Harold Gillies, an otolaryngologist, established a center for maxillofacial reconstruction of the many allied military casualties. V. H. Kazanjian is well known for his application of dental techniques to these injuries. In particular, he pioneered the use of prosthetic devices to maintain soft tissue as delayed bony reconstructions were performed. Upon entry of the United States in World War I, Vilray P. Blair of St. Louis became consultant to the surgeon general of the army and organized units to care for soft tissue wounds and maxillofacial reconstruction.

In the period between World Wars I and II academic societies of plastic surgeons were established and the training of plastic surgeons both in the United States and Europe became more formalized. As in World War I, centers for plastic surgery were established during World War II, including centers for reconstructive surgery of the hand. Ever since World War II a unifying theme in the clinical and research work of plastic surgeons has been the transfer of tissue, both autologous and allogeneic. The biology of the transplantation and rejection of skin allografts was studied by Gibson and Medawar, who set the stage for the first successful renal transplantation in monozygotic twins by Joseph Murray and coworkers in 1954 at the Peter Bent Brigham Hospital in Boston.

The techniques of moving tissues regionally within a patient were advanced by the development of axial pattern flaps, including the deltopectoral flap by Bakamjian in 1965 and the groin flap by McGregor and Jackson in 1972. Transferring muscle and muscle-skin units had been described in the nineteenth century by Tanzini. Lost and forgotten, the work was rediscovered by McCraw, Orticochea, Ger, Vasconez, Bostwick, and others. Since 1974 a great number of muscle and muscle-skin flaps have been developed.

Also during this time the development of microsurgical techniques by Buncke in the United States, Ackland in the United Kingdom, Harii and Ohmori in Japan, and Taylor in Australia made possible the transfer of tissues from distant donor sites. Among the earliest “free flaps” described were transfer of the omentum and groin flap. By 1990 microsurgical transfer of a great number of muscle as well as skin flaps was commonplace, and replantation of traumatically amputated extremities and digits became a standard of care.

In France, Paul Tessier and coworkers in the 1960s pioneered new techniques for reconstruction of deformities of the craniofacial skeleton. This new discipline of craniofacial surgery was rapidly adopted in the United States, and multidisciplinary centers for the treatment of craniofacial disorders were established.

GENERAL PRINCIPLES

Skin Incisions

Perhaps more than in any other surgical specialty the operative results in plastic surgery are readily apparent, to both the trained and the untrained eye. Although a patient may not have an opinion about the adequacy of resection after colectomy for colon carcinoma, the same patient would almost certainly have a strong opinion about the results of rhinoplasty or the repair of a child’s cleft lip. Hence in any plastic surgery procedure skin incisions are planned to optimize the resultant scar. Langer has described the skin lines of minimal tension that generally run perpendicular to the long axis of underlying muscles (Fig. 43-1). Gravitational forces also work to produce these lines. The lines of minimal tension, or *Langer’s lines*, are very obvious in the face of an older person, in whom they can be seen as the lines of facial expression (Fig. 43-2). An incision planned within or parallel to these lines will result in a narrower and less conspicuous scar than one that runs perpendicular to the lines of minimal tension. Elective incisions may also be placed in areas less obvious to the observer, such as the hairline or eyebrow, or within the eyelids, nares, or mouth.

As the scar matures there is an attendant contraction of the wound. If an incision must cross a joint surface, the direction of the incision should be altered so a linear contracture does not develop that may restrict joint motion. An incision placed across a joint in an oblique or transverse fashion will lengthen the overall scar and change the direction of wound contracture to optimize range of motion.

Excisions of lesions small enough to be closed primarily are frequently performed by an elliptical excision. The long and short axes should be in a ratio of approximately 4:1. Excisions by an ellipse with a long axis that is too short will often result in excess tissue at either end of the closure, or a *dog ear*. Removal of this excess tissue can be performed with a number of techniques, all of which result in a longer but much more acceptable scar. A common technique involves extending the incision in longitudinal fashion and then excising the redundant tissue on one side of the incision (Fig. 43-3).

Wound Closure

Wound healing is a complex process dependent on local and systemic factors as well as surgical technique. Before closing any wound, basic preparation includes debridement of devitalized tissue and removal of foreign bodies and gross wound contaminants. Debridement of devitalized tissue is considered by many to be the single most important factor in the management of a contaminated wound. Nonviable tissue within a wound acts as a culture medium for bacterial growth and also affects leukocyte function by promoting an anaerobic environment within the wound. Tissue viability and the extent of debridement required are best judged by thorough inspection of the wound. In fresh traumatic or surgical wounds skin viability can often be difficult to assess accurately. In these instances the injection of fluorescein dye can be helpful in demarcating devitalized skin. In general, however, viable skin edges and muscle tissue bleed when debrided, and viable muscles contract when stimulated.

Ideally skin should be coapted with minimal or no tension (Fig. 43-4). Excessive tension across a wound can lead to delayed wound healing and a wide scar. The tension across a wound closure can be modulated by undermining the surrounding skin, closing the wound in a layered fashion such that sutures in a deep dermal or subcutaneous plane reduce the tension at the skin edge, and by transposing additional skin into the wound in the form of a flap.

Skin and subcutaneous tissues may be coapted with suture material, staples, or surgical tape. A great variety of suture material, absorbable and nonabsorbable, is available to the surgeon. Common absorbable sutures include surgical gut, chromic gut, polyglactin-910 (Vicryl), polydioxanone (PDS), and polyglycolic acid (Dexon). Plain gut, derived from sheep or beef intestine, loses most of its strength within 7 days. This is commonly used in areas where wound healing is expected to be rapid and suture removal may be difficult. Fine gut suture (6-0) is commonly used for skin closure in the periorbital regions and in facial wounds in children. Chromic gut is treated to resist breakdown by tissue and maintains its tensile strength longer than plain gut. We frequently use this suture within the mouth and nose, where suture removal can be difficult. Tissue reaction to chromic gut can be significant. It is rarely used on the skin of the face. Polyglactin and polydioxanone sutures can be used in subcutaneous tissues. Polyglactin will have lost approximately 70 percent of its tensile strength at 3 weeks, and polydioxanone will maintain approximately 50 percent of its tensile strength at 4 weeks and 25 percent at 6 weeks.

Nonabsorbable sutures include surgical silk, surgical cotton, surgical steel, nylon, polypropylene, and polyethylene terephthalate. Nylon and polypropylene are relatively inert within tissue. These sutures are appropriate for skin closure in both interrupted and running subcuticular “pull-out” fashion, tendon repairs, and microsurgery. A continuous running polypropylene suture in a subcuticular fashion allows the accurate coaption of skin edges and can be left in place without creating suture marks in the skin. A subcuticular polypropylene pull-out suture is our standard for general skin closure. The point where a suture pierces the epidermis will begin to epithelialize at 3 days. This process is responsible for the permanent suture marks sometimes seen after wound repair. To prevent these permanent suture marks, skin sutures on the face should be removed at 3 to 5 days and the wound stabilized with sterile tape.

Skin Grafts

Skin grafting is one of the most powerful techniques in plastic surgery. By transplanting a sheet of epidermis and a variable thickness of dermis a surgeon is able to close wounds that would otherwise require a more complex flap or free tissue transfer. The technique of skin grafting also allows large areas to be resurfaced, as in the treatment of burns. A skin graft is a sheet of skin including epidermis and a variable thickness of dermis that is completely freed from its native blood supply and transplanted to a recipient site to be closed. Skin grafts may be either full thickness, including all epidermis and dermis, or partial (split) thickness, in which the skin is harvested at some level within the dermis (Fig. 43-5). Partial-thickness grafts are frequently harvested at a thickness of 1/1200 to 1/2000 of an inch. Harvesting skin at this level will leave some dermis as well as epidermal appendages such as hair follicles and sweat glands in the donor site. These epidermal appendages then allow regeneration of new epidermis to close the donor site.

A full-thickness graft, because it includes the entire epidermis and dermis at the donor site, requires that the donor site be closed primarily or left open to granulate and contract. In

a full-thickness graft there are no epidermal or dermal remnants to allow epithelialization as in a partial-thickness graft (Fig. 43-6). As a general rule, split-thickness grafts have a greater chance of surviving but will contract up to 40 percent on average. In practice, the type of skin graft is matched to the given clinical situation. Partial-thickness grafts are usually used to resurface burns, as the donor areas will rapidly close and can be reharvested. A full-thickness graft is usually selected to resurface relatively small and critical areas such as an eyelid so that contraction of the grafted tissue is minimal.

There are a number of commonly used donor sites for full-thickness grafts. In selecting a donor site, the surgeon must consider which anatomic location provides the best skin color and texture match to the anatomic area being grafted and the consequences of a scar in the region of the donor site. In general, donor sites that are closest anatomically to the site to be grafted will have a closer match in skin color and texture. Postauricular skin is suitable for grafting the skin of the face. Smaller grafts in this area may be taken and the donor site closed primarily. If necessary the entire non-hair-bearing skin of the postauricular and mastoid region can be harvested and grafted with split-thickness skin, as this area is relatively concealed from a frontal view. The supraclavicular region is another appropriate donor site for a full-thickness graft for the face. This skin will often maintain a pinkish tone to match the face. The antecubital crease and inner aspect of the arm are appropriate donor sites for full-thickness grafts for the hand or finger. These donor sites are usually already prepared in the field during hand cases but have the disadvantage of a frequently prominent scar. Transverse incisions in the area of the wrist flexion crease can be especially troublesome to patients if mistaken by others as self-inflicted or an apparent suicide attempt. The hip flexion crease in the inguinal area is an excellent donor site for full-thickness grafting to the upper extremities and hands. The linear scar usually remains well hidden in a natural skin fold. A limited split-thickness graft may be harvested from the hypothenar region, which gives the best match and texture for the hand. Redundant upper eyelid skin may be harvested with or without underlying orbicularis muscle for full-thickness grafting of the contralateral lid. This is also a well-concealed donor site when the donor incision is oriented within the tarsal fold.

When larger areas are to be grafted, split-thickness skin is frequently harvested from the buttocks, thighs, and abdominal wall. It is desirable whenever possible to keep this donor site in an area that will eventually be hidden by undergarments and bathing suits. The scalp and suprapubic areas provide possible donor sites for split-thickness skin. With the regrowth of hair in the donor site any resultant scar remains well concealed. Thick grafts should be avoided in these areas to prevent alopecia. Scalp donor sites reepithelialize rapidly and permit frequent reharvesting. This technique is useful in treating extensive burns.

Cultured autologous keratinocytes can provide sheets of epithelium that will cover patients who have massive burns. Promising techniques are being developed to provide collagen or dermis matrix to enhance the stability of cultured keratinocytes.

For a skin graft to survive it must be revascularized by the recipient bed. Radiation-damaged tissue and relatively avascular tissue such as tendon or bone are therefore poor recipient sites for skin grafts. For the first 48 h that a skin graft is in contact with the recipient bed the graft survives by a process known as *plasmatic imbibition*. During this time nutrients diffuse from the underlying bed through the surrounding extracellular fluid and directly into the capillaries and cells of the skin graft. During this period fibrin bonds fix the graft to the recipient bed. At 48 h, vascular buds in the bed have made contact with capillaries within the graft, and circulation is established by a process known as *inosculation*. In this process vessels in the graft bed randomly attach to capillaries in the skin graft or recipient site blood vessels grow into the graft itself and within some of the preexisting capillaries. It remains controversial whether neovascular ingrowth or random inosculation of existing vessels is the primary method of revascularization.

For a graft to be successfully revascularized and survive, a number of conditions must be satisfied. As discussed above, as a rule the recipient bed must be well vascularized. An exception to this is the occurrence of the “bridging” phenomenon, in which a small area of skin graft over avascular tissue is perfused via collateral vessels from adjacent vascularized skin graft. In this event only several millimeters of vascularity are bridged and even then may eventually break down.

Wounds containing more than 1×10^5 bacteria per gram of tissue will not support a skin graft. Heavily colonized wounds are also chronic and poorly vascularized. In these instances simple examination of the wound is often not adequate and quantitative bacteriology may be helpful. A rapid slide technique to predict the number of bacteria per gram of tissue homogenate is examined microscopically and the bacteria per gram of tissue reliably determined.

Contact of the skin graft and the recipient bed must be maintained for the skin graft to survive. Any shear force between the graft and its bed will disrupt revascularization. When a skin graft is applied to the extremities, splinting of proximal and distal joints should be considered to limit motion at the site of the skin graft. Compressive dressings also serve to firmly fix the skin graft to its bed. Any collection of fluid, such as serum, blood, or purulent material, between the skin graft and its bed will also prevent revascularization. Here again a compressive dressing such as a tie-over dressing can be helpful. Meshing or “pie crusting” the skin graft with a number of small slits in the graft will permit the egress of fluid beneath the graft into the adjacent dressing. Unless a tie-over dressing is applied, examination of the graft on the second postoperative day is recommended. At this point if any fluid collection is noted beneath the graft, that area can be incised and the fluid evacuated.

Split-thickness skin grafts can be harvested by hand or by using a number of power instruments. A drum dermatome such as the Reese dermatome employs an adhesive-coated cylinder that is applied to the skin. As this cylinder is rotated away from the surface of the skin a blade harvests a sheet of skin in a very controlled manner.

Humby knives and Weck knives use a guarded, disposable blade that harvests skin as the knife is advanced tangentially across the body surface with a back-and-forth movement. Power-driven instruments permit the rapid harvesting of large surface areas. These include the Brown or Zimmer air-driven dermatome (Fig. 43-7) and the Padgett electric dermatome. Full-thickness grafts are usually harvested with a No. 10 or a No. 15 blade by elliptical excision. Any fat is then trimmed from the dermal surface to permit dermal contact with the recipient bed. This can be facilitated by rolling the graft around the surgeon's finger with the dermal side exposed and then using a Metzenbaum-type scissors to excise subcutaneous fat.

A number of mechanical devices are available to mesh skin grafts. All these devices produce multiple uniform slits in the graft. When tension is applied perpendicular to these slits, openings appear in the graft and the effective surface of the area is increased. These openings will eventually epithelialize. Meshed skin grafts allow large surface areas to be resurfaced and are particularly useful in large burns. They contour well and allow the egress of any fluid collection beneath the graft during the first 24 h. Meshed skin grafts contract more than sheet grafts, however, and the aesthetic result is often inferior to that with sheet grafts.

Skin grafts can be secured to the recipient bed by staples, sutures, skin tapes, or the dressing alone. For large burns skin stapling is rapidly done and decreases operating time. We favor fine chromic sutures in the younger child, in whom staple or suture removal can be problematic. Skin grafts are usually dressed with a layer of nonadherent material such as Xeroform or Adaptic next to the graft's surface followed by a layer of sterile absorbent gauze. The dressing is then secured by an occlusive circumferential wrapping of an extremity and tape. A tie-over dressing is helpful for providing compression in areas such as the face where the circumferential wrapping is not possible (Fig. 43-8). Sutures placed circumferentially at the edge of the skin graft are left long and then tied over a standard skin graft dressing beneath a layer of cotton or Dacron fiber. This technique is also used to secure intraoral skin grafts. We substitute a soft dental compound for the fiber bolus in oral tie-over dressings.

The split-thickness graft donor site heals by reepithelialization from epidermal appendages in the dermis. Epithelial cells from sweat glands, hair follicles, and sebaceous glands in the dermis migrate across the surface of the donor area. The donor site is not unlike a partial-thickness burn wound; a donor site from which a thick skin graft has been taken will heal more slowly than one from which a thin graft has been harvested.

The healing process may take from 7 to 21 days, depending on the thickness of the graft. This process is facilitated by a clean, moist environment. Donor sites may be dressed with a layer of Xeroform and Adaptic followed by absorbent gauze or pads. After 24 h the dressing is taken down to the level of nonadherent dressing material. As healing occurs at the donor site the edges of the nonadherent material separate from the underlying skin and may be trimmed. The donor site usually dries after it is left open for 1 or 2 days. We do not favor the use of heat lamps or hair dryers to dry a donor site because it is painful and impedes reepithelialization. The transparent semipermeable dressings, such as Op-Site, are also excellent dressings for donor sites. These have been shown to reduce pain remarkably at the donor site postoperatively. They can be left in place for 1 to 2 weeks as epithelialization proceeds. If an excessive amount of fluid collects under a transparent dressing it may be drawn off with a syringe and the dressing “patched” with another, smaller piece of transparent dressing using sterile technique.

We prefer to take down the skin graft dressing on the second postoperative day and examine the graft for any fluid collection or signs of infection. If the likelihood of these complications is thought to be small the dressing may be left undisturbed for 5 days for a split-thickness graft and 7 to 10 days for a full-thickness graft. Any unexplained fever, local pain, erythema, or purulent drainage necessitates the immediate removal of all dressings.

Infection can convert a donor site from a partial-thickness to full-thickness skin wound. We treat any signs of infection at the donor site or skin graft site aggressively with frequent dressing changes and topical and appropriate systemic antibiotic therapy. Patients who are debilitated, immunosuppressed, or elderly are also at risk for having the donor site convert to a full-thickness injury. When this risk is judged to be high, the donor site can be grafted with a small amount of widely meshed split-thickness skin.

When the skin grafts and donor sites are epithelialized, any commercially available skin lotion containing lanolin or cocoa butter may be used. Compressive garments can control edema and scar hypertrophy at both the skin graft and the donor site.

Lower-extremity skin grafts are maintained in a nondependent position for 7 days postoperatively. Subsequently, activity is slowly increased from leg dangling to non-weight-bearing to full-weight-bearing ambulation over the next 1 to 2 weeks. During this time frequent examination of the graft is essential and compressive garments or wraps are used for the first week. Activity is then increased as tolerated, with frequent monitoring of the graft for any sign of breakdown. Elastic (Ace) bandages are essential for 4 weeks.

A skin graft undergoes contraction in two phases. Primary contraction occurs as the skin graft is lifted from its donor site. In this phase contraction is caused by elastic fibers within the dermis of the skin graft. Thin grafts with less dermis contract less than thick grafts. The secondary phase of contraction is largely a result of contraction of the recipient bed. This phase begins after 7 to 10 days and continues for as long as 6 months. A number of factors alter the extent of the secondary contraction. Thin grafts contract more than thick grafts, and full-thickness grafts undergo little contraction. Meshed grafts or grafts in which only partial take has occurred contract more as the open areas heal by contraction and epithelialization. Soft, mobile recipient beds tend to contract more than fixed, rigid beds. For instance, skin grafts placed on flexor surfaces often yield troublesome contractures.

Skin grafts, especially those taken below the clavicle, often become darker as they mature. This is often more pronounced with thinner grafts. To minimize hyperpigmentation skin grafts should be protected from ultraviolet light by clothing or sunscreen for a full year postoperatively.

The return of sensation to a skin graft is dependent on the recipient bed and its anatomic location. Heavily scarred beds or beds with poor native sensibility, such as granulation tissue over bone or periosteum, allow minimal regeneration of peripheral nerves into the graft. Two-point discrimination in a skin graft, however, will approach that of surrounding skin if regenerating nerves are not impeded at the level of the recipient bed. The return of sensation can be considered maximal at 2 years.

Skin grafts may be effectively banked by immediately replacing them on their donor site. Grafts stored in this manner may be reharvested easily up to 48 h later. Skin grafts may

also be banked by placing them in saline- soaked gauze sponges and storing them at 4°C. If sterility is maintained, grafts stored in this manner remain viable for up to 21 days.

A composite graft is a free graft containing at least two different tissue elements. The most common composite graft is skin and cartilage transferred to a small wedge taken from the ear. This composite graft is commonly used to reconstruct alar defects of the nose (Fig. 43-9). Common composite grafts include skin and cartilage from the auricular concha, skin and fat from the earlobe, and skin and muscle from the eyelid. Revascularization of a composite graft is similar to that in a split-thickness skin graft except that the large ratio of volume to absorptive surface area precludes nourishment from plasmatic imbibition. The general principles involved in skin graft survival also apply to composite grafts. A well-vascularized recipient bed and meticulous technique, including graft immobilization and hemostasis, are imperative for graft survival. As a general rule, all points within the composite graft should be less than 1 cm from the recipient bed, and grafts with a diameter of less than 1.5 cm are safest.

Flaps

A flap is a tongue of tissue transferred from one anatomic site to another. Vascularity of the transferred tissue is maintained by nutrient vessels within the flap pedicle. The pedicle may either remain attached at its origin or be divided during transfer and reanastomosed to recipient vessels using microsurgical techniques. Microsurgical transfer of tissue is also known as a *free flap*. Flaps are useful for closing defects too large for primary closure and where skin grafting is inadequate. Examples include exposed structures such as brain, blood vessels, bone and joint surfaces, and wounds with poor vascularity, where skin grafting would likely fail. Full- thickness defects in the head and neck are often unacceptable in appearance and function when reconstructed with skin grafts. In these instances flaps may provide better form and function. The reconstructive surgeon may transfer functional units of bone, muscle, and neural tissue for reconstruction of complex defects. Flaps, especially those containing muscle, have proved useful in clearing infection at the recipient site. Muscle flaps have proved to be very immunologically active and are a mainstay in the treatment of problem infections such as osteomyelitis, sternal wound infections, and infected prosthetic materials.

Flaps are classified according to their blood supply. A *random flap* receives blood through the dermal and subdermal plexus (Fig. 43-10). The subdermal plexus receives its blood supply from vessels extending from underlying muscles to the skin. These vessels are sometimes referred to as *musculocutaneous perforators* . Skin flaps that receive their blood supply from cutaneous arteries and veins longitudinally oriented within the substance of the flap are known as *axial pattern flaps* (Fig. 43-11). A *fasciocutaneous flap* contains skin and underlying fascia and is supplied by the vascular plexus within the fascia. A *musculocutaneous flap* contains skin, fascia, and muscle and is supplied by perforating vessels from the muscle to the subdermal plexus (Fig. 43-12). Arterialized flaps have a richer vascular supply and allow for reliable transfer of a greater amount of tissue.

Random Flaps

Random flaps receive their blood supply from the subdermal plexus via musculocutaneous perforators at the base of the flap. These flaps are usually categorized by their type of movement. The *advancement flap* is moved into a defect without lateral or rotational movement. To close a rectangular defect, three sides of an adjoining rectangle of tissue can be incised and undermined with a layer of subcutaneous fat. The rectangle of skin is then advanced into the defect. Small rectangles of tissue at the base of the flap, known as Burow's triangles, may be excised to facilitate advancement. A triangle of tissue may also be advanced as a V-Y advancement flap (Fig. 43-13). With this technique a rectangular defect is closed by incising an adjacent triangle of tissue, which is advanced into the defect, closing the secondary defect as a linear incision behind the flap. Advancement flaps generally allow limited movement of tissue. This movement is facilitated in anatomic areas where a relative excess of skin exists or in the elderly patient with loose, mobile skin.

A *rotation flap* is a semicircular flap rotating about a pivot point to close a triangular defect (Fig. 43-14). The secondary defect may then be closed primarily or skin grafted. A small backcut into the base of the flap can facilitate movement of a rotation flap. A *transposition flap* is typically a rectangular flap that rotates about its base to fill an adjacent defect (Fig. 43-15). This allows closure of the defect without undue tension. Here again, a small backcut into the base of the flap may facilitate movement. Any backcut into the base carries some risk of compromising the vascular supply of the flap.

The *Z-plasty* is a particularly useful type of transposition flap (Fig. 43-16). The Z-plasty consists of two triangular flaps, each of which is rotated into the defect left by the other flap. This is frequently used to transfer tissue into a scar or contracture and lengthen it or to reposition a scar within lines of minimal tension. The sides of the two triangles must be equal in length, and the angles may vary between 30 and 90 degrees. For any given angle, longer triangles will transfer more tissue and provide greater lengthening. Larger angles will also transfer more tissue; however, clinical experience has shown that 60 degrees provides optimal lengthening.

The *rhomboid flap*, or Limberg flap, is another common transposition flap design (Fig. 43-17). To close a rhomboid-shaped defect with internal angles of 60 and 120 degrees, a rhomboid-shaped flap of identical dimensions is designed on the side of the defect judged to have the most available tissue. A *bilobed flap* is a series of two transposition flaps (Fig. 43-18). A properly designed bilobed flap is rotated into the defect as a primary flap, and a secondary flap, whose diameter is half that of the primary flap, is rotated into the defect left by the primary flap. The defect left by the secondary flap is closed primarily. For a bilobed flap to be successful the secondary flap should come from an area of relative skin laxity. An *interpolation flap* is similar to a transposition flap in that a tongue of skin and subcutaneous tissue is rotated about an axis. An interpolation flap, however, is transposed across an intervening bridge of tissue. The deltopectoral flap, discussed below, is an example of an interpolation flap.

Axial Flaps

Axial pattern flaps contain cutaneous vessels running in the longitudinal axis of the flap. They are better vascularized and more reliable than random flaps. The possible length of an axial pattern flap is determined by the length of the cutaneous artery and an additional distal area of skin that is randomly supplied by the subdermal plexus. The midline forehead flap is based on supratrochlear vessels and is commonly used for nasal reconstruction. The deltopectoral flap receives blood supply from perforating branches of the internal mammary artery through the second, third, and fourth intercostal spaces (Fig. 43-19). Historically, this flap has been a mainstay in the reconstruction of head and neck defects. The groin flap is based on the superficial circumflex iliac artery and has been used to resurface wounds of the hand and upper extremity (Fig. 43-20). An axial pattern flap can be designed around the radial artery to involve skin and subcutaneous tissues of skin and bone. The radial forearm flap can then be transferred to defects of the upper extremity or transferred with microsurgical techniques to distant sites. The entire greater omentum may be taken as an axial flap based on either the right or left gastroepiploic arteries. The omentum is then available for closing defects of the chest wall and mediastinum (Fig. 43-21). Because of its large surface area, the omentum is particularly useful for extensive wounds such as radiation injuries and sternal wound infections. The omentum readily accepts split- thickness grafts to complete closure.

Fasciocutaneous Flaps

A fasciocutaneous flap consists of skin, subcutaneous tissues, and underlying fascia. Blood supply is derived at the base of the flap from musculocutaneous perforators or direct branches of major arteries. Because this flap includes the vascular plexus immediately superficial to the fascia, it is more reliable than random flaps. Fasciocutaneous flaps are frequently used in the lower extremity. A fasciocutaneous flap can be designed to include perforators from the medial head of the gastrocnemius muscle to close difficult wounds of the middle and proximal thirds of the lower leg. The posterior thigh fasciocutaneous flap is frequently used for closure of ischial pressure sores. This flap includes fascia lata, subcutaneous skin, and tissue of the posterior thigh and includes the descending branch of the inferior gluteal artery. This vessel runs parallel to the posterior cutaneous nerve of the thigh and accounts for the reliability of this flap.

Muscle and Musculocutaneous Flaps

Muscles can be transferred into adjacent defects if their native vascular supply is preserved. The utility of any given muscle flap is limited by the size of the muscle and the length of its vascular pedicle and hence the distance it can be transferred. The defect, both functional and cosmetic, created by the muscle transfer must also be considered. Muscle flaps transfer richly vascularized and very immunologically active tissue into wounds that are ischemic or infected. The bulk of muscle flaps allows contour defects to be resurfaced.

Skin can be transferred with the underlying muscle. Vessels extending directly from the muscle to the overlying skin, known as *musculocutaneous perforators* , account for a significant portion of cutaneous circulation. These perforating vessels define independent vascular territories within the skin that can be transferred as units with muscle. This principle has been rediscovered several times in the past century. Manhot, in 1889, noted that the skin received much of its blood supply from the underlying muscle. Tanzini, in 1906, described the latissimus dorsi myocutaneous unit for breast reconstruction. The concept was again lost until McCraw and colleagues described a number of musculocutaneous flaps and their vascular territories in 1974.

Delay

Partial disruption of the vascular supply to a flap in a preliminary procedure is known as a *delay procedure*. The purpose of a delay procedure is to increase the length of the random portion of the flap that will survive after transfer. In a delay procedure skin incisions defining the flap partially disrupt the subdermal plexus but maintain the primary pedicle. The flap is then replaced in its bed. One week later the flap is lifted out of its native bed and transferred to the recipient site. Clinical experience suggests that delay procedures can increase the amount of tissue a given pedicle can carry. The mechanism of the delay phenomenon remains incompletely understood. There are two general schools of thought. One theory suggests that delay acclimatizes the flap to ischemia, permitting it to survive with less blood flow than would normally be required. Another theory suggests that delay improves vascularity by increasing flow through preexisting vessels, reorganizing the pattern of blood flow to more ischemic areas. In reality, probably both mechanisms contribute to the delay phenomenon.

Free-Tissue Transfer

The development of microsurgery was one of the great advances in plastic surgery of this century. Reconstructive surgeons previously were constrained by the arc of rotation of any given vascular pedicle. Microvascular surgery allows the complete division of a known vascular pedicle and reanastomosis to recipient vessels at a distant site.

Replantation of severed upper extremities was reported by Malt and McKhan in 1964. The first clinical example of free-tissue transfer was reported in 1972, when McLean and Buncke transferred greater omentum to resurface a large scalp defect. Since that time microsurgery has become standard in most large medical centers, and successful reimplantations of severed parts occur on a daily basis. Many series of elective free-tissue transfers report success rates higher than 95 percent. Free-tissue transfer has applications in all areas of reconstruction and is the standard for reconstructing large defects of the head and neck.

In its simplest form microvascular surgery involves the anastomosis of vessels 0.5 to 3.0 mm in diameter in a manner that maintains patency of the anastomosis. This involves not only surgical technique but also controlling the dynamics of the blood flow, the coagulation system, and inflammatory mediators.

Flow through a blood vessel can be described by Poiseuille's law, whereby the volume flow rate through a capillary tube varies directly with the pressure drop and the fourth power of the radius of the tube, and inversely with the length of the tube and the coefficient of viscosity. Clinically blood pressure must be adequately maintained throughout a microvascular procedure. Hypotension or the use of systemic vasoconstrictors to support blood pressure can be disastrous. Blood viscosity is another factor that can be manipulated by the microvascular surgeon by altering hematocrit levels. A hematocrit level between 30 and 35 percent maximizes flow and oxygen-carrying capacity. Poiseuille's law suggests that vessel radius is the single most important factor in determining flow rate. For this reason the largest possible vessels should be selected for anastomosis. When

dealing with very small vessels any problem with technique or vasospasm will likely result in failure of the anastomosis.

The coagulation system is best controlled by avoiding traumatized vessels and by performing an anastomosis that coapts intact intima to intact intima. Microanastomoses that expose subendothelial collagen invite platelet aggregation and activate the coagulation cascade. A traumatized artery responds to mediators of inflammation by contraction of the smooth muscle within the vessel wall. Some vasospasm occurs even with careful dissection of vessels. If vasospasm is persistent, it can be controlled by irrigation of the vessel with a solution of papaverine and lidocaine. The spasm of severely traumatized vessels is often refractory. For this reason microsurgeons prefer to operate out of the "zone of injury," the region near the defect that, although grossly intact, is undergoing the metabolic changes of injury at a cellular level. The mediators of inflammation within the zone of injury can lead to vasoconstriction and platelet aggregation. If a surgeon is unable to perform the anastomosis out of the zone of injury, a pharmacologic agent to modulate the coagulation system can be administered. The most commonly used agents are aspirin, heparin, and dextran. These medications can be given prophylactically before operation or used selectively intraoperatively.

Aspirin blocks prostaglandin synthesis at the level of cyclooxygenase. At lower doses this effect appears to be selective for thromboxane synthesis and platelet aggregation. Aspirin is used in some microsurgical centers as a single preoperative or intraoperative prophylactic dose. Heparin inhibits thrombin production, the production of fibrin from fibrinogen, and the second phase of platelet aggregation. Heparin is usually used selectively in circumstances judged to be unfavorable such as reoperation for thrombosis of a microvascular anastomosis, anastomoses performed with an extensive zone of injury, and replantation. Dextran decreases platelet adhesiveness, alters the structure of fibrin clot to make it more unstable, and enhances blood flow by reducing viscosity and increasing vascular volume. Dextran is frequently given routinely by microsurgeons as an intraoperative bolus at the time of anastomosis, followed by several days of continuous infusion. The role of any agents used to manipulate the coagulation system nevertheless remains poorly defined in microsurgery. Although dextran is still widely used, the trend in elective free flap transfers has been toward less frequent use of anticoagulation therapy. In situations in which mechanical or metabolic factors are not optimal for microvascular anastomosis, such as replantation surgery, extensive zones of injury proximal to the area to be reconstructed, and long periods of ischemia, anticoagulant therapy is commonly used.

Response to Ischemia

Any operation involving free-tissue transfer requires the donor tissue to undergo a variable period of ischemia. Skin and subcutaneous tissue are relatively tolerant of ischemia. Skin can tolerate 24 h of anoxia. Muscle, however, is relatively sensitive to ischemia. Irreversible changes are apparent in muscle tissue after 6 h of warm ischemia. Peripheral nerves are intermediate in their tolerance to ischemia, with irreversible changes appearing at 8 h of warm ischemia. In general those tissues with a low metabolic rate, such as connective tissue, are more resistant to ischemia. Cooling of tissue prolongs viability during ischemia. Traumatically amputated parts should be wrapped in saline-moistened gauze and placed in a plastic bag. The bag should then be packed in ice. When handled in this manner successful digital replantation has been reported after 36 h.

Vascular endothelium appears to be the tissue most sensitive to hypoxia. Changes can be seen in the endothelium after only several minutes of warm ischemia. Hypoxic damage to endothelial cells is thought to be responsible for the "no-reflow phenomenon," in which damage to the sodium pump of endothelial cells leads to cellular swelling of the vascular endothelium and leakage of intravascular fluid to the interstitium. Adenosine triphosphate- depleted red blood cells are less deformable and not able to pass through microcirculation. A combination of endothelial swelling, interstitial edema, and aggregation of rigid red blood cells leads to obstruction of the microcirculation. There is additional evidence that arachidonic acid release by injured endothelial cells and aggregation of granulocytes are also part of the mechanism of the no-reflow phenomenon.

Clinically, the no-reflow phenomenon might be seen in a free muscle flap with an extended ischemic period. Upon microvascular anastomosis, the flap would show progressive decrease in arterial inflow because of increasing arterial resistance within the flap. Finally, despite widely patent anastomoses there would be no arterial inflow or venous outflow. There is no specific treatment for the no-reflow phenomenon at this time. Current research revolves around the use of nonsteroidal anti-inflammatory agents and control of oxygen free radicals.

Microsurgical Technique

Microsurgery involves the use of an operating microscope. This instrument usually includes foot pedals to control focus, a beam splitter with independent objectives to allow the surgeon and assistant to assume positions 180 degrees from each other, a fiberoptic light source, and a solid base to minimize unwanted motion of the microscope during the procedure. Microsurgical instruments perform the same basic functions as macrosurgical instruments, but are particularly small and delicate. They must be handled carefully; simply dropping a microsurgical instrument can damage it beyond repair. A jeweler's forceps is used almost constantly in the surgeon's nondominant hand. This instrument is useful for handling perivascular tissue and sutures. The tips must be precisely aligned, kept free of dried blood and tissue, and meet evenly or suture cannot be reliably handled by the instrument. There are a variety of needle holders available. Most have tips that are angled in relation to the handles. A curved-tip microscissors is useful for dissection. A straight microscissors is use for trimming adventitia of vessels. A vessel dilator is a modified jeweler's forceps with a smooth, tapered, blunt tip. This instrument is helpful for dilating vessel ends for irrigation. Vessel clamps may be either single- or double-approximating clamps. Double-approximating clamps may have a suture holding frame that permits stay sutures to be used during the anastomosis. Closing pressures of microvascular clamps must be below 30 g/mm² to prevent extensive damage to the endothelium. Suture material is usually polypropylene or monofilament nylon. The size of the suture used is dependent on the vessel size and ranges from 8-0 to 11-0. Microneedles are usually tapered needles ranging from 30 to 100mm in diameter.

Vascular Anastomosis

The end-to-end anastomosis is the most common anastomosis performed by the microsurgeon. The vessels to be anastomosed are placed in microclamps and approximated without tension (Fig. 43-22). Adventitia is removed from the vessel ends using the microscissors. The adventitia may be removed as a sleeve by pulling the adventitia off the end of the vessel in the long axis of the vessel and trimming it with straight microscissors. The vessel dilator is then used to dilate the end of the vessel, and the lumen is irrigated with heparinized saline. Frequently a piece of blue or green plastic sheeting is placed behind the microclamps and vessel ends. This prevents adjacent tissue from obstructing the operative field and permits easy visualization of the suture material.

The needle holder is then used to grasp the needle, which is most stable when held perpendicularly to the jaws of the needle holder. The needle should be passed perpendicularly through the vessel wall, released, and regripped in the vessel lumen. The needle should be passed through the tissue along the curve of the needle. With time, it may be difficult to refrain from passing the needle through both vessel ends in one stroke. This is discouraged, however, as needle placement in the second vessel wall is less precise. The suture is then drawn through the vessels until approximately 10 mm is apparent from the vessel end. To tie a knot, the longer end of the suture is grasped with the nondominant-hand forceps approximately 10 mm from the vessel end. A loop is made by winding the dominant-hand forceps around the suture or creating a loop with the opposite-hand forceps or a combination of these movements. The dominant-hand forceps is then passed through the loop, and the free, short end of the suture is grasped. The free end is pulled back through the loop and the knot secured. A second throw of the square knot can be completed by making a loop in the opposite direction. The loop should be formed near the short end so that the distance traveled by the instruments is small. As the second half knot is made, the nondominant-hand forceps never releases the suture. For most purposes, three half knots will be sufficient. The suture is then cut short end first followed by the long end between forceps and knot. In this manner the suture is always controlled and the needle can be retrieved easily by pulling the needle into the field with the nondominant hand.

A second suture may be placed 180 degrees from the first and the anterior wall completed by placing sutures bisecting the space between each suture. Alternatively the triangulation technique may be used with the first three sutures.

On completion of the anterior wall the microclamps are rotated 180 degrees and the anastomosis is inspected. This is an opportunity to note any suture placed in the back wall and again to irrigate any blood or clot within the lumen. The posterior wall of the anastomosis is then completed in an identical manner.

If the discrepancy of lumen size in the vessels to be anastomosed is greater than 2:1, an end-to-side anastomosis should be performed. The same microsurgical techniques also apply to an end-to-side anastomosis. The adventitia is removed from the area of the anastomosis using a curved microscissors. The media of the vessel is grasped with a jeweler's forceps and the arteriotomy or venotomy made with a straight microscissors. It may be necessary to change the microscissors from the one hand to the other to create two clean scissors cuts. If placed properly, the two scissors cuts will make a neat elliptical hole in the vessel wall. The donor vessel should be trimmed at an angle to reduce turbulence. The end of the donor vessel is often trimmed in a convex form to increase the lumen size at the area of anastomosis (Fig. 43-23). The posterior wall is normally completed first. If the anterior wall is completed first it is often difficult to fully visualize the sutures in the posterior wall. Suturing of anastomosis can be facilitated if the donor vessel can be moved across the field as each side of the anastomosis is completed.

Although not universally accepted, continuous-suture techniques can be used in microsurgery. The principal advantage of a continuous-suture technique is that anastomosis can be completed in less time; in addition, it may be the preferred technique in the anastomosis of vessels with different- sized lumens. The usual disadvantages of continuous-suture techniques include purse-stringing, narrowing of the anastomosis, and suture breakage requiring repeat anastomosis.

When the distance between recipient and donor vessels is too large to allow direct coaptation of vessel ends, vein grafts are required. Vein grafts may be used in head and neck reconstruction, where therapeutic neck dissection or irradiation preclude the use of vessels adjacent to the defect. Lower-extremity trauma with extensive zones of injury also may require vein grafts. The vein graft selected should be as close in size to the recipient vessel as possible. Donor sites for vein grafts include the saphenous vein in the leg and the superficial veins in the dorsum of the hand and ventral forearm.

Monitoring Free-Tissue Transfers

The success rate for free-tissue transfer, in most series, is greater than 90 percent. Given the large number of free-tissue transfers performed yearly, this success rate still yields a sizable number of failed grafts. The purpose of any postoperative monitoring system is to identify failing grafts early enough that an appropriate intervention might salvage the graft.

Examination by an experienced physician is the single most reliable test of graft viability. The flap should be examined for temperature, color, capillary refill, and if any doubt remains, response to pinprick. A pale, cool flap with delayed or absent capillary refill indicates arterial obstruction. Flaps that are purplish with rapid capillary refill and prolonged bleeding of dark blood upon needle stick likely have a venous obstruction. Although the clinical examination is usually an accurate assessment of flap viability, it is often not practical to have an experienced physician observe the flap continuously for the first several days postoperatively. Remote monitoring systems include percutaneous Doppler flow monitoring, laser Doppler, temperature probes, transcutaneous measurements of tissue oxygen tension, and intravenous fluorescein. Clinical experience has not indicated any single monitoring system to be superior. The inexpensive, reliable, and simple free-flap monitoring system has yet to be described.

Tissue Expansion

Human tissue has a well-known ability to expand. This is observed during growth and pregnancy as well as in response to pathologic processes such as tumor growth and morbid obesity. Neumann in 1957 described the first attempt at controlled clinical expansion of soft tissue. He successfully resurfaced an ear reconstruction by placing a subcutaneous balloon in the postauricular area and inflating it over 2 months. Radovan, in a series of reports beginning in 1976, described tissue expansion with silicone devices. Tissue

expansion has become an increasingly important part of reconstructive surgery. The technique provides sensate tissue of similar color and texture to tissue adjacent to the defect.

A wide variety of expansion devices are available to the reconstructive surgeon. In its simplest form the device consists of a silicone reservoir attached to a self-sealing injection port via a variable length of silicone tubing. Using a 23-gauge or smaller needle, saline is injected into the injection port and the reservoir is expanded beneath skin and subcutaneous tissue. Ports may be placed external to the skin or in a subcutaneous position. They should be placed in an accessible location and, if subcutaneous, should be easily palpable. Tissue expanders are also currently available with the injection port integrated into the wall of the expander. These devices require accurate localization of the injection port by palpation or a magnetic locator to prevent inadvertent puncture and deflation of the expander.

The biology of tissue expansion has been extensively studied. When expansion forces are applied to skin, collagen fibers are noted to orient parallel to the expansion force and lose their native orientation. In the first week after placement of the tissue expander epidermal thickening can be noted. This rapidly returns to baseline and probably represents postoperative edema. An increase in melanocyte activity is noted during active expansion and may account for the hyperpigmentation sometimes seen during tissue expansion. A decrease in dermal thickness is noted during and after expansion. The population of myofibroblast cells increases, and elastic tissues become more prominent. Elastin fibers serve to align collagen. With excessive expansion forces elastin fibers are disrupted, leaving collagen in an unnatural orientation. Disruption of elastin fibers might account for stria formation and the permanence of expansion when the expander is removed. Skin with a great deal of intact elastin fibers tends to contract more after expansion.

Tissue expansion increases the vascularity of skin. New vessels are seen in the area adjacent to the capsule. Skin flaps raised in expanded tissue behave similarly to delayed flaps. Compared to unexpanded skin, increased survival length of skin flaps can be expected.

A number of mechanisms account for the new skin formed with tissue expansion. When serial stress loads are applied to skin, collagen and elastin fibers align themselves parallel to the stress, as noted above. As this alignment takes place, the skin stretches. This phenomenon is known as *creep*. As an expander is inflated, adjacent skin is recruited into the area of the expander. It is probable that new tissue is also formed during the expansion process. Increased epidermal mitoses are noted during expansion, and collagen synthesis increases in the dermis.

Suction Lipectomy

Suction lipectomy is said to be the most frequently performed aesthetic procedure in the United States. The procedure involves removal of subcutaneous fat via a blunt-tipped cannula attached to a vacuum pump. The cannula is introduced through small stab incisions, and radial passes are made across the anatomic region where fat removal is desired. The ideal patient for suction lipectomy is young, with localized fat deposits that are unresponsive to diet and exercise.

Suction cannulas range from 1.5 to 10 mm in diameter, and the most popular models have three openings set back from the tip at 120 degrees around the axis of the cannula. A vacuum pump able to generate a negative pressure of 1 atm is then connected to the cannula by flexible tubing. The area to be suctioned is usually injected with a solution of 0.25% lidocaine and 1:400,000 epinephrine. Hyaluronidase (Wydase) is sometimes added to improve the diffusion of the injection fluid. A small stab incision is made along Langer's lines, typically in the groin or suprapubic area for abdominal suction and within the gluteal fold for hip and buttock suction. With the vacuum pump on, the cannula is passed in the subcutaneous plane, creating tunnels radiating from the skin incision. This process is repeated through a second incision such that the radiating subcutaneous tunnels intersect one another. As the procedure progresses the area being contoured is constantly palpated. As a general rule the procedure is terminated when less than 2 cm of skin and subcutaneous tissue can be pinched between the index finger and thumb in the extremities and trunk. The wounds are closed with fine absorbable sutures and a compressive garment is worn for 10 days postoperatively.

Suction lipectomy requires close monitoring of the patient's hemodynamic status. For every 150 mL of fat removed, a 1 percent drop in the patient's hematocrit level can be expected. Fluid shifts dictate that crystalloid must replace fat removed in a 3:1 fluid-to-fat ratio. If more than 1500 mL of fat is removed, autologous blood replacement must be considered. Tumescence technique involves the injection of large volumes of dilute lidocaine and epinephrine solution into the subcutaneous tissue before liposuction. This technique decreases transfusion requirements in cases in which large volumes of fat are removed.

The procedure is generally safe and well tolerated. In 1987 the American Society of Plastic and Reconstructive Surgeons ad hoc committee on new procedures reported a 5-year experience with more than 100,000 suction lipectomies. Eleven deaths and nine major complications were reported within this group. Most cases of death or major morbidity result from hypovolemia, overwhelming sepsis, or embolism, including fat emboli. Although rare, these complications underscore the need for appropriate patient selection and intravenous fluid therapy. Most surgeons give prophylactic antibiotics, but available clinical and laboratory data do not favor any specific prophylactic agent for the fat emboli syndrome.

HEAD AND NECK

Cleft Lip

Two theories are generally held concerning the embryogenesis of cleft lip. The classic theory holds that a cleft lip results from failure of the nasomedial and nasolateral process of the embryo to fuse. The mesodermal penetration theory notes that the embryo is an epithelial bilayer in the region of the face until mesoderm migrates between the bilayers, forming the facial processes. Failure of this mesodermal migration results in clefting. Available evidence favors the mesodermal penetration theory in the etiology of cleft lip.

The primary palate is considered to include the lip, the alveolus, and the hard palate to the incisor foramen. The secondary palate includes the hard palate posterior to the incisor foramen and the soft palate. A cleft lip, then, would also be considered a cleft of the primary palate.

The incidence of cleft lip varies across races (approximately 1/1000 for whites and 0.41/1000 for blacks in the United States). The etiology of cleft lip is thought to be multifactorial. Factors that may increase the incidence of cleft lip include increased parental age, drug use and infections during pregnancy, and smoking during pregnancy. Parents with a cleft lip have an increased risk of producing a child with a cleft lip. The chance of producing a cleft-lipped child with one affected parent is approximately 4 percent. If both parents are unaffected but have a single child with a cleft lip, the risk of the second child's being affected is also approximately 4 percent.

A cleft lip may be unilateral or bilateral and may or may not be associated with a cleft palate (Figs. [43-24](#) and [43-25](#)). Cleft lips are termed *complete* if the cleft extends into the nostril floor and *incomplete* if a bridge of tissue connects the central and lateral lip. This tissue bridge is sometimes referred to as *Simonart's band*.

Cleft lips are associated with characteristic nasal deformities. For a unilateral cleft lip this consists of inferior and posterior displacement of the alar cartilage on the cleft side. The maxilla and alar base on the cleft side is also deficient, resulting in posterior displacement of the alar base. In a bilateral cleft lip these nasal deformities are bilateral and the columella is usually short. The nose tends to be broad and vertically short.

Before definitive lip repair is carried out the patient may undergo a course of presurgical oral orthopaedic procedures. In conjunction with an orthodontist, acrylic splints may be fabricated to better align the alveolar segments and effectively narrow the cleft. By adding acrylic or removing it in certain areas of the splint, the orthodontist is able to move the alveolar arches and soft tissues of the lip to facilitate repair. As an alternative to these passive appliances, pin-retained appliances may be used. The appliance is fixed to the palate with metal pins and moves the alveolar segments with a connecting screw that is rotated daily.

Definitive lip repair is carried out when the child's general health and weight permit the safe induction of general anesthesia. The time-honored "rule of tens" is a useful guide to the timing of surgery. According to this rule, lip repair is carried out when the child has attained a weight of 10 lbs, is 10 weeks old, and has a hemoglobin concentration higher than 10 mg/dL. Cleft lip repair is being considered at even earlier ages in otherwise healthy babies.

The cleft lip repair is usually performed under general anesthesia. After marking the proposed flaps, the lip is injected with dilute epinephrine (1:400,000) to facilitate hemostasis. Many different types of cleft lip repair have been proposed. An effective repair must realign the vermilion and cupid's bow of the lip, reconstruct the upper lip and philtrum, and reapproximate the orbicularis oris muscle within the repair. The rotation- advancement repair as described by Millard is the most popular technique for unilateral cleft lip repair (Fig. [43-26](#)). This technique uses a medial rotation flap to realign the vermilion of the lip. The triangular C flap is inset into the defect created by the rotation flap and is used to lengthen the columella. An advancement flap of the lateral side closes the upper lip and nostril sill. At the time of primary lip repair some surgeons will attempt to raise mucoperiosteal flaps within the alveolar cleft. When these flaps are reapproximated across the cleft, a tunnel of periosteum is formed that may support bone growth across the cleft. This technique, known as *gingivoperiosteoplasty*, is facilitated by the use of presurgical oral orthopaedics.

Bilateral clefts of the lip present a great challenge. No technique yet described provides consistently satisfying results. The central portion of the upper lip (prolabium) and maxilla (premaxilla) often project anteriorly away from the lateral lip elements. Repositioning of the premaxilla is carried out preoperatively with an intraoral appliance or with external traction from tapes or head gear. Current techniques for closure of the bilateral cleft lip use the probabium to form the upper central lip, closing the cleft with either a straight-line closure or some application of the Z-plasty principle. The probabium contains no mature muscle, and orbicularis oris muscle continuity must be restored.

Attempts have traditionally been made to protect the cleft lip repair postoperatively. Some surgeons advocate the use of stiff plastic arm cuffs that prevent the child from touching the lip. Feeding by cups and catheter also avoids possible trauma by nippleing. This view, however, is not universally accepted. Breast feeding has not been shown to be deleterious to lip repair.

Correction of the cleft lip nasal deformity may accompany primary lip repair. Proponents of this approach typically mobilize the lower lateral cartilage from the overlying nasal skin and reapproximate the cartilage medially with a suture. Timing of the nasal correction remains controversial. Some authors believe that dissecting the nasal cartilages at the time of lip repair carries the risk of injury to these fragile structures and possible disturbances in nasal growth. Another approach to the cleft lip nasal deformity is to perform a rhinoplasty at the age of five or six, before the child enters school. At this time further revisions in the cleft lip repair are also often completed.

Cleft Palate

The hard palate develops from the fusion of the two lateral palatine processes that meet initially in the region of the incisive foramen and then continue to fuse posteriorly until, by the twelfth week of gestation, the uvula is formed. These processes are initially on either side of the tongue but swing upward into a more horizontal position as the tongue migrates down. Clefting results when the fusion of the two palatal processes is incomplete. Clefts of the lip and palate are more common on the left side, probably because this palatal process is the last to assume a horizontal position. The most common cleft is a cleft uvula. This occurs in approximately 2 percent of the population. The second most common type of cleft palate is a left unilateral cleft palate and left unilateral cleft lip or cleft of both the primary and secondary palates (Fig. [43-27](#)). The incidence of cleft palate alone in the general population is approximately 1/2000.

A unilateral cleft of the primary palate extends from the incisive foramen anteriorly between a canine and an adjacent incisor to the lip. Clefts of the primary palate may be unilateral or bilateral. A cleft of the secondary palate extends from the incisive foramen to the uvula. An incomplete cleft palate does not extend entirely to the incisive foramen, involving mainly the soft palate.

As with cleft lip, the etiology of cleft palate is multifactorial. Animal models have demonstrated that vitamin A, corticosteroids, and phenytoin produce cleft palate when given during pregnancy. Cleft palates commonly are associated with other anomalies at birth.

During speech and swallowing the palate moves both superiorly and posteriorly against the posterior pharyngeal wall to separate the oral pharynx and nasal pharynx. This is known as *velopharyngeal closure*. Effective velopharyngeal closure allows intraoral pressure to be increased for release with certain sounds during speech. A cleft palate allows air to escape through the nose during speech; this is known as hypernasal speech. Because the oral and nasal cavities cannot be effectively partitioned in the presence of a cleft palate, it is also difficult for the patient to create a negative intraoral pressure. This results in an ineffective suck and feeding problems in the infant with cleft palate. Children with cleft palate have an increased incidence of otitis media. This may be related to abnormal insertions of palate musculature and ineffective opening of the eustachian tube. The recurrent otitis media in cleft palate patients can result in hearing loss.

Timing of cleft palate repair remains somewhat controversial. The cleft palate is usually closed when the patient is between 6 and 18 months of age. The dissection involved in cleft palate closure may have a negative effect on facial growth. For this reason many surgeons prefer late closure of the cleft palate. Children who have undergone early closure of their cleft palate, however, tend to develop more normal speech.

Cleft palate closure is performed under general anesthesia using endotracheal intubation. The head is hyperextended, and a retractor such as the Dingman mouthgag is placed to hold the mouth open and retract the tongue and endotracheal tube. The palate is injected with 0.25% lidocaine with 1:400,000 epinephrine solution. Prophylactic intravenous antibiotics are usually administered. A number of techniques are currently used for cleft palate closure. The von Langenbeck palatoplasty is the oldest technique for cleft palate closure still in use today. In this procedure, relaxing incisions are made bilaterally medial to the alveolar ridges, and incisions are made along the margin of the cleft (Fig. 43-28). Bipedicle mucoperiosteal flaps are then raised based on the greater palatine vessels. These flaps are then closed in the midline in two layers. Mucosal flaps may be also raised from the vomer to aid in closure of the nasal layer. Most surgeons believe that it is also important to reconstitute the sling of the levator veli palatini muscle during cleft palate closure. In patients with cleft palate this muscle abnormally inserts along the margins of the cleft. In the normal palate the levator veli palatini muscle extends from the base of the skull to the middle of the palate, where it decussates with the opposite paired muscle to form a continuous sling. This sling moves the palate posteriorly and superiorly during speech and swallowing (Fig. 43-29). Reconstitution of the levator muscle sling is known as an *intravelar veloplasty*. The bare areas of the hard palate are then left for closure by secondary intention or may be closed primarily in some cases. Critics of the von Langenbeck technique note that it does nothing to lengthen the palate to aid in velopharyngeal closure and that fistulas tend to develop anteriorly where a number of suture lines meet.

The Veau-Wardill-Kilner palatoplasty uses a W-shaped relaxing incision to create mucoperiosteal flaps that are then advanced in a V-Y fashion to close and lengthen the palate. This technique also involves a confluence of incisions anteriorly, and many question whether any permanent lengthening of the palate is achieved.

We prefer Furlow's double-opposing Z-plasty technique for most cleft palates in our practice (Fig. 43-30). This operation uses a Z-plasty on both the nasal and the oral side of the palate to lengthen the palate and to realign the levator musculature. With this technique, relaxing incisions may still be required for closure of the hard palate.

Breakdown of the cleft palate closure will result in a palatal fistula. This allows food and fluid to escape into the nose. Air may also pass through the palatal fistula, resulting in speech disturbances. Small fistulas may be closed by raising adjacent flaps of mucoperiosteum. Large palatal fistulas often require flaps of gingiva or tongue tissue.

After cleft palate closure the palate may still not be able to achieve velopharyngeal closure. The incidence of velopharyngeal incompetence varies among studies, depending on the surgical technique and methods of speech evaluation involved. Historically, 20 percent of patients exhibit velopharyngeal insufficiency after a cleft palate closure. The majority of these patients can achieve satisfactory speech with speech therapy. The remainder require a second procedure to lengthen the palate. The most popular surgical technique for correction of velopharyngeal insufficiency is a superiorly based posterior pharyngeal flap inserted into the free edge of the soft palate. Orticochea has described a pharyngoplasty designed to create a dynamic velopharyngeal sphincter. In this technique the palatopharyngeus muscles are raised from the lateral pharynx and joined in the posterior midline with an inferiorly based flap of mucosa and the superior constrictor muscle.

A well-trained speech pathologist is best able to evaluate velopharyngeal incompetence and success after corrective surgery. Other useful adjuncts in diagnosis include radiography and nasal endoscopy.

Craniofacial Surgery

Craniofacial surgery principally involves movements of the craniofacial skeleton for reconstruction of acquired or congenital defects. Many of the procedures commonly performed by craniofacial surgeons today were considered too dangerous or the bony movements too unstable until the pioneering work of Dr. Paul Tessier of Paris. These advances were in part made possible by parallel advances in the fields of anesthesia and critical care. In a series of publications during the late 1960s and 1970s Tessier not only clarified the classification of craniofacial anomalies but also described access to the entire cranium and facial skeleton through less conspicuous incisions, and the bony movements necessary to reconstruct craniofacial anomalies.

Craniosynostosis

Craniosynostosis is the premature closure of one or more cranial sutures. This occurs in approximately 1/1000 births. The etiology of craniosynostosis remains incompletely defined and is probably multifactorial. Some cases of craniosynostosis, particularly craniosynostosis as part of a craniofacial syndrome, may be secondary to an abnormal cranial base. In these cases abnormal forces from the cranial base could be transmitted to the involved suture through ligamentous structures. Isolated cases of craniosynostosis may also be secondary to abnormal extrinsic forces in utero, such as abnormal fetal position, or weak intrinsic growth forces, such as microcephaly.

Premature closure of a cranial suture results in limitation of growth perpendicular to the line of the suture and compensatory growth parallel to the direction of the suture. Complications of craniosynostosis follow from the brain growing within a restricted bony vault. These problems include intracranial hypertension, optic atrophy, and mental retardation. Hydrocephalus may be secondary to a generalized stenosis of the cranial base.

Craniosynostosis is often classified according to the shape of the skull (Fig. 43-31). It is important to remember that these terms refer not to a disease entity but simply to the morphology of the skull. For instance, a child with plagiocephaly does not necessarily have premature fusion of the coronal suture. *Scaphocephaly* refers to increased length of the skull in the sagittal plane with bitemporal narrowing. This skull morphology is seen with isolated synostosis of the sagittal suture. *Trigonocephaly* describes a midline, keel- like prominence of the forehead that is seen with premature closure of the metopic suture. *Plagiocephaly* is an asymmetric obliquity or flattening of one side of the skull. This may be either anterior or posterior. Plagiocephaly is seen with synostosis of the coronal or lambdoid sutures. *Brachycephaly* is shortening of the skull in the anteroposterior direction. This is seen with premature fusion of both coronal sutures. *Turricephaly* presents an excessive skull height and can be seen in pansynostosis.

Craniofacial Dysostosis

Craniofacial dysostosis refers to a variety of syndromes, most involving some degree of midface (maxilla, orbits) deficiency and craniosynostosis. These syndromes involve complex deformities that are a supreme challenge for the craniofacial surgeon. Reconstruction of these craniofacial syndromes is usually staged, with several procedures during childhood and early adolescence.

Crouzon's Syndrome

Crouzon's syndrome is characterized by hypoplasia of the orbits, zygomas, and maxillae and variable craniosynostosis. The coronal suture is more frequently involved. The syndrome is inherited in an autosomal dominant pattern with variable penetrance. The incidence in the general population is 1/25,000. Patients with Crouzon's syndrome can suffer all the complication of craniosynostosis as well as problems related to exorbitism such as corneal exposure. Midface hypoplasia can lead to malocclusion, airway disturbances, and significant facial deformity (Fig. 43-32).

Apert's Syndrome

Apert's syndrome involves a morphology similar to Crouzon's syndrome with syndactyly of the hands. Most cases of Apert's syndrome are sporadic, although autosomal dominant transmission has been reported. Advanced paternal age is thought to be a factor in some cases. The deformities of the midface are generally more severe than in Crouzon's syndrome. There is also a significant incidence of cleft palate.

Treacher Collins Syndrome

Treacher Collins syndrome, also known as *mandibulofacial dysostosis*, is characterized by hypoplasia of structures derived from the first and second branchial arches. Typically this involves hypoplasia or clefting of the zygomas, external ear deformities, and hypoplasia of the mandible. The syndrome has an autosomal dominant inheritance pattern with incomplete penetrance.

Facial Clefts

A number of rare craniofacial clefts have been described. As with cleft lip and palate, these probably represent failure of mesenchymal migration or incomplete fusion of facial processes. A number of confusing classification schemes have been proposed. In 1976 Tessier proposed a classification system that remains in wide use. In Tessier's system clefts are assigned a number based on their position relative to the sagittal midline (Fig. 43-33). Within this classification system, the standard cleft lip is part of clefts 1, 2, and 3, and Treacher Collins syndrome exhibits some expression of clefts 6, 7, and 8.

Craniofacial Techniques

The majority of craniofacial surgery can be performed through a combination of three well-concealed incisions. A buccal sulcus incision can be made within the mouth to provide wide subperiosteal exposure of the anterior maxilla, the piriform aperture, and the anterior zygoma. A subciliary incision placed just beneath the lower eyelashes is also well concealed and provides access to the orbital floor, the medial and lateral orbital walls, and the anterior maxilla (Fig. 43-34). A bicoronal incision extended from ear to ear transversely across the scalp allows the surgeon to expose the entire cranial vault, and by reflecting the bicoronal flap anteriorly and dissecting in a subperiosteal plane, the entire bony orbit can be exposed (Fig. 43-35).

Historically, craniiosynostosis was treated by strip craniectomy in an attempt to create a patent cranial suture. Simple craniectomy fails to address other abnormalities in the cranial vault and has a high rate of reoperation, as the craniectomy site rapidly reossifies. Tessier pioneered the advancement of the superior orbit and frontal bone in one piece. This increases intracranial volume to alleviate increased intracranial pressure and also corrects the brachycephaly commonly seen in craniofacial syndromes or coronal synostosis. At the time of frontal bone advancement, the anterior cranial vault might also be remodeled by shaping and repositioning the bone of the anterior cranium. In a typical patient with coronal synostosis frontal orbital advancement and cranial vault remodeling would be performed as a single procedure before the child is 1 year old (Fig. 43-36). Children with more complex cranial deformities and severe craniiosynostosis syndromes may require repeat frontal bone advancement or further cranial vault remodeling in early childhood. In those patients with midface hypoplasia such as in Apert's and Crouzon's syndromes, a second phase of treatment involves advancement of the midface. By performing osteotomies at the nasofrontal junction, the floor of the orbit, the lateral maxilla, and the pterygomaxillary fissure, the entire midface may be advanced in an anterior direction. This is known as a LeFort III midfacial advancement after the facial fracture of the same name. Advancement of the midface in this manner serves to correct exophthalmos and improve airway difficulties, and it has a powerful effect on facial appearance. The timing of the LeFort III midfacial advancement is also controversial and varies from early childhood to adolescence. Treatment of the patient with a craniofacial syndrome would typically then be completed by a course of orthodontic treatment and jaw surgery, if necessary, in adolescence.

Hypertelorism

Orbital hypertelorism refers to increased distance between the bony orbits. This is seen with midline craniofacial clefts and certain craniofacial syndromes, including Crouzon's and Apert's syndromes. Significant hypertelorism is best managed by an intracranial approach as described by Tessier. In this procedure a bicoronal incision provides exposure for an anterior craniotomy. Osteotomies are then made around the entire orbit and paramedian bone is removed. The orbits are then mobilized and fixed in a more medial position (Fig. 43-37).

Distraction Osteogenesis

Distraction osteogenesis has become a standard part of the craniofacial surgeons' armamentarium. This technique is adapted from the original work of Ilizarov in orthopaedic reconstruction. For some time the healing callus of a surgically created cut in the long bones of extremities are stretched slowly by an extended frame to lengthen the bone. This typically yields 1 mm of lengthening per day of distraction.

After the pioneering work of McCarthy and colleagues, this technique is now widely applied to the mandible of children with hemifacial microsomia and micrognathia. This has proved to be a powerful technique for augmenting the facial bones and even soft tissue of children with severe microsomia.

Orthognathic Surgery

Disproportion of the maxillae and mandible is commonly seen with craniofacial problems. Maxillary hypoplasia is particularly common with cleft lip and palate and may in part be related to the dissection and subsequent scarring of repair. Acquired maxillomandibular disproportion is usually the result of trauma. Maxillomandibular disproportion can result in malocclusion and ineffective chewing. These patients may also have a significant aesthetic deformity.

The patient with maxillomandibular disproportion is best evaluated by team approach. Essential members of this team include a plastic surgeon, an orthodontist, and a prosthodontist. Evaluation includes dental impressions from which plaster models can be made. If surgery is planned, these models can then be cut and segments repositioned during model surgery. This allows the plastic surgeon and the orthodontist to plan accurately what bony movements are necessary to achieve proper occlusion. Cephalograms or standardized frontal and lateral radiographs of the face and skull provide an outline of the skeletal and soft tissue profile. Measurements between standardized points on the radiograph can then be compared to a database of normal values for age and sex. Minor problems of maxillomandibular disproportion and malocclusion can often be handled by orthodontia alone. Significant skeletal disproportion requires both surgical and orthodontic treatment.

Deformities of the jaws may be classified according to their dental and skeletal relationships. Dental occlusal relationships are normally described by Angle's classification. In this classification *mesial* means close to the dental midline, *distal* means away from the dental midline, *buccal* means toward the outer aspect of the dental arch, and *lingual* means toward the inner aspect of the dental arch. In normal occlusion the mesial buccal cusp of the maxillary first molar lies within the buccal groove of the mandibular first molar. This is known as *class I occlusion*. In *class II occlusion* the mesial buccal cusp of the maxillary first molar is found mesial to the buccal groove of the mandibular first molar, as might be seen with an abnormally small mandible. In *class III occlusion* the mesial buccal cusp of the maxillary first molar is found distal or posterior to its normal position. This relationship may be seen with an abnormally large mandible or small maxillae.

Micrognathia refers to an abnormally small mandible. *Retrognathia* refers to a mandible of normal size located in an abnormally posterior position. Both retrognathia and micrognathia would typically present with class II occlusion and the appearance of a weak or recessed chin. Mandibular advancement is normally performed by creating osteotomies in the sagittal plane through both mandibular rami. The distal mandible is then advanced into a more favorable position and plates or screws used to hold the advancement rigidly. This technique carries some risk of injury to the inferior alveolar nerve. Nerve transection has been reported in approximately 2 percent of cases, and up to 15 percent of patients may have some evidence of long-term dysfunction (Fig. 43-38).

Prognathia denotes an abnormally large mandible. These patients will often have class II occlusion. Surgical treatment of mandibular prognathism may also be by a sagittal split of the mandibular rami, with posterior repositioning of the distal mandibular segments. Alternatively a vertical osteotomy can be made from the sigmoid notch to just anterior to the angle of the mandible. The distal mandibular segment can then be positioned posteriorly and lingual to the proximal segment of the mandible. This technique has a lower risk of inferior alveolar nerve injury, but fixation is more difficult. Most patients will be placed in intermaxillary fixation for 4 to 6 weeks after vertical osteotomy of the mandible.

The soft-tissue contour of the mandible can be further altered by bony movement of the symphysis. A genioplasty or horizontal osteotomy of the mandible allows the anterior projection of the chin to be augmented or reduced. By advancing or recessing the mentum, the soft-tissue contour of the chin can be significantly altered without changing dental occlusion. Plates or lag screws are used to rigidly fix the bony movement. Chin augmentation can also be achieved with the use of a small silicone prosthesis.

Maxillary deformities of hyperplasia and hypoplasia are also commonly seen by the craniofacial surgeon. Maxillary hypoplasia commonly presents as maxillary retrognathia and class II or class III occlusion. This is commonly seen in cleft lip and palate patients. Maxillary hyperplasia often presents as vertical maxillary excess, also known as the *long face syndrome*. These patients have excessive gingival exposure when smiling and poor lip competence.

Most maxillary deformities can be approached by the LeFort I osteotomy. An incision is made in the upper buccal sulcus and dissection is carried out in the subperiosteal plane. The surgeon then performs a horizontal osteotomy just above the level of the piriform aperture bilaterally. When additional cuts are completed through the base of the nasal septum, the medial wall of the maxillary sinus, and the pterygomaxillary fissure, the maxilla is then free to be repositioned in three dimensions. This technique is usually performed in patients with permanent dentition. Horizontal osteotomy of the maxilla risks damage of tooth buds in patients with deciduous dentition.

In treating maxillary hypoplasia the surgeon would reposition the maxillae anteriorly and rigidly fix the segments with plates and screws according to the preoperative plan. Treating vertical maxillary excess will require maxillary bone to be removed and the segments repositioned superiorly.

Facial Fractures

Facial fractures are most frequently caused by motor vehicle accidents and physical assaults. Although the specifics of facial fracture treatment have evolved since the advent of craniofacial techniques in the 1980s, the principles remain the same. The patient with facial fracture is often a multiple trauma patient and requires thorough evaluation. Although the facial fracture is rarely a surgical emergency, other associated injuries frequently require emergency care. For instance, the patient with a gunshot wound to the angle of the mandible is first and foremost a patient with a penetrating neck wound, and the patient with a depressed supraorbital fracture is first and foremost a patient with closed head trauma.

Physical Examination

Examination for facial trauma must be systematic and complete (Fig. 43-39). Proceeding from superior to inferior, the forehead and supraorbital ridge is palpated for any fracture. The orbital rims are also evaluated for any stepoff or irregularity. The ophthalmologic examination must include pupillary responses, extraocular muscle function, and visual acuity. Any enophthalmos, inferior displacement of the globe, or limitation in extraocular muscle function is a sign of orbital fracture. Any patient with signs of direct injury to the ocular globe or an orbital fracture should have a formal ophthalmologic examination. The nose is examined for any gross deformity, the septum is examined for hematoma, and patency of the airway is ascertained. The zygomatic arches are palpated for fractures. Depressed fractures of the zygomatic arch can impinge on the temporalis muscle. It is important to note whether there is pain associated with movement of the mandible or any limitation in mandibular excursion. Fractures of the maxilla frequently run through the infraorbital foramen, and these patients frequently have paresthesia of the upper lip and teeth. Midface stability is assessed by grasping the upper incisors and alveolar ridge with one hand and palpating the nasofrontal junction and anterior maxilla with the other. An attempt is then made to gently rock the alveolar ridge anteriorly and posteriorly. Occlusion is noted and the patient is asked if his teeth "fit together" as usual. The mandible is palpated for any tenderness or fracture.

Radiologic Examination

The radiologic examination is guided by the history and physical examination (Table 43-1). Most patients will need clear views of the cervical spine, including odontoid, posteroanterior, and lateral to include the first thoracic vertebrae. The standard facial bone series includes posteroanterior, lateral, and Waters views. Standard views of the mandible include oblique, lateral, and Towne views. A Panorex radiograph will provide an image of the entire mandible, including the temporomandibular joints. This film is necessary in most cases of suspected mandible fracture. Submental vertex films free the zygomatic arches of other bony silhouettes and are useful in diagnosing zygomatic arch fractures (Fig. 43-40).

Computed tomography is the mainstay in diagnosing and planning treatment for orbital fractures. Standard axial slices should be reformatted to provide coronal views of the orbit. These coronal images are extremely helpful in diagnosing orbital floor fractures.

Treatment

Mandibular Fractures

Specific mandibular fractures are treated according to the characteristics of the fracture and its anatomic location. The goal in treating any mandibular fracture is restoration of function occlusion with firm healing at the fracture site. Any open mandible fracture requires operative irrigation, debridement, and intravenous antibiotics.

A number of techniques may be employed in the management of mandibular fractures. *Intermaxillary fixation* is the ligation of the maxilla to the mandible with the teeth in occlusion, serving to immobilize the mandible. This may be accomplished by ligating metal arch bars to the upper and lower dental arch and then fixing these bars to each other by means of heavy elastics or wire. Alternatively, when the teeth are in occlusion, mandibular and maxillary teeth may be directly wired together. Intermaxillary fixation requires a patent nasal airway and must be used with caution in alcoholics and other patients prone to loss of consciousness or vomiting. Use of intermaxillary fixation is contraindicated in patients with a seizure disorder. *Open reduction and internal fixation* has become the treatment of choice for most unstable fractures of the mandible. Most fractures can be exposed intraorally through a buccal sulcus incision and rigidly fixed with plates and screws. This precludes the need for intermaxillary fixation and allows the patient to resume oral intake on the first postoperative day. It also reduces the possibility of limited mandibular excursion secondary to prolonged intermaxillary fixation. External fixation devices are employed only rarely, for large bony defects and recurrent infections.

Mandibular fractures may be classified as *favorable* or *unfavorable*, according to the manner in which the muscles of mastication serve to distract the fracture. Fractures of the body of the mandible that course obliquely and anteriorly when followed from inferior to superior are termed unfavorable. In these fractures the masseter muscle, by pulling on the posterior segment, will displace the fracture (Fig. 43-41). Fractures of the body that course posteriorly in the inferior to superior direction are termed favorable. In these fractures the masseter muscle, by pulling superiorly on the posterior segment, serves to appose the mandibular segments.

Fractures of the rami and mandibular condyles are frequently not greatly displaced by the muscles of mastication. These may be treated with intermaxillary fixation or a soft diet alone if not widely displaced and occlusion is adequate. Widely displaced fractures of the condyle may require open reduction and internal fixation. Fractures of the coronoid process usually do not interfere with occlusion unless the fracture fragments interfere with mandibular excursion. Coronoid process fractures usually may be treated with a soft diet. Fractures of the angle and body of the mandible are subjected to significant stress and motion by the muscles of mastication. These fractures usually require open reduction and internal fixation. Minimally displaced fractures may be treated by internal maxillary fixation. Nondisplaced fractures can be treated with a soft diet alone in selected patients. Fractures of the symphysis and parasymphyseal area also are usually displaced by the muscles of mastication and normally require open reduction and internal fixation. Fractures of the alveolar process may be seen alone or in combination with other fractures of the mandible. These are usually treated by ligating the teeth of the fractured segment to adjacent stable teeth by the application of arch bars.

Nasal Fractures

A nasal fracture is usually diagnosed on the basis of gross deformity and nasal airway obstruction. In patients with craniofacial trauma any deformity should be noted and the nasal airway should be examined with a speculum for patency and the presence of a septal hematoma. Septal hematomas require prompt incision, drainage, and nasal packing to prevent subsequent infection and septal perforation. With early treatment nasal fractures usually can be reduced with closed techniques under local or general anesthesia. Late treatment of nasal fractures usually requires osteotomies to mobilize and adequately reposition the nasal bones.

More severe nasal fractures often occur when the nasofrontal area strikes the steering wheel during a motor vehicle accident. These fractures involve not only the nasal bones but also the nasofrontal junction, the ethmoid sinuses bilaterally, and the medial orbits. These fractures are also known as nasoorbital-ethmoid fractures. Typically the patient with a nasoorbital-ethmoid fracture has a depressed and deformed nasal contour, and *telecanthus*, an increased distance between the medial canthi. The medial canthus inserts into the frontal process of the maxilla and the lacrimal bone in the medial orbit. In a nasoorbital-ethmoid fracture these bones are comminuted and displaced laterally, increasing the distance between the medial canthi (Fig. 43-42). Nasoorbital-ethmoid fractures are best treated by extensive exposure through a bicoronal incision. The multiple small fracture fragments are accurately reduced and wired or plated in place. If the medial canthi remain attached to a large fracture fragment, accurate reduction and rigid fixation of this fragment will reposition the medial canthi. A transnasal wire placed posterior and superior to the native insertion of the medial canthus can also be used to reduce the traumatic telecanthus. This wire is placed as a loop through each medial canthus and then tightened, repositioning the canthi in a more favorable and medial position.

Zygoma Fractures

Fractures of the zygoma commonly occur as isolated fractures of the zygomatic arch or as zygomaticomaxillary complex fractures (Fig. 43-43). Fractures of the zygomatic arch, when depressed, may impinge on the temporalis muscle, causing pain with mandibular excursion. The depression in the lateral face may also be a significant distortion in appearance. Noncomminuted zygomatic arch fractures sometimes may be elevated by sliding an instrument beneath the arch via an incision in the temporal scalp. Comminuted zygomatic arch fractures usually require plating via a bicoronal incision. Zygomaticomaxillary complex fractures result in a depressed malar eminence and infraorbital rim. The fracture lines are typically along the zygomatic frontal suture, zygomatic arch, and infraorbital rim and floor of the orbit, through the infraorbital foramen (Fig. 43-44). These patients typically present with an inferiorly and laterally displaced eye, enophthalmos, and paresthesia of the infraorbital nerve. Malocclusion also can result if the fracture extends along the anterior wall of the maxilla into the dental arch. These fractures require open reduction and internal fixation of the fracture fragments and, if any significant defect exists, reconstruction of the orbital floor.

Orbital Floor Fractures

Orbital floor fractures may occur as a result of a direct blow to the eye. An increase in the infraorbital pressure can result in herniation of intraorbital contents through a fracture in the orbital floor (Fig. 43-45). These “blowout” fractures may involve entrapment of intraorbital fat and occasionally extraocular muscles within the orbital floor defect. These patients typically present with enophthalmos, opacification of the maxillary sinus on radiography, and diplopia, particularly on upward gaze. The orbital floor defect, in effect, increases the orbital volume, allowing the eye to settle posteriorly and inferiorly, resulting in enophthalmos. Diplopia and limitation of upward gaze is the result of entrapment of orbital fat or extraocular muscles within the orbital floor defect. Trauma to the extraocular muscles and resultant edema also can cause diplopia.

The indications for operation on orbital floor fractures have always been controversial. Enophthalmos of 2 mm or more is generally recognized as an indication for operation. It is important to remember, however, that edema after trauma can mask significant enophthalmos. Patients with a normally positioned globe in the immediate posttrauma period may become significantly enophthalmic several weeks later as edema and inflammation subside. We attempt to individualize the decision to operate based on enophthalmos after fully discussing the risks and benefits with each patient. Central to this discussion is the principle that minor degrees of enophthalmos after trauma that are easily correctable can evolve to significant enophthalmos that is difficult to correct once bone and soft tissues are well healed. Entrapment and limitation of extraocular muscles is also a general indication for operation. In some of these patients the limitation of gaze will resolve spontaneously, but as a general rule a floor defect large enough to limit ocular movement will also eventually leave the patient enophthalmic.

Exploration of the orbital floor is performed through an eyelid incision or a transconjunctival incision in the subperiosteal plane. The limits of the floor defect are defined and all periorbital herniated through the fracture are reduced. Most craniofacial surgeons favor repair of floor defects with onlay split calvarial bone grafts. Available data suggest that for moderate and small defects of the orbital floor alloplastic materials also may be safely used. Orbital floor implants are available composed of titanium, hydroxyapatite, silicone, Teflon, and other materials. Any patient with an orbital floor fracture should have a forced duction test performed pre- and postoperatively. Tetracaine is instilled locally to anesthetize the conjunctiva, and then the conjunctiva and insertion of the inferior rectus muscle are grasped with a fine-toothed forceps. Attempting to rotate the globe superiorly will reveal any limitation caused by entrapment. The forced duction test is mandatory as a means of evaluating reduction of entrapped periorbital.

Maxillary Fractures

Maxillary fractures may be classified according to the LeFort classification system (Fig. 43-46). LeFort recorded the pattern of fracture in fresh cadaver skulls subjected to blunt trauma. He noted several characteristic patterns of maxillary fracture. The LeFort I fracture extends from the piriform aperture laterally to the pterygomaxillary fissure. This fracture separates the lower maxilla, hard palate, and maxillary teeth from the remainder of the skull. The LeFort II fracture extends from the pterygomaxillary fissure superiorly across the anterior maxilla to the nasofrontal junction through the inferior orbital rim. This fracture separates the lower maxilla and nose from the craniofacial skeleton. The craniofacial disjunction or LeFort III fracture separates the entire midface from the cranium by fracturing through the pterygomaxillary fissure and zygomatic frontal sutures, along the floor of the orbit, and through the nasofrontal junction. LeFort fractures are normally treated by placing the patient in intermaxillary fixation and then rigidly fixing the fractures with plates and screws. Typically, 1.5-mm plates are placed on the nasomaxillary and zygomaxillary buttresses in LeFort I fractures. Additional fixation is provided at the nasofrontal junction and zygomatic arches in LeFort III fractures. The intermaxillary fixation may then be released once rigid fixation is established.

Ear Reconstruction

Microtia refers to a congenitally absent or small ear. This condition is often seen with other abnormalities of the first and second branchial arches, including micrognathia, hemifacial microsomia, and lateral facial clefts. The development of middle ear structures in a microtic ear is variable. Hearing is normally decreased in an affected ear. For this reason otitis media in the contralateral ear must be treated aggressively to preserve hearing. Reconstruction of a microtic ear normally begins at age five or six, before the child enters school (Fig. 43-47). The reconstruction is staged beginning with the insertion of a cartilaginous framework sculpted from costochondral cartilage. This stage requires meticulous technique to create a natural-looking ear framework and to avoid hematoma or infection. Subsequent stages include rotation of the native ear lobule and elevation of the ear from the mastoid fascia. A final stage often includes excavation of the concha and reconstruction of the tragus using a composite skin/cartilage graft from the contralateral ear.

Prominent ears are a congenital abnormality, resulting from some combination of an excessively large concha and a poorly developed antihelical fold. These ears tend to project excessively from the lateral skull. Otoplasty for prominent ears is undertaken to recreate an antihelical fold and decrease ear prominence. The Stenstrom technique takes advantage of the natural tendency of cartilage to bend away from any disruption of the perichondrium. When the perichondrium is scored anteriorly along the natural course of the antihelical fold, the ear cartilage will tend to bend posteriorly in this area, recreating the antihelical fold. The Mustarde technique involves placing sutures along the posterior auricular cartilage to recreate the antihelical fold. This cartilage is then fully mobilized and tubed to recreate the antihelical fold. Skin is also sometimes removed from the posterior surface of the ear, and the ear also may be sutured to the mastoid fascia (Fig. 43-48).

Nasal Reconstruction

Nasal defects commonly occur with the excision of squamous and basal cell carcinomas of the nose (Fig. 43-49). Depending on the location and size of these defects, they may be handled by skin grafting or local flaps. Defects of the alar rim may be reconstructed with a composite graft of cartilage and skin from the ear. Total nasal reconstruction remains one of the greatest challenges in plastic surgery. This requires a stable bony framework, often in the form of an iliac bone graft secured to the nasofrontal junction. The lining of the nasal airway may be provided by local flaps in the region of the nasolabial fold or by skin grafts. The nasal contour may be restored with a midline forehead flap, scalp flap, or free-tissue transfer.

Lip Reconstruction

Most defects of the lip encountered by the reconstructive surgeon may be closed primarily. Defects of up to one-third of the lower lip and one-fourth of the upper lip can usually be closed directly without the need for more complex reconstructive techniques. When closing lip defects that cross the vermilion-cutaneous border, it is imperative to align perfectly the vermilion and white roll edge of the vermilion. Any notching of the vermilion tends to catch the eye of an observer and is usually cosmetically unacceptable. A number of local flaps are available for reconstructing larger defects. The Abbe flap, or cross-lip flap, is an extremely useful technique to transfer full-thickness lip (Figs. 43-50 and 43-51).

Eyelid Reconstruction

Full-thickness defects of 25 percent of the upper and lower eyelid often may be closed primarily. When closing lid margins the tarsal plate should be approximated with fine absorbable suture and the gray line of the lid margin accurately approximated. In any laceration medial to the punctum of the lacrimal apparatus a lacrimal injury must be ruled out. Larger defects of the upper lid can be closed by advancing adjacent and lateral eyelid and temporal skin. Lower eyelid skin also can be advanced or rotated into an upper eyelid defect. At a second operation the flap is completely divided from the lower lid to complete the reconstruction. Large defects of the lower lid also may be closed by advancing lateral eyelid and cheek skin. Conjunctival lining can be provided by a free mucosal graft taken from the nasal septum with or without cartilage.

Eyebrow lacerations usually can be closed primarily with precise realignment of hair follicles (Fig. [43-52](#)). Complete reconstruction of an eyebrow is usually done with thin strips of free scalp grafts or a thin island flap based on the temporal artery (Figs. [43-53](#) and [43-54](#)).

Eyelid Ptosis

Elevation of the upper eyelid is achieved through contraction of the levator palpebrae superioris muscle and Müller's muscle. The levator muscle is innervated by the oculomotor nerve (IIIrd cranial nerve) and Müller's muscle receives sympathetic innervation. If either of these muscles functions poorly or not at all, the upper eyelid will fail to elevate adequately, which is known as *ptosis*. Ptosis may be congenital or acquired. Congenital ptosis is usually idiopathic. Acquired ptosis may be a result of dysfunction of the oculomotor nerve or the sympathetic chain. Horner's syndrome includes upper eyelid ptosis, miosis, and decreased sweating on the affected side of the face. Horner's syndrome is seen with disruption of the sympathetic tract and may be a presenting sign of malignancy. Myasthenia gravis can present as eyelid ptosis. Most cases of acquired ptosis are the result of trauma. In these cases disruption or stretching of the levator aponeurosis is usually seen.

Treatment of eyelid ptosis is generally determined by the degree of levator function. Levator function is assessed by immobilizing the frontalis muscle with gentle pressure above the eyebrows. Upper lid excursion is then measured as the patient's gaze moves from inferior to superior. Levator function is classified as excellent (12 to 15 mm), good (8 to 12 mm), fair (5 to 7 mm), poor (less than 4 mm). Most cases of congenital ptosis have poor levator function, whereas cases of acquired ptosis generally have fair to good excursion of the upper lid. Minimal ptosis with good levator function can be treated with transconjunctival excision of part of the tarsal plate and Müller's muscle. This is known as the *Fasanella-Servat procedure*. More significant ptosis with fair to good levator function is usually treated by resection of the levator muscle through an upper eyelid incision. The shortened levator muscle is then approximated to the tarsal plate. Resection or plication of the levator muscle is usually accomplished through an upper eyelid skin incision. In cases of severe congenital ptosis with absent levator function, suspension of the upper eyelid to the frontalis muscle usually is required. Fascia lata or alloplastic material may be used for suspension of the lid.

Skull and Scalp Reconstruction

Partial defects of the scalp and calvaria are usually the result of trauma or surgical defects after tumor excision. Acute coverage of small or moderate- sized scalp defects can be accomplished with scalp flaps. Orticochea has described raising the entire scalp as three or four flaps that are then transposed into the defect. By scoring the galea aponeurotica on the undersurface of each flap, increased length can be gained. Large defects can also be closed with split-thickness skin grafts. Split-thickness skin grafts placed on the calvaria provide unstable coverage in the long term, however, and are prone to breakdown. Acute coverage of large, full- thickness defects involving both scalp and calvaria may require free-tissue transfer. The calvaria may be reconstructed with split calvarial or rib grafts that are in turn covered by a free flap. The omentum and latissimus dorsi or serratus anterior muscle are frequent choices for free flaps in reconstruction of the scalp and calvaria. These flaps are thin and broad and easily support a skin graft.

Tissue expansion is a useful technique for reconstructing partial defects of the scalp. By effectively increasing the amount of hair-bearing scalp, areas of burn alopecia or prior skin grafts placed on the scalp may be excised and closed with adjacent hair-bearing scalp.

Total scalp avulsion is usually the result of hair becoming entangled in machinery. Because the galea is a strong layer that resists separation, scalp avulsions frequently separate between the galea and the periosteum. The avulsed segment usually includes the eyebrow and upper ear. This devastating injury should be treated by microvascular replantation whenever possible. As the scalp is relatively resistant to hypoxia, long delays in treatment do not preclude replantation. Contraindications to scalp replantation include an unstable patient or severe damage to the avulsed scalp.

Facial Reanimation

The facial nerve is both a motor and a sensory nerve. The sensory component of the nerve includes both an afferent division, which is responsible for taste from the anterior two-thirds of the tongue via the chorda tympani nerve, and an efferent division with secretory fibers to the lacrimal gland and, also via the chorda tympani, to the sublingual and submandibular glands. The motor component innervates the muscles of facial expression and the stapedius muscle in the middle ear. Facial nerve paralysis is seen in a wide variety of disease stages. Congenital facial nerve paralysis is seen in Möbius' syndrome. Bell's palsy is thought to be the result of swelling within the bony canal. It is frequently idiopathic and occurs most commonly in young adults. Penetrating trauma can lacerate the nerve at any point after it exits from the stylomastoid foramen (Fig. [43-55](#)). Facial nerve paralysis can also be seen with fractures of the temporal bone.

A lacerated facial nerve should be explored and repaired as soon as possible. Using microsurgical techniques, the nerve ends are directly coapted whenever possible. Nerve grafts are used for segmental nerve loss when direct repair is not possible. Facial nerve resection as part of a cancer operation also should be reconstructed with nerve grafts. Planned postoperative radiation therapy does not contraindicate nerve grafting. Fractures of the temporal bone usually do not disrupt the facial nerve. When the facial nerve is affected by temporal bone fractures, partial facial nerve paralysis or late appearance of the facial nerve paralysis indicates that the nerve has not been completely disrupted. Nonoperative management should be pursued in these cases. With immediate and complete facial paralysis after a temporal bone fracture, decompression of the facial nerve within the bony canal should be undertaken.

Paralysis of the facial musculature results in functional problems that include incompetence of the oral and ocular sphincter mechanisms (Fig. [43- 56](#)). Paralysis of the orbicularis oculi muscle results in an inability to close the eyelid. This may lead to exposure keratitis. Paralysis of the orbicularis oris muscle can result in oral incompetence and abnormal speech. Facial symmetry and facial expression are also profoundly affected by facial paralysis. When the distal facial nerve–muscle unit is thought to be intact, nerve grafts from the contralateral intact facial nerve or from a partial transection of the ipsilateral hypoglossal nerve may be used to reanimate the affected side of the face (Fig. [43-57](#)). These techniques must be employed within 2 years of the time of injury, as motor endplates degenerate in the absence neural stimulation. Static techniques have long been used to suspend the paralyzed face (Fig. [43-58](#)). These techniques typically involve slings of fascia lata to the eyelids and mouth. More dynamic techniques include transfer of part of the temporalis or masseter musculature to the oral commissure and eyelid. As these muscles are innervated by the Vth cranial nerve, considerable practice is required by the patient to produce a coordinated smile (Fig. [43-59](#)).

Microneurovascular muscle transfers have yielded some of the most promising results in facial reanimation surgery. With these techniques a cross-facial nerve graft is performed and the advancing Tinel's sign followed. Approximately 1 year later a free vascularized muscle flap is performed and its motor nerve anastomosed to the cross-facial nerve graft. The gracilis muscle is the most common donor muscle. The pectoralis minor and latissimus dorsi muscles have also been used. Typically the gracilis muscle would be sutured to the oral commissure and the body of the zygoma in an effort to produce a more normal smile (Fig. [43-60](#)). The advantages of this technique are that facial movement is controlled by the contralateral facial nerve and that new vascularized muscle with well- preserved motor endplates is brought to the face. The main disadvantages of the technique are that two long operations are required, donor site scars results, and at least 2 years are usually required before facial movement returns.

Reconstruction After Tumor Extirpation

Surgical extirpation of head and neck tumors can create large complex defects that are potentially very disfiguring. Reconstruction of these defects requires not only restoration of form but also attention to the functions of speech, oral competence, and alimentation. Through the 1970s large head and neck defects were of necessity reconstructed by regional transposition flaps. The most common and reliable local flaps available to the head and neck included the deltopectoral flap and the pectoralis, latissimus, and trapezius myocutaneous flaps. The advent of microsurgical techniques has revolutionized head and neck reconstruction. Free flaps can provide thin, pliable oral lining, bone, and skin without bulky pedicles or disfiguring donor sites. For this reason free-tissue transfer has become the first choice in reconstruction of most complex head and neck defects.

Intraoral defects require reconstruction with thin, pliable tissue. For small defects split-thickness skin grafting is sometimes adequate. Care must be taken that, with secondary contraction of the skin graft, tongue mobility is maintained. Flaps of tongue tissue occasionally may be employed, but they too may adversely affect tongue mobility and limit speech or swallowing. Large defects of oral lining are usually reconstructed by free-tissue transfer. The free radial forearm flap based on the radial artery provides a flap with relatively thin skin and subcutaneous tissue and a long pedicle. Other available free flaps include scapular and parascapular flaps, based on cutaneous branches of the circumflex scapular artery, and the groin flap, based on the superficial circumflex iliac artery. Salvage techniques after free-flap failure or for patients thought not to be candidates for free-tissue transfer include deltopectoral and pectoralis major flaps.

Resection of more advanced oral cancers or primary tumors of bone may include defects of the mandible or maxilla. Small defects sometimes may be reconstructed with some combination of a soft-tissue flap and nonvascularized bone graft. Nonvascularized bone grafts do poorly in radiated beds and also poorly tolerate subsequent radiation therapy. When radiation therapy is part of the clinical picture, or for very large defects, free vascularized bone flaps are the procedure of choice. Donor sites include the fibula, based on the peroneal vessels, the lateral border of the scapula, based on circumflex scapular vessels, and the iliac crest, based on the deep circumflex iliac vessels. These free bone flaps may also be taken with an associated skin island to provide tissue for reconstruction of skin or mucosal defects.

The cervical esophagus must often be resected as part of the treatment for a stage T3 or T4 carcinoma of the hypopharynx and larynx. Tubed myocutaneous flaps have been used to reconstruct the cervical esophagus, but they are bulky and prone to fistula formation. The stomach can also be delivered into the neck and an esophagogastrostomy performed for reconstruction of cervical esophageal defects. A segment of jejunum transferred as a free vascularized flap, however, is our preferred method. This flap is reliable in a radiated wound and also tolerates radiation therapy. It is performed in a single stage and allows early resumption of oral feeding (Fig. [43-61](#)).

Hemangiomas and Vascular Malformations

Hemangiomas are common pediatric tumors that enlarge by cellular proliferation. Historically, hemangiomas have been given a number of names based on their clinical appearance. *Capillary hemangiomas* and *strawberry hemangiomas* have a similar biology. *Cavernous hemangiomas* occur deeper in subcutaneous tissue or muscle and hence appear darker or blue from the skin surface. Hemangiomas usually present within the first month of life and typically undergo a period of proliferation lasting 6 to 18 months. The lesion then begins to involute and this process may continue for several years. The majority of hemangiomas require no treatment. Even large lesions of the face frequently resolve with minimal or no cosmetic sequelae. Treatment is indicated for lesions that obstruct the visual axis, airway, or bilateral external auditory canals. Less commonly hemangiomas may be associated with hemorrhage or platelet consumption (Kasabach- Merritt syndrome), which requires treatment. Treatment modalities include systemic or intralesional

steroids. Recent reports have also indicated that interferon-alpha may be effective in rapidly expanding hemangiomas. Invasive treatment generally entails surgical excision for small tumors in areas where a surgical scar does not pose a cosmetic problem.

Vascular malformations are abnormal vascular structures formed during embryogenesis. These lesions are not neoplastic and do not exhibit the rapid cellular proliferation of hemangiomas. Vascular malformations may be predominantly capillary, venous, or lymphatic in composition. Frequently some combination of these components is present. Vascular malformations also may be present as abnormal connections between arteries and veins, known as *arteriovenous malformations*. These may be high-flow lesions that demonstrate bruits or thrills. Arteriovenous malformations can obstruct cervical viscera or massively bleed on dental manipulation if they arise in the head and the neck. Treatment is by intraarterial embolization under radiographic control followed by surgical removal when possible. These lesions can be extensive, and morbidity and mortality can be correspondingly high.

Intradermal capillary malformations are also known as *port-wine stains*. These present at birth and do not undergo resolution with time. Port-wine stains frequently present in the skin area innervated by the trigeminal nerve. Small lesions may be directly excised. Larger intradermal lesions usually benefit from treatment with laser.

Lymphatic malformations often present with some component of venous malformation. Lymphatic malformations most commonly occur in the neck and upper chest. *Cystic hygroma* is a form of lymphatic malformation commonly found in the neck, with large cystic cavities and abnormal venous connections. Treatment of lymphatic malformations is usually by direct surgical excision. Indications for treatment include recurrent infection and airway obstruction. Because these lesions frequently extensively invest vital structures of the neck, it is frequently possible only to debulk the lesion at the time of operation. Involution of cystic hygroma has been observed in some centers.

BREAST

Postmastectomy Breast Reconstruction

Breast cancer is a major public health concern in the United States. Currently most patients with breast cancer are treated by local excision of the cancer followed by radiation or mastectomy and axillary dissection. Even women who have undergone breast-conserving procedures frequently must make a decision about breast reconstruction.

The ideal candidate for breast reconstruction is a woman with a full understanding of the surgical procedures, possible complications, and realistic goals of breast reconstruction. Patients with a favorable prognosis are also usually good candidates for breast reconstruction, although an unfavorable prognosis does not necessarily rule out reconstruction. Possible benefits of breast reconstruction include an improved body image, preservation of feminine identity, elimination of external prostheses, and lessened psychological impact of mastectomy. Available data suggest that breast reconstruction does not delay detection of breast cancer recurrence.

Breast reconstruction often may be performed at the time of the mastectomy. Immediate reconstruction allows the patient to recover from anesthesia with a breast mound in place. Studies suggest that patients who have undergone immediate reconstruction experience less psychological disturbance. Immediate reconstruction requires that the patient undergo a longer initial period of anesthesia. Complications of the mastectomy or reconstruction could also conceivably delay subsequent chemotherapy or radiation therapy. For this reason many reconstructive surgeons prefer to offer immediate reconstruction to patients in whom the likelihood of advanced disease is small. For patients who choose delayed breast reconstruction, the reconstructive procedure usually is performed after chemotherapy or radiation therapy is complete. Advanced age and disease states are not absolute contraindications to reconstruction.

Breast reconstruction techniques may involve autologous or alloplastic materials. Alloplastic materials generally involve some combination of tissue expander and permanent breast prosthesis. Typically a patient choosing reconstruction with an expander/implant would have a tissue expander placed via the original mastectomy incision. Tissue expanders are generally placed in the submuscular position after raising the pectoralis major muscle, the serratus anterior muscle, and the anterior rectus sheath as a unit. Expansion is then carried out postoperatively, allowing for slight overexpansion relative to the desired final breast size. The tissue expander is exchanged for a permanent breast prosthesis during a second surgical procedure. In some patients there is adequate skin after a mastectomy for insertion of an implant large enough to match the contralateral breast without the use of a tissue expander, but this obtains only rarely.

Patients with significant deficiencies in skin and subcutaneous tissue may need to have a latissimus dorsi myocutaneous flap transposed into the surgical defect. The expander/prosthesis may then be positioned beneath the latissimus dorsi flap (Fig. 43-62). Complications following breast reconstruction with a tissue expander or implant include hematoma, infection, exposure of the implant, and capsular contraction. Contraction of the fibrous capsule that forms around a breast implant can cause significant deformity or pain. The true incidence of capsular contraction is unknown but probably ranges from 5 to 20 percent. Capsular contraction may be seen less frequently with textured implants.

Breast reconstruction with autologous tissue offers the patient a warm, natural-feeling breast without the use of alloplastic materials. The abdomen is an excellent source of autologous tissue for breast reconstruction. The entire lower abdominal ellipse of skin and subcutaneous tissue may be lifted as a myocutaneous flap based on perforators from the rectus abdominis muscles. These muscles receive most of their blood supply from the superior epigastric vessels, and care must be taken to preserve these vessels as the rectus abdominis muscle is dissected to the costal margin. The *transverse rectus abdominis musculocutaneous (TRAM)* flap is then passed under the skin of the upper abdomen and set into the surgical defect. The flap is shaped to match the contralateral breast. Because of the abundant tissue supplied to the TRAM flap a prosthesis is rarely necessary to complete the reconstruction (Fig. 43-63).

In experienced hands, the TRAM flap is extremely reliable and has a low overall complication rate. The incidence of partial flap loss and ventral hernia has been low. The TRAM flap also may be transferred as a free flap pedicled on the inferior epigastric vessels. These are typically anastomosed in the axilla to the thoracodorsal or axillary vessels. The free TRAM flap has the theoretical advantages of improved vascularity of the flap and less disturbance of abdominal wall integrity with a harvest of only a small block of rectus abdominis muscle. The inferior gluteal musculocutaneous unit is less commonly used as a donor site for free-flap breast reconstruction. The inferior gluteal donor site is acceptable but generally provides less tissue and is a more difficult dissection than the TRAM flap.

Nipple reconstruction is usually delayed for at least 6 weeks after breast reconstruction. The nipple itself may be reconstructed using local dermal flaps or grafted tissue from the contralateral nipple. The areola may be reconstructed with a combination of full-thickness skin grafts and tattooing.

Prophylactic mastectomy is considered for some patients with a high risk of breast cancer or extreme mastodynia. When a subcutaneous mastectomy is offered to these patients it must be well understood by both surgeon and patient that a small amount of breast tissue remains. The remaining breast tissue is usually in the region of the nipple areola. Immediate reconstruction after subcutaneous mastectomy can be carried out with breast implants. Because of the paucity of soft tissue between the skin surface and the breast implant, these reconstructions are frequently unsatisfactory. These patients frequently complain of a cold, unnatural-feeling breast and capsular contracture. In these instances, a latissimus dorsi flap may be transposed to supplement the reconstruction.

Gynecomastia

Gynecomastia is enlargement of the male breast secondary to an increase in ductal tissue. This is seen in a number of disease states, including testicular tumors and liver disease, and is a side effect of use of certain drugs, notably cimetidine and cannabis. Gynecomastia is normally observed in many adolescent males. If the gynecomastia is judged to be severe and interferes with psychosocial development, the excess breast tissue is usually carried out through a periareolar incision. Suction-assisted lipectomy is also helpful in treating mild cases.

Breast Reduction

Macromastia refers to an abnormal enlargement of the breast. This may be due to genetic factors, hormonal imbalances, or obesity. Complications of macromastia include back pain, brassiere strap furrowing, skin irritation in the inframammary folds, and breast pain, particularly in the rapidly enlarging breast.

Treatment of macromastia involves a physician-directed weight-loss program for obese patients. Reduction mammoplasty may be offered to the patient with macromastia after a thorough discussion of the risks and benefits of surgery. Most reduction techniques involve long scars that can be difficult to conceal completely. Lactation and nipple sensibility can be affected by the procedure. Other complications include hematoma, fat necrosis, nipple necrosis, and hypertrophic scar formation. There are a wide number of reduction techniques employed by reconstructive surgeons to reduce the size of the breast and reposition the nipple-areola complex. The nipple and areola are carried on a dermal pedicle that may be based superiorly, inferiorly, or medially (Fig. 43-64). The central pedicle technique maintains the nipple areola complex at the apex of the breast mound (Fig. 43-65). Breast tissue is excised with relatively more removed from the inferior and lateral poles of the breast. Excess breast skin is also excised and skin redraped around the new breast mound and nipple-areola complex (Fig. 43-66). The blood supply to the breast is from branches of the internal mammary arteries, intercostal perforators, lateral thoracic artery, and branches of the thoracoacromial trunk. Any technique of breast reduction selected must preserve the blood supply to the breast parenchyma and nipple. The central breast pedicle has the theoretical advantage of preserving all these arterial contributions.

The patient with extremely large breasts will necessarily have a long pedicle from the chest wall to the nipple-areola complex. This makes the blood supply to the nipple more tenuous and thus carries a higher risk of nipple necrosis. In these patients it is safest to move the nipple as a free nipple graft.

Breast Augmentation

Hypomastia refers to an abnormally small breast. This may be the result of genetic influences. Hypomastia is also seen in Poland's syndrome, along with the absence of the pectoralis major muscle. Breast augmentation can do much to enhance self-image in women with hypomastia. Most techniques of breast augmentation involve placing a prosthesis in a subglandular or submuscular position. Breast prostheses are saline-filled silicone shells or silicone shells filled with a combination of saline and silicone gel (Fig. 43-67). A breast implant may be inserted via an axillary, periareolar, or inframammary incision. Capsular contraction is the most frequent complication of breast augmentation and may be reduced when the implant is placed in a submuscular position deep to the pectoralis major muscle, or with textured implants. Rare complications of breast augmentation include hematoma and infection.

Implants filled with silicone gel are no longer available for augmentation procedures. These implants were the subject of much controversy in the early 1990s, prompting the Food and Drug Administration to remove them from the market. Concerns included migration of the silicone material and possible secondary connective tissue diseases. A number of large retrospective studies have failed to show an increased incidence of connective tissue disease in women who had undergone breast augmentation compared to the general population. Currently silicone gel implants are available for postmastectomy reconstruction only and within the confines of a research protocol. Only saline-filled silicone shells are available for breast augmentation.

Mastopexy

Breast ptosis refers to a sagging breast in which the nipple has descended below the inframammary fold. This condition is due to a relative excess of breast skin compared to breast parenchyma. Breast ptosis is usually the result of some combination of weight loss, aging, and postpartum atrophy. Mastopexy techniques are similar to breast reduction techniques in that the nipple is repositioned superiorly and excess breast skin excised. Breast parenchyma is preserved and repositioned.

TRUNK

Chest Wall Defects

Defects of the chest wall have the potential to expose vital organs and compromise respiration. These defects may be seen after tumor extirpation or trauma, or as a result of radiation necrosis. A number of local flaps are available to close chest-wall defects. The pectoralis major muscle may be mobilized to cover defects of the anterior chest wall and sternum. The latissimus dorsi muscle, because of its length and wide arch of rotation, is available to cover most defects of the back and anterolateral chest wall. The rectus abdominis muscle is frequently transposed into sternal defects. Because of its large surface area the greater omentum is available to resurface large areas of the anterior and lateral chest wall. This flap is particularly useful with extensive radiation injury.

Full-thickness defects of the chest wall may be reconstructed as a single stage with rib grafts or prosthetic materials used for skeletal reconstruction. Alloplastic materials used for skeletal reconstruction of the chest wall include Prolene mesh, Marlex Mesh Gore-Tex, and methyl methacrylate. Prosthetic materials, however, should not be used in radiated wounds.

Sternal Wounds

The infected sternal wound poses a particular challenge in chest-wall reconstruction. The overall infection rate after median sternotomy for coronary artery bypass procedures is generally reported to be less than 2 percent. Given the large number of these procedures performed annually, however, a significant number of infected sternal wounds require treatment. Historically, the infected sternal wound was treated by debridement and open packing. This yielded mortality rates as high as 50 to 70 percent and long hospitalizations. In the late 1960s and 1970s debridement and reclosure of the sternum over closed irrigation systems reduced the mortality rate of sternal wound infections to 20 percent. Jurkiewicz and associates reported in 1980 their experience with the treatment of infected median sternotomy wounds by debridement and closure with muscle flaps, usually in a single stage (Fig. 43-68). This technique has reduced mortality to less than 5 percent and decreased the length of hospitalization significantly. Muscle flaps available for closure of sternal wounds include the pectoralis major and the rectus abdominis muscles, and occasionally the latissimus dorsi muscles. The omentum also has been effectively used in the treatment of infected sternal wounds but requires a laparotomy.

Use of the internal mammary artery as a conduit for coronary artery bypass influences the treatment of sternal wound infections. Blood supply to the ipsilateral pectoralis and rectus muscles are affected by internal mammary artery harvest. In these instances, most surgeons select a muscle flap contralateral to the internal mammary artery graft. More aggressive sternal debridement is usually required in these cases as the ipsilateral sternum is significantly devascularized by internal mammary artery harvest.

Back

Reconstruction of the back is typified by meningomyelocele closure. This variant of spina bifida requires early closure to prevent infection and preserve remaining neural function. These defects are usually closed with some combination of gluteus and latissimus dorsi myocutaneous flap advancement. The latissimus dorsi muscle may be transferred on its thoracodorsal pedicle or reversed and transposed into the defect based on paraspinous perforators. Tissue-expansion techniques have been useful in this disorder when time is not of the essence.

Pressure Sores

Unrelieved pressure over bony prominences results in tissue necrosis if the sustained pressure exceeds the arterial capillary pressure of 32 mmHg. These so-called pressure sores frequently follow spinal cord injuries. They also occur in patients subjected to prolonged anesthesia. *Decubitus ulcers* refer to pressure sores that develop when the patient is recumbent for long periods. These wounds typically occur at the sacrum, the back of the head, and the greater trochanters. Pressure sores that occur over the ischial tuberosities in a sitting patient therefore are not decubitus ulcers. A typical pressure sore involves a cavernous wound overlying a larger area of subcutaneous fat or muscle necrosis. When chronic, these wounds form large bursas lined with granulation tissue.

Small superficial sores may be treated with local care and avoidance of pressure. More extensive sores usually require flap coverage. Flap selection is individualized to the location of the event of recurrence. Sacral sores may be closed with a transverse back flap or gluteus myocutaneous units. Wounds over the ischial tuberosities are often closed with a gluteus myocutaneous flap, posterior thigh flap, or biceps femoris musculocutaneous flap (Fig. 43-69). The tensor fascia lata musculocutaneous flap is often the first choice for closure of greater trochanteric wounds. After flap closure of pressure sores patients are normally kept in air-fluidized or low-air-loss beds for 3 weeks. Mobilization is then begun and the wound carefully observed for any signs of breakdown. Positive nitrogen balance established preoperatively must be maintained in the postoperative period. Behavior patterns must be changed to avoid recurrence. Despite these efforts recurrence rate is high, and recurrence in the young patient with traumatic paraplegia is the rule rather than the exception.

Abdominal Wall

Flaps available for anterior abdominal wall reconstruction are limited. Superiorly or inferiorly based rectus abdominis muscle flaps can reach most of the anterior abdomen, but these muscles are frequently involved in the abdominal wall defect. The tensor fascia lata and rectus femoris muscle flaps may be rotated into the defects of the lower anterior abdomen. Particularly useful for closing wounds of the irradiated perineum are the gracilis muscle and portions of gluteus muscle, usually with overlying skin as musculocutaneous unit.

LOWER EXTREMITY

Difficult wounds of the lower extremity are usually seen in the setting of other disease states, including venous or arterial insufficiency and diabetes. Patients with significant soft tissue defects over the tibia frequently require flap closure. These include patients with type III and IV open tibial fractures. Local gastrocnemius and soleus muscle flaps are preferred for proximal wounds when possible. Wounds of the distal one-third of the leg with exposed bone usually require free-tissue transfer. Posttraumatic chronic osteomyelitis of the tibia also usually requires debridement and muscle flap coverage. Cure rates higher than 90 percent are routinely reported for this difficult problem when muscle flaps are employed. Large defects of the tibia are frequently treated with free vascularized bone flaps. Segmental defects larger than 6 cm may be reconstructed by microsurgical transfer of fibula or iliac crest grafts. This technique typically yields success rates in the range of 80 to 90 percent.

Stasis Ulcers

Patients with venous stasis ulcers should be evaluated for treatment of their venous insufficiency. This may include ligation of incompetent perforators or venous bypass. The majority of patients are treated with compressive garments. Venous stasis ulcers that do not close with conservative care are usually excised and skin grafted. Local flaps are reserved for the most severe stasis ulcers. Free-tissue transfer has not been effective in the treatment of venous stasis ulcers.

Diabetic Ulcers

Recent advances have been made in the care of diabetic foot ulcers and limb salvage. A certain degree of fatalism has always existed concerning the care of diabetic foot ulcers, rooted in the misconception of "small-vessel disease." Historically, diabetic patients have been thought to develop occlusive disease at the level of the arteriole. The concept of small-vessel disease stems from postmortem studies demonstrating proliferative changes and hyaline deposition in vessels of diabetic patients. More recent prospective studies have failed to confirm these findings.

A large number of fasciocutaneous flaps have been described in the foot for closure of diabetic wounds. A number of small series have also reported success with free-tissue transfer for closure of large diabetic foot ulcers. The success of any reconstructive effort in the diabetic foot depends on a thorough vascular evaluation and bypass of flow-limiting lesions. Foot reconstruction may be performed at the time of revascularization. Free flaps have been performed at the time of distal revascularization, and theoretically they may improve patency of distal bypass grafts by providing additional outflow for the graft.

Lymphedema

Lymphedema, the abnormal collection of lymph in the interstitial space, is classified as *primary* when it occurs independently of other disease processes or surgery. Primary lymphedema is termed *lymphedema congenitum* when it is present at birth, *lymphedema praecox* when it occurs before the age of 35 years, and *lymphedema tardum* when it presents after the age of 35 years. *Secondary* lymphedema refers to lymphatic obstruction or ablation usually from malignancies, surgery, infection, or radiation.

Treatment of lymphedema is usually nonsurgical. The patient must be instructed to avoid even minor trauma to the extremity and to wear only well-fitting clothing and shoes. Elevation, when possible, and individually fitted compressive garments are the mainstay of treatment. In severe cases, surgical options are limited. Procedures involving skin flaps and skin grafts have been used since the early part of this century. Kondoleon in 1912 described the excision of deep muscle fascia beneath large skin flaps to facilitate lymphatic drainage directly into muscle tissue. This was largely unsuccessful and was modified by Sistrunk in 1918 to include staged subcutaneous excisions of edematous tissue beneath large skin flaps. Charles in 1912 described excision of skin and subcutaneous tissue with immediate regrafting of fascia and muscle of the limb with split- or full- thickness grafts. Full-thickness grafts generally provide a better result than split-thickness grafts, which tend toward unstable coverage, excessive scarring, and hyperkeratotic changes. The Charles technique, in the main, has been abandoned.

A number of procedures have been proposed to bypass or reconstruct the lymphatic system. These include dermal flaps buried in muscle, flaps of omentum placed subcutaneously in the lymphedematous extremity, demucosalized segments of ileum placed over lymph node basins, and microsurgical lymphaticovenous anastomoses. Clinical data and experience remain inadequate to demonstrate efficacy in these techniques. None of these procedures has gained widespread acceptance to date. Staged excision of the subcutaneous tissue and excess skin followed by rigorous adherence to use of pressure garments provide the best palliation.

AESTHETIC SURGERY

Facial Aging

Increased laxity of skin and subcutaneous tissues of the face is a normal part of aging. This typically presents as "bags" in the lower eyelids, the appearance of jowls along the jaw line, a deepened nasolabial fold, and excess skin and wrinkling in the face and neck. These changes can be treated effectively by surgical techniques to improve the appearance of the face.

The facelift, or *rhytidectomy*, is designed to excise excess skin and reposition the sagging soft tissue of the face (Fig. 43-70). The procedure may be performed under general or local anesthesia. The face and neck are infiltrated with 0.25% lidocaine with 1:400,000 epinephrine. The standard incision extends from a point in the temporal scalp approximately 6 cm superior to the root of the helix along the anterior margin of the ear around the lobe and along the posterior sulcus of the ear, terminating in hair-bearing scalp. In this manner all but a very fine scar anterior to the ear is hidden by hair-bearing scalp. The skin of the face and neck is then dissected from underlying muscles of facial expression.

A thin facial layer known as the superficial muscular aponeurotic system (SMAS) extends over the facial musculature and is contiguous with the platysma inferiorly and the frontalis muscle and temporal fascia superiorly. The SMAS may be dissected with the skin flap or as a separate layer and tightened by excising redundant SMAS or by plicating it to itself. Tightening the SMAS theoretically gives a more effective lift and longer-lasting results. Additional techniques include dissection in a deeper plane just superficial to the branches of the facial nerve as they exit the parotid gland, or in a subperiosteal plane. After excess SMAS and skin have been excised, the SMAS is sutured to fascia anterior to the ear and over the mastoid process. The skin is then closed. Drains may or may not be used. Ancillary techniques often used with the facelift procedure include suction-assisted lipectomy of the face, direct submental lipectomy through an incision beneath the chin, and tightening the platysma by suturing it laterally or in the midline should there be any diastasis.

Care must be taken to preserve vascularity to the skin flaps. Delayed wound healing may be seen in the posterior auricular incision, especially in patients who smoke. Hematomas sometimes occur and may range from small collections that resolve spontaneously to large hematomas requiring surgical drainage. The overall rate of significant hematoma formation is approximately 4 percent. Facial nerve injury is rare, occurring in less than 1 percent of patients.

As in other surgical specialties, minimally invasive surgery is becoming a standard part of the plastic surgeon's repertoire. Most of the commonly used endoscopic plastic surgery procedures are in the realm of cosmetic surgery. The endoscopic brow lift is one of the most common such procedures. Where previously a coronal incision was used to access the brow musculature and excise lax scalp, now three endoscopic ports are made in the scalp in the temporal and frontal regions (Fig. 43-71). Through these ports the brow musculature can be altered, usually by resection of the corrugator and procerus muscles in the nasofrontal region.

Long-term results in endoscopic aesthetic surgery remain controversial. Difficulty in standardizing pre- and postoperative results contributes to the difficulty. Despite this, patient preference for the limited incision of endoscopic approaches makes the endoscopic brow lift a mainstay of aesthetic surgery.

The fine wrinkles seen in the perioral region are generally not improved by rhytidectomy. These wrinkles may be addressed, however, by dermabrasion or chemical peel. *Dermabrasion* is a technique employing a rotating abrasive cylinder on a hand-held power tool. With this instrument skin surface irregularities such as shallow acne scars and fine wrinkles are removed. Dermabraded skin heals as a superficial partial-thickness burn. The wound is cared for in an open manner with antibiotic ointment. Pigmentation changes can occur, and it is important to avoid sunlight for several months after the procedure to diminish this risk. Phenol and trichloroacetic acid have been used to induce a superficial chemical burn to the face. This technique, known as a *chemical face peel*, has also been used to treat fine wrinkles around the mouth and eyes. When performed by experienced practitioners, the complication rates in chemical face peels are very low. Hypertrophic scarring has been reported. Pigmentation changes are the most common complication.

Similarly, the CO₂ laser also may resurface the face by removing a portion of the epidermis and superficial dermis. Some surgeons feel that this is a more precise technique than dermabrasion or chemical peel.

Eyelid

Facial aging can result in laxity in the soft tissues of the eyelid. The patient may complain of “bags” under the eyes and with excessive upper eyelid skin may actually have the visual field reduced on superior gaze. These changes are the result of excessive eyelid skin, redundant festoons of orbicularis oculi muscle, and prominence of intraorbital fat. The technique of blepharoplasty attempts to restore the eyelids to a more youthful appearance. This technique involves some combination of skin removal and fat excision. Incisions for lower eyelid blepharoplasty may be transconjunctival if no skin excision is planned, or in a subciliary position, immediately below the lower eyelid lashes. The incision for an upper eyelid blepharoplasty is typically in the supratarsal skin fold. The overall complication rate of blepharoplasty is low. Excessive lower lid skin excision in the lower lid may result in ectropion, especially in the elderly patient with poor lower-lid tone. Excessive fat excision results in a sunken or hollow look to the eyes. Blindness, however rare, has been reported in the literature from retrobulbar hemorrhage.

Nasal Deformity

Rhinoplasty is undertaken to alter the form of the nose or the function of the nasal airway. Patient complaints frequently include a dorsal nasal hump, broad nasal tip, asymmetry of the tip or dorsum, and nasal airway obstruction. Rhinoplasty may be performed under general or local anesthesia with open or closed technique. An open rhinoplasty involves an incision in the skin at the base of the columella that is then carried inside the nose along the alar rim. Through this incision the skin of the columella and nasal tip is dissected away from the alar cartilages. The surgeon then may reduce or modify alar and tip cartilage under direct vision. Changes in the bony pyramid of the nose are made by using a small osteotome to rasp the dorsum (to reduce any nasal hump) and the infracture of the nasal bones arising from the maxillae (to narrow the nose). The closed technique employs incisions inside the nose near or within the alar cartilages. The surgeon then modifies the nasal cartilages or bony pyramid of the nose as desired (Fig. 43-72). Because the closed technique does not allow these changes to be made under direct vision (Fig. 43-73), many surgeons prefer the open technique for difficult problems of the nasal tip. If nasal airway obstruction is present, the surgeon may elect to resect or modify the nasal septum at the time of rhinoplasty. Excision of hypertrophic inferior turbinates may also improve the nasal airway.

Abdomen, Thighs, and Buttocks

Pregnancy, weight loss, and aging can result in redundancy of abdominal skin. *Abdominoplasty* involves direct excision of a lower abdominal ellipse of skin to remove excesses of abdominal skin and fat (Fig. 43-74). The umbilicus is repositioned in the superior skin flap as it is advanced over the lower abdomen. If a diastasis recti or ventral hernia is present it may be repaired at the same time. This technique typically leaves an acceptable scar at the level of the pubis that may be concealed under clothing and swimwear.

The most common aesthetic complaint in the region of the abdomen, buttocks, and thighs is localized fat refractory to exercise and diet. For males this fat is typically found in the lower abdomen; females frequently complain of “saddlebags” of the lateral hips and thighs. Suction-assisted lipectomy has proved very helpful for these problems. Direct excision of the redundant skin of the buttocks and thighs is generally reserved for the patient who has undergone massive weight loss, as may be seen after gastric bypass for morbid obesity.

(Bibliography omitted in Palm version)

[Back to Contents](#)

CHAPTER 44 - Minimally Invasive Surgery

John G. Hunter

Minimally invasive surgery describes an area of surgery that crosses all traditional disciplines, from ophthalmology to podiatric surgery. It is not a discipline unto itself but more a philosophy of surgery, a way of thinking. Minimally invasive surgery is to perform major operations through small incisions, often using miniaturized high-tech imaging systems, to minimize the trauma of surgical exposure. Some believe that *minimal access surgery* more accurately describes the small incisions generally necessary to gain access to surgical sites in high-tech surgery, but John Wickham's term *minimally invasive surgery (MIS)* is widely used because it describes the paradox of postmodern high-tech surgery, small holes, big operations—the “minimalness” of the access and the invasiveness of the procedures, captured in three words.

HISTORICAL BACKGROUND

While the term *minimally invasive surgery* is relatively recent, the history of its component parts is nearly a hundred years old. The newest and most popular variety of minimally invasive surgery, laparoscopy, is in fact the oldest. Primitive laparoscopy, placing a cystoscope within an inflated abdomen, was first performed by Kelling in 1901. Illumination of the abdomen required hot elements at the tip of the scope and was dangerous. In the late 1950s Hopkins described the rod lens, a method of transmitting light through a solid quartz rod with no heat and little light loss. Around the same time, thin quartz fibers were discovered to be capable of trapping light internally and conducting it around corners, opening the field of fiber optics and allowing the rapid development of flexible endoscopes. In the 1970s the application of flexible endoscopy grew faster than that of rigid endoscopy except in a few fields such as gynecology and orthopaedics. By the mid-1970s rigid and flexible endoscopes made a rapid transition from diagnostic instruments to therapeutic instruments. The explosion of video surgery in the past 10 years was a result of the development of compact high-resolution charge- couple devices, which could be mounted on the internal end of flexible endoscopes or to the external end of a Hopkins telescope. Coupled with bright light sources, fiberoptic cables, and high-resolution video monitors, the video endoscope has changed our understanding of surgical anatomy and changed the shape of surgical practice.

While optical imaging produced the majority of minimally invasive surgical procedures, other (traditionally radiologic) imaging technologies allowed the development of innovative procedures in the 1970s. Fluoroscopic imaging allowed the adoption of percutaneous vascular procedures, the most revolutionary of which was balloon angioplasty. Balloon-based procedures spread into all fields of medicine, assisting in a minimally invasive manner to open up clogged cylinders. Stents were then developed that were used in many disciplines to keep the newly ballooned segment open. The culmination of fluoroscopic balloon and stent proficiency is exemplified by the *transvenous intrahepatic portosystemic shunt (TIPS)* (Fig. [44-1](#)).

Minimally invasive surgical procedures using ultrasound imaging have been limited to fairly crude exercises such as fragmenting kidney stones and freezing liver tumors because of the relatively low resolution of ultrasound devices. Newer, high-resolution ultrasound methods with high-frequency crystals may act as a guide while performing minimally invasive resections of individual layers of the intestinal wall.

Axial imaging, such as *computed tomography (CT)*, has allowed the development of an area of minimally invasive surgery that is not often recognized because it requires only a CT scanner and a long needle. CT- guided drainage of abdominal fluid collections and percutaneous biopsy of abnormal tissues are minimally invasive means of performing procedures that previously required a celiotomy.

The most powerful noninvasive method of imaging that will allow the development of the least invasive—and potentially noninvasive—surgery is *magnetic resonance imaging (MRI)*. *MRI* is an extremely valuable diagnostic tool, but it is only slowly coming to be of therapeutic value. One obstacle to the use of *MRI* for minimally invasive surgery is that image production and refreshment of the image as a procedure progresses are slow. Another is that all instrumentation must be nonmetallic when working with the powerful magnets of an *MRI* scanner. Moreover, *MRI* magnets are bulky and limit the surgeon's access to the patient. “Open magnets” have been developed that allow the surgeon to stand between two large *MRI* coils, obtaining access to the portion of the patient being scanned. The advantage of *MRI*, in addition to the superb images produced, is that there is no radiation exposure to patient or surgeon. Some neurosurgeons are accumulating experience using *MRI* to perform frameless stereotactic surgery.

PHYSIOLOGY

With the least invasive of the minimally invasive surgical procedures no significant physiologic alterations occur. Many minimally invasive procedures require no sedation or minimal sedation, and there are few alterations to the cardiovascular, endocrinologic, or immunologic systems. The least invasive of such procedures include stereotactic biopsy of breast lesions and flexible gastrointestinal endoscopy. Minimally invasive procedures that require general anesthesia have a greater physiologic impact because of the anesthetic agent, the incision (even if small), and the induced pneumoperitoneum.

Laparoscopy

The unique feature of endoscopic surgery in the peritoneal cavity is the need to lift the abdominal wall from the abdominal organs. Two methods have been devised for achieving this. The first, used by most surgeons, is the induction of a pneumoperitoneum. Throughout the early twentieth century intraperitoneal visualization was achieved by inflating the abdominal cavity with air, using a sphygmomanometer bulb. The problem with using air insufflation is that nitrogen is poorly soluble in blood and is slowly absorbed across the peritoneal surfaces. Air pneumoperitoneum was believed to be more painful than nitrous oxide pneumoperitoneum. Subsequently, carbon dioxide and nitrous oxide were used for inflating the abdomen. N₂O had the advantage of being physiologically inert and rapidly absorbed. It also provided better analgesia for laparoscopy performed under local anesthesia when compared with CO₂ or air. The disadvantage of N₂O when compared to CO₂ was that it did not suppress combustion. CO₂ suppresses combustion and is rapidly absorbed and therefore is the preferred gas for laparoscopy (Fig. [44-2](#)).

The physiologic effects of CO₂ pneumoperitoneum can be divided into two areas: (1) gas-specific effects and (2) pressure-specific effects. CO₂ is rapidly absorbed across the peritoneal membrane into the circulation. In the circulation, CO₂ creates a respiratory acidosis by the generation of carbonic acid. Body buffers, the largest reserve of which lies in bone, absorb CO₂ (up to 120 L) and minimize the development of hypercarbia or respiratory acidosis during brief endoscopic procedures. Once the body buffers are saturated, respiratory acidosis develops rapidly, and the respiratory system assumes the burden of keeping up with the absorption of CO₂ and its release from these buffers.

In patients with normal respiratory function this is not difficult; the anesthesiologist increases the ventilatory rate or vital capacity on the ventilator. If the respiratory rate required exceeds 20 breaths per minute there may be less efficient gas exchange and increasing hypercarbia. Conversely, if vital capacity is increased substantially there is a greater opportunity for barotrauma and greater respiratory-motion-induced disruption of the upper abdominal operative field. In some situations, it is advisable to evacuate the pneumoperitoneum or reduce the intraabdominal pressure to allow time for the anesthesiologist to adjust for hypercarbia. While mild respiratory acidosis probably is an insignificant problem, more severe respiratory acidosis leading to cardiac arrhythmias has been reported. Hypercarbia also causes tachycardia and increased systemic vascular resistance, which elevates blood pressure and increases myocardial oxygen demand.

The pressure effects of the pneumoperitoneum on cardiovascular physiology also have been studied. In the hypovolemic individual, excessive pressure on the inferior vena cava and a reverse Trendelenburg position with loss of lower-extremity muscle tone may cause decreased venous return and cardiac output. This is not seen in the normovolemic patient. The most common arrhythmia created by laparoscopy is bradycardia. A rapid stretch of the peritoneal membrane often causes a vagovagal response with bradycardia and, occasionally, hypotension. The appropriate management of this event is desufflation of the abdomen, administration of vagolytic agents (e.g., atropine), and adequate volume replacement.

With the increased intraabdominal pressure compressing the inferior vena cava, there is diminished venous return from the lower extremities. This has been well documented in the patient placed in the reverse Trendelenburg position for upper abdominal operations. Venous engorgement and decreased venous return promote venous thrombosis. Many series of advanced laparoscopic procedures in which *deep venous thrombosis (DVT)* prophylaxis was not used demonstrate the occurrence of pulmonary embolus. This usually is an avoidable complication with the use of sequential compression stockings, subcutaneous heparin, or low- molecular-weight heparin. In short-duration laparoscopic procedures, such as appendectomy, hernia repair, or cholecystectomy, the risk of *DVT* may not be sufficient to warrant extensive *DVT* prophylaxis.

The increased pressure of the pneumoperitoneum is transmitted directly across the paralyzed diaphragm to the thoracic cavity, creating increased central venous pressure and increased filling pressures of the right and left sides of the heart. If the intraabdominal pressures are kept under 20 mmHg, the cardiac output usually is well maintained. The direct effect of the pneumoperitoneum on increasing intrathoracic pressure increases peak inspiratory pressure, pressure across the chest wall, and also the likelihood of barotrauma. Despite these concerns, disruption of blebs and consequent pneumothoraces are rare after uncomplicated laparoscopic surgery.

Increased intraabdominal pressure decreases renal blood flow, glomerular filtration rate, and urine output. These effects may be mediated by direct pressure on the kidney and the renal vein. The secondary effect of decreased renal blood flow is to increase plasma renin release, thereby increasing sodium retention. Increased circulating *antidiuretic hormone (ADH)* levels also are found during the pneumoperitoneum, increasing free water reabsorption in the distal tubules. Although the effects of the pneumoperitoneum on renal blood flow are immediately reversible, the hormonally mediated changes, such as elevated *ADH* levels, decrease urine output for up to 1 h after the procedure has ended. Intraoperative oliguria is common during laparoscopy, but the urine output is not a reflection of intravascular volume status; intravenous fluid administration during an uncomplicated laparoscopic procedure should not be linked to urine output. Because fluid losses through the open abdomen are eliminated with laparoscopy, the need for supplemental fluid during a laparoscopic surgical procedure is rare.

The hemodynamic and metabolic consequences of pneumoperitoneum are well tolerated by healthy individuals for a prolonged period and by most individuals for at least a short period. Difficulties can occur when a patient with compromised cardiovascular function is subjected to a long laparoscopic procedure. It is during these procedures that alternative approaches should be considered or insufflation pressure reduced. Alternative gases that have been suggested for laparoscopy include the inert gases—helium, neon, and argon. These gases are appealing because they cause no metabolic effects, but they are poorly soluble in blood (unlike CO₂ and N₂O) and are prone to create gas emboli if the gas has direct access to the venous system. Gas emboli are rare but serious complications of laparoscopic surgery. They should be suspected if hypotension develops during insufflation. Diagnosis may be made by listening (with an esophageal stethoscope) for the characteristic “mill wheel” murmur. The treatment of gas embolism is to place the patient in a left lateral decubitus position with the head down to trap the gas in the apex of the right ventricle. A rapidly placed central venous catheter then can be used to aspirate the gas out of the right ventricle.

In some situations minimally invasive abdominal surgery should be performed without insufflation. This has led to the development of an abdominal lift device that can be placed through a 10- to 12-mm trocar at the umbilicus. These devices have the advantage of creating little physiologic derangement, but they are bulky and intrusive. The exposure and working room offered by lift devices also are inferior to those accomplished by pneumoperitoneum. Lifting the anterior abdominal wall causes a “pinching in” of the lateral flank

walls, displacing the bowel medially and anteriorly into the operative field. A pneumoperitoneum, by its well- distributed intraabdominal pressure, provides better exposure. Abdominal lift devices also cause more postoperative pain, but they do allow the performance of minimally invasive surgery with standard (nonlaparoscopic) surgical instruments.

Early it was predicted that the surgical stress response would be significantly lessened with laparoscopic surgery, but this is not always the case. Serum cortisol levels after laparoscopic operations are often higher than after the equivalent operation performed in through an open incision. In terms of endocrine balance, the greatest difference between open and laparoscopic surgery is the more rapid equilibration of most stress- mediated hormone levels after laparoscopic surgery. Immune suppression also is less after laparoscopy than after open surgery. There is a trend toward more rapid normalization of cytokine levels after a laparoscopic procedure than after the equivalent procedure performed by celiotomy.

Thoracoscopy

The physiology of thoracic minimally invasive surgery (thoracoscopy) is different from that of laparoscopy. Because of the bony confines of the thorax it is unnecessary to use positive pressure when working in the thorax. The disadvantages of positive pressure in the chest include decreased venous return, mediastinal shift, and the need to keep a firm seal at all trocar sites. Without positive pressure, it is necessary to place a double-lumen endotracheal tube so that the ipsilateral lung can be deflated when the operation starts. By collapsing the ipsilateral lung, working space within the thorax is obtained.

Extracavitary Minimally Invasive Surgery

Many new minimally invasive surgical procedures are creating working spaces in extrathoracic and extraperitoneal locations. Laparoscopic inguinal hernia repair usually is performed in the anterior extraperitoneal Retzius space. Laparoscopic nephrectomy often is performed with retroperitoneal laparoscopy. Lower extremity vascular procedures and plastic surgical endoscopic procedures require the development of working space in unconventional planes, often at the level of the fascia, sometimes below the fascia, and occasionally in nonanatomic regions. Some of these techniques use insufflation of gas, but many use balloon inflation to develop the space followed by low-pressure gas insufflation or lift devices to maintain the space (Fig. 44-3). These techniques produce fewer and less severe adverse physiologic consequences than does the pneumoperitoneum, but the insufflation of gas into extraperitoneal locations can spread widely, causing subcutaneous emphysema and metabolic acidosis.

Anesthesia

The most important factors in appropriate anesthesia management are related to CO₂ pneumoperitoneum. The laparoscopic surgeon can influence cardiovascular performance by releasing intraabdominal retraction and dropping the pneumoperitoneum. Insensible fluid losses are negligible, and therefore intravenous fluid administration should not exceed a maintenance rate. Minimally invasive surgical procedures usually are outpatient procedures, and short-acting anesthetic agents are preferable. Because the factors that require hospitalization after laparoscopic procedures include the management of nausea, pain, and urinary retention, the anesthesiologist should minimize the use of agents that provoke these conditions and maximize the use of medications that prevent such problems. Critical to the anesthesia management of these patients is the use of nonnarcotic analgesics (e.g., ketorolac) and the liberal use of antiemetic agents.

GENERAL PRINCIPLES OF ACCESS

The most natural ports of access for minimally invasive surgery are the anatomic portals of entry and exit. The nares, mouth, urethra, and anus are used to access the respiratory, gastrointestinal, and urinary systems. The advantage of using these points of access is that no incision is required. The disadvantages lie in the long distances between the orifice and the region of interest.

Access to the vascular system may be accomplished under local anesthesia by “cutting down” and exposing the desired vessel, usually in the groin. Increasingly, vascular access is obtained with percutaneous techniques using a small incision, a needle, and a guide wire, over which are passed a variety of different-sized access devices. This approach, known as the Seldinger technique, is most frequently used by general surgeons for placement of Hickman catheters but also is used to gain access to the arterial and venous system for performance of minimally invasive procedures. Guide-wire-assisted, Seldinger-type techniques also are helpful for gaining access to the gut for procedures such as percutaneous endoscopic gastrostomy, for gaining access to the biliary system through the liver, and for gaining access to the upper urinary tract.

In thoracoscopic surgery, the access technique is similar to that used for placement of a chest tube. In these procedures general anesthesia and split-lung ventilation are essential. A small incision is made over the top of a rib and, under direct vision, carried down through the pleura. The lung is collapsed, and a plastic trocar is inserted across the chest wall to allow access with a telescope. Once the lung is completely collapsed, subsequent access may be performed with direct puncture, viewing all entry sites through the videoendoscope. Because insufflation of the chest is unnecessary, simple plastic sheaths that keep the small incisions open are all that is required to allow repeated access to the thorax.

The requirements for laparoscopy are more involved, because the creation of a pneumoperitoneum requires that instruments of access (trocars) contain a valve to maintain abdominal inflation. Two methods are used for establishing abdominal access during laparoscopic procedures. The first, direct puncture laparoscopy, begins with the elevation of the relaxed abdominal wall with two towel clips or a well-placed hand. A small incision is made in the umbilicus, and a specialized spring-loaded (Veress) needle is placed in the abdominal cavity (Fig. 44-4). With the Veress needle, two distinct pops are felt as the surgeon passes the needle through the abdominal wall fascia and the peritoneum. The umbilicus usually is selected as the preferred point of access because in this location the abdominal wall is quite thin, even in obese patients. The abdomen is inflated with a pressure-limited insufflator. CO₂ gas usually is used with maximal pressures in the range of 14 to 15 mmHg. Laparoscopic surgery can be performed under local anesthesia, but general anesthesia is preferable. Under local anesthesia, N₂O is used as the insufflating agent and insufflation is stopped after 2 L of gas is insufflated or when a pressure of 10 mmHg is reached.

After peritoneal insufflation, direct access to the abdomen is obtained with a 5- or 10-mm trocar. The critical issues for safe direct-puncture laparoscopy include the use of a vented stylet for the trocar or a trocar with a safety shield. The trocar must be pointed away from the sacral promontory and the great vessels. For performance of laparoscopic cholecystectomy, the trocar is angled toward the right upper quadrant.

Occasionally the direct peritoneal access (Hasson) technique is advisable. With this technique, the surgeon makes a small incision just below the umbilicus and under direct vision locates the abdominal fascia. Two Kocher clamps are placed on the fascia, and with a curved Mayo scissors a small incision is made through the fascia and underlying peritoneum. A finger is placed into the abdomen to make sure that there is no adherent bowel. A sturdy suture is placed on each side of the fascia and secured to the wings of a specialized trocar, which is then passed directly into the abdominal cavity (Fig. 44-5). Rapid insufflation can make up for some of the time lost with the initial dissection. This technique is preferable for the abdomen of patients who have undergone previous operations, in which small bowel may be adherent to the undersurface of the abdominal wound. The close adherence of bowel to the peritoneum in the previously operated abdomen does not eliminate the possibility of intestinal injury but should make great vessel injury extremely unlikely. Because of the difficulties in visualizing the abdominal region immediately adjacent to the primary trocar, it is recommended that the telescope be passed through a secondary trocar in order to inspect the site of initial abdominal access. This examination usually is performed at the end of the operative procedure.

Secondary punctures are made with 5- and 10-mm trocars. For safe access to the abdominal cavity, it is critical to visualize all sites of trocar entry. At the completion of the operation, all trocars are removed under direct vision and the insertion sites are inspected for bleeding. If bleeding occurs, direct pressure with an instrument from another trocar site or balloon tamponade with a Foley catheter placed through the trocar site generally stops the bleeding within 3 to 5 min. When this is not successful, a full-thickness abdominal wall suture has been used successfully to tamponade trocar site bleeding.

It is generally agreed that 5-mm trocars need no site suturing. Ten- millimeter trocars placed off the midline and above the transverse mesocolon do not require repair. Conversely, if the fascia has been dilated to allow the passage of the gallbladder, all midline 10-mm trocar sites should be repaired at the fascial level with interrupted sutures. Specialized suture delivery systems similar to crochet needles have been developed for mass closure of the abdominal wall in obese patients, in whom it is difficult through a small skin incision to visualize the fascia. Failure to close lower abdominal trocar sites that are 10 mm in diameter or larger can lead an incarcerated hernia.

Access for Subcutaneous and Extraperitoneal Surgery

There are two methods for gaining access to nonanatomic spaces. For retroperitoneal locations, balloon dissection is effective. This access technique is appropriate for the extraperitoneal repair of inguinal hernias and for retroperitoneal surgery for adrenalectomy, nephrectomy, lumbar discectomy, or para-aortic lymph node dissection. The initial access to the extraperitoneal space is performed in a way similar to direct puncture laparoscopy except that the last layer (the peritoneum) is not traversed. Once the transversalis fascia has been punctured, a specialized trocar with a balloon on the end is introduced. The balloon is inflated in the extraperitoneal space to create a working chamber. The balloon then is deflated, and a Hasson trocar is placed. An insufflation pressure of 10 mmHg usually is adequate to keep the extraperitoneal space open for dissection. Higher gas pressures force CO₂ into the soft tissues and may contribute to hypercarbia. Extraperitoneal endosurgery provides less working space than laparoscopy but eliminates the possibility of intestinal injury, intestinal adhesion, herniation at the trocar sites, and ileus. These issues are important for laparoscopic hernia repair because extraperitoneal approaches protect the small bowel from sticking to the prosthetic mesh.

Subcutaneous surgery, the newest method of access in minimally invasive surgery, uses the creation of working room in nonanatomic spaces. This technique has been most widely used in cardiac, vascular, and plastic surgery. In cardiac surgery, subcutaneous access has been used for saphenous vein harvesting, and in vascular surgery for ligation of subfascial perforating veins (Linton procedure). With minimally invasive techniques the entire saphenous vein above the knee may be harvested through a single incision (Fig. 44-6). Once the saphenous vein is located, a long retractor that holds a 5-mm laparoscope allows the coaxial dissection of the vein and coagulation or clipping of each side branch. A small incision above the knee also can be used to ligate perforating veins in the lower leg.

Subcutaneous access also is used for plastic surgical procedures. Minimally invasive approaches are especially well suited to cosmetic surgery, in which attempts are made to hide the incision. It is easier to hide several 5-mm incisions than one long incision. The technique of blunt dissection along fascial planes combined with lighted retractors and endoscope holding retractors is most successful for extensive subcutaneous surgery. Some prefer gas insufflation of these soft tissue planes. The primary disadvantage of soft tissue insufflation is that subcutaneous emphysema can be created.

Imaging Systems

Two methods of video imaging are widely used. The first of these is flexible video endoscopy. With flexible video endoscopy, a *charge-couple device (CCD)* is placed on the internal end of a long flexible endoscope. In the second method, thin quartz fibers are packed together in a bundle, and the *CCD* camera is mounted on the external end of the endoscope. Most standard gastrointestinal endoscopes have the *CCD* chip at the distal end, but small, delicate choledochoscopes and nephroscopes are equipped with fiberoptic bundles. Distally mounted *CCD* chips were developed for laparoscopy but are unpopular.

Imaging for laparoscopy, thoracoscopy, and subcutaneous surgery uses a rigid metal telescope, usually 30 cm in length. This telescope contains a series of quartz optical rods with differing optical characteristics that provide a specific “character” to each telescope (Fig. 44-7). These metal telescopes vary in size from 2 to 10 mm in diameter. Since light transmission is dependent on the cross-sectional area of the quartz rod, when the diameter of a rod/lens system is doubled, the illumination is quadrupled. Little illumination is needed in highly reflective small spaces such as the knee, and a very small telescope will suffice. When working in the abdominal cavity, especially if blood is present, the full illumination of a 10-mm telescope usually is necessary.

Rigid telescopes may have a flat or angled end. The flat end provides a straight view (0 degrees), and the angled end provides an oblique view (30 or 45 degrees). Angled scopes allow greater flexibility in viewing a wider operative field through a single trocar site; rotating an angled telescope changes the field of view. The use of an angled telescope has distinct advantages for most videoendoscopic procedures, particularly in visualizing the common bile duct during laparoscopic cholecystectomy or visualizing the posterior esophagus or the tip of the spleen during laparoscopic fundoplication.

Light is delivered to the endoscope through a fiberoptic light cable. These light cables are highly inefficient, losing more than 90 percent of the light delivered from the light source. Extremely bright light sources (300 watts) are necessary to provide adequate illumination for video endosurgery.

Video cameras come in two basic designs. The one-chip camera has a black-and-white video chip that has an internal processor capable of converting gray scales to approximate colors. Perfect color representation is not possible with a one-chip camera, but perfect color representation is rarely necessary for endosurgery. The most accurate color representation is obtained using a three-chip video camera. A three-chip camera has *red, green, and blue (RGB)* input and is identical to the color cameras used for television production. *RGB* imaging provides the highest fidelity but is probably not necessary for everyday use. An additional feature of new video cameras is digital enhancement. Digital enhancement detects edges, areas where there are drastic color or light changes between two adjacent pixels. By enhancing this difference, the image appears sharper and surgical resolution is improved. Digital enhancement is available on one- and three-chip cameras. Priorities in a video system for minimally invasive surgery are illumination first, resolution second, and color third. Without the first two attributes, video surgery is unsafe.

There has been a recent interest in three-dimensional endoscopy. Three-dimensional laparoscopy provides the additional depth of field that is lost with 2-D endosurgery and allows greater facility for novice laparoscopists performing complex tasks of dexterity, including suturing and knot tying. The advantages of 3-D systems are less obvious to experienced laparoscopists. Additionally, because 3-D systems require the flickering of two similar images, which are resolved with special glasses, the image edges become fuzzy and resolution is lost. The optical accommodation necessary to rectify these slightly differing images negates any advantage offered by the additional depth of field.

Energy Sources for Endoscopic Surgery

Minimally invasive surgery uses conventional energy sources, but the requirement of bloodless surgery to maintain optimal visualization has spawned new ways of applying energy. The most common energy source is *radio frequency (RF)* electrosurgery using an alternating current with a frequency of 500,000 cycles/s (Hz). Tissue heating progresses through the well-known phases of coagulation (60°C), vaporization and desiccation (100°C), and carbonization (>200°C).

The two most common methods of delivering *RF* electrosurgery are with monopolar and bipolar electrodes. With monopolar electrosurgery a remote ground plate on the patient's leg or back receives the flow of electrons that originate at a point source, the surgical electrode. A fine-tipped electrode causes a high current density at the site of application and rapid tissue heating. Monopolar electrosurgery is inexpensive and easy to modulate to achieve different tissue effects. A short-duration, high-voltage discharge of current (coagulation current) provides extremely rapid tissue heating. Lower-voltage, higher-wattage current (cutting current) is better for tissue desiccation and vaporization. When the surgeon desires tissue division with the least amount of thermal injury and least coagulation necrosis, a cutting current is used.

With bipolar electrosurgery, the electrons flow between two adjacent electrodes. The tissue between the two electrodes is heated and desiccated. There is little opportunity for tissue cutting when bipolar current is used, but the ability to coapt the electrodes across a vessel provides the best method of small-vessel coagulation without thermal injury to adjacent tissues.

Another method of delivering radio frequency electrosurgery is argon beam coagulation. This is a type of monopolar electrosurgery in which a uniform field of electrons is distributed across a tissue surface by the use of a jet of argon gas. The argon gas jet distributes electrons more evenly across the surface than does spray electrofulguration. This technology has its greatest application for coagulation of diffusely bleeding surfaces, such as the cut edge of liver or spleen. It is of less use in laparoscopic procedures because the increased intraabdominal pressures created by the argon gas jet can increase the chances of a gas embolus.

Gas, liquid, and solid-state lasers have been available for medical application since the mid-1960s. The CO₂ laser (wavelength 10.6 mm) is most appropriately used for cutting and superficial ablation of tissues. It is most helpful in locations unreachable with a scalpel, such as excision of vocal cord granulomas. The CO₂ laser beam must be delivered with a series of mirrors and is therefore somewhat cumbersome to use. The next most popular laser is the neodymium:yttrium-aluminum-garnet (Nd:YAG) laser. Nd:YAG laser light is 1.064 mm (1064 nm) in wavelength. It is in the near-infrared portion of the spectrum and, like the CO₂ laser, is invisible to the naked eye. A unique feature of the Nd:YAG laser is that 1064-nm light is poorly absorbed by most tissue pigments and therefore travels deep into tissue. Deep tissue penetration provides deep tissue heating (Fig. 44-8). For this reason the Nd:YAG laser is capable of the greatest amount of tissue destruction with a single application. Such capabilities make it the ideal laser for destruction of large fungating tumors of the rectosigmoid or tracheobronchial tree. A disadvantage is that the deep tissue heating may cause perforation of a hollow viscus.

When it is desirable to coagulate flat lesions in the cecum, a different laser should be chosen. The frequency-doubled Nd:YAG laser, also known as the KTP laser (potassium thionyl phosphate crystal is used to double the Nd:YAG frequency), provides 532-nm light. This is in the green portion of the spectrum; at this wavelength, selective absorption by red pigments in tissue (such as hemangiomas and arteriovenous malformations) is optimal. The depth of tissue heating is intermediate between those of the CO₂ and the Nd:YAG lasers. Coagulation (without vaporization) of superficial vascular lesions can be obtained without intestinal perforation.

In flexible gastrointestinal endoscopy, the CO₂ and Nd:YAG lasers have largely been replaced by heater probes and endoluminal stents. The heater probe is a metal ball that is heated to a temperature (60 to 100°C) that allows coagulation of bleeding lesions without perforation.

A unique application of laser technology provides extremely rapid discharge (<10⁻⁶ seconds) of large amounts of energy (>10³ volts). These high-energy lasers, of which the pulsed dye laser has seen the most clinical use, allow the conversion of light energy to mechanical disruptive energy in the form of a shock wave. Such energy can be delivered through a quartz fiber and, with rapid repetitive discharges, can provide sufficient shock-wave energy to fragment kidney stones and gallstones. Shock waves also may be created with miniature electric spark-plug discharge systems, known as electrohydraulic lithotrippers. These devices also are inserted through thin probes for endoscopic application. Lasers have the advantage of pigment selectivity, but electrohydraulic lithotrippers are more popular because they are substantially less expensive and are more compact.

Methods of producing shock waves or heat with ultrasonic energy also is of interest. Extracorporeal shockwave lithotripsy creates focused shock waves that intensify as the focal point of the discharge is approached. When the focal point is within the body, large amounts of energy are capable of fragmenting stones. Slightly different configurations of this energy can be used to provide focused internal heating of tissues. Potential applications of this technology include the ability noninvasively to produce sufficient internal heating to destroy tissue without an incision.

A third means of using ultrasonic energy is to create rapidly oscillating instruments that are capable of heating tissue with friction; this technology represents a major step forward in energy technology. An example of its application is the *laparoscopic coagulation shears (LCS)* device (Harmonic Scalpel), which is capable of coagulating and dividing blood vessels by first occluding them and then providing sufficient heat to weld the blood vessel walls together and to divide the vessel. This nonelectric method of coagulating and dividing tissue with a minimal amount of collateral damage has facilitated the performance of numerous endosurgical procedures. It is especially useful in the control of bleeding from medium-sized vessels that are too big to manage with monopolar electrocautery and require bipolar desiccation followed by cutting.

Balloons and Stents

All minimally invasive procedures, from coronary angioplasty to palliation of pancreatic malignancy, invoke the use of endoluminal balloon dilators and prostheses. Endoluminal balloon dilators may be inserted through an endoscope, or they may be fluoroscopically guided. Balloon dilators all have low compliance—that is, the balloons do not stretch as the pressure within the balloon is increased. The high pressures achievable in the balloon create radial expansion of the narrowed vessel or orifice, usually disrupting the atherosclerotic plaque, the fibrotic stricture, or the muscular band (e.g., esophageal achalasia).

Once the dilation has been attained, it is frequently beneficial to hold the lumen open with a stent. Stenting is particularly valuable in treating malignant lesions and endovascular procedures (Fig. 44-9). Stenting usually is not applicable for the long-term management of benign gastrointestinal strictures, except in patients with limited life expectancy.

A variety of stents are available, but they may be divided into two basic categories, plastic stents and expandable metal stents. Plastic stents came first and are used widely as endoprotheses for temporary bypass of obstructions in the biliary or urinary systems. Metal stents generally are delivered over a balloon and expanded with the balloon to the desired size. These metal stents usually are made of titanium or nitinol. Although great progress has been made with expandable metal stents, two problems remain. The first is the propensity for tissue ingrowth through the interstices of the stent. Ingrowth may be an advantage in preventing stent migration, but such tissue ingrowth may occlude the lumen and cause obstruction anew. This is a particular problem when stents are used for palliation of gastrointestinal malignant growth and may be a problem for the long-term use of stents in vascular disease. Filling the interstices with silastic or other materials may prevent tumor ingrowth but also makes stent migration more likely. Stent designs have incorporated hooks and barbs in an attempt to minimize migration, addressing the second problem.

Hand Instruments

Hand instruments for minimally invasive surgery usually are duplications of conventional surgical instruments made longer, thinner, and smaller at the tip. Certain conventional instruments such as scissors are easy to reproduce with a diameter of 3 to 5 mm and a length of 20 to 45 cm, but other instruments, such as forceps and clamps, cannot provide remote access. Different configurations of graspers were developed to replace the various configurations of surgical forceps and surgical clamps. Standard hand instruments are 5 mm in diameter and 30 cm long, but smaller and shorter hand instruments are now available for pediatric surgery, for microlaparoscopic surgery, and for arthroscopic procedures. A unique laparoscopic hand instrument is the monopolar electrical hook. This device usually is configured with a suction and irrigation apparatus to eliminate smoke and blood from the operative field. The monopolar hook allows tenting of tissue over a bare metal wire with subsequent coagulation and division of the tissue.

Room Setup

Nearly all minimally invasive surgery, whether using fluoroscopic, ultrasound, or optical imaging, incorporates a video monitor as a guide. Occasionally two images are necessary to adequately guide the operation, as in such procedures as *endoscopic retrograde cholangiopancreatography (ERCP)*, laparoscopic common bile duct exploration, and laparoscopic

ultrasonography. When two images are necessary, the images should be mounted on two adjacent video monitors or projected on a single screen with a “picture in picture” effect. The video monitor(s) should be set across the operating table from the surgeon. The patient should be interposed between the surgeon and the video monitor; ideally, the operative field also lies between the surgeon and the monitor. In pelviscopic surgery it is best to place the video monitor at the patient's feet, and in laparoscopic cholecystectomy, the monitor is placed at the 10 o'clock position (relative to the patient) while the surgeon stands on the patient's left at the 4 o'clock position. The insufflating and patient monitoring equipment also is ideally placed across the table from the surgeon, so that the insufflating pressure and the patient's vital signs and end tidal CO₂ tension can be monitored.

Trocars for the surgeon's left and right hand should be placed at least 10 cm apart. For most operations it is possible to orient the telescope between these two trocars and slightly retracted from them. The ideal trocar orientation creates an equilateral triangle between the surgeon's right hand, left hand, and the telescope with 10 to 15 cm on each leg. If one imagines the target of the operation (e.g., the gallbladder or gastroesophageal junction) oriented at the apex of a second equilateral triangle built on the first, these four points of reference create a diamond (Fig. 44-10). The surgeon stands behind the telescope, which provides optimal ergonomic orientation but frequently requires that a camera operator (or robotic arm) reach between the surgeon's hands to guide the telescope.

The position of the operating table should permit the surgeon to work with both elbows in at the sides, with arms bent 90 degrees at the elbow. It usually is necessary to alter the operating table position with left or right tilt with the patient in the Trendelenburg or reverse Trendelenburg position, depending on the operative field.

Patient Positioning

Patients usually are placed in the supine position for laparoscopic surgery. When the operative field is the gastroesophageal junction or the left lobe of the liver, it is easiest to operate from between the legs. The legs may be elevated in Allen stirrups or abducted on leg boards to achieve this position. When pelvic procedures are performed, it usually is necessary to place the legs in Allen stirrups to gain access to the perineum. A lateral decubitus position with the table flexed provides the best access to the retroperitoneum when performing nephrectomy or adrenalectomy. For laparoscopic splenectomy, a 45-degree tilt of the patient provides excellent access to the lesser sac and the lateral peritoneal attachments to the spleen. For thoracoscopic surgery, the patient is placed in the lateral position.

When the patient's knees are to be bent for extended periods or the patient is going to be placed in a reverse Trendelenburg position for more than a few minutes, deep venous thrombosis prophylaxis should be used. Sequential compression of the lower extremities during prolonged (more than 90 min) laparoscopic procedures increases venous return and provides inhibition of thromboplastin activator inhibitor.

MINIMALLY INVASIVE GENERAL SURGICAL PROCEDURES (Table 44-1)

Cholecystectomy

Although flexible endoscopy introduced many general surgeons to video- assisted procedures (e.g., percutaneous endoscopic gastrostomy), it was laparoscopic cholecystectomy that provided the cornerstone around which minimally invasive surgery units were built.

The benefits of laparoscopic cholecystectomy are that it causes little pain and minimal scarring, requires only a short rehabilitation period, and usually can be performed as an outpatient procedure. Common bile duct injury occurs more frequently with laparoscopic cholecystectomy because it frequently is mistaken for the cystic duct. Five steps that help minimize the risk of bile duct injury are: (1) use of an angled (30-degree) scope, (2) maximal cephalic retraction of the gallbladder fundus, (3) lateral retraction of the infundibulum, (4) complete dissection of the gallbladder at its junction with the cystic duct, and (5) liberal use of fluoroscopic cholangiography (Fig. 44-11).

Appendectomy

The indications for laparoscopic appendectomy depend on the surgeon's confidence in the diagnosis, the age, sex, and weight of the patient, and the stage of appendicitis. Most agree that patients who have right iliac fossa tenderness of uncertain origin might benefit from diagnostic laparoscopy. If uncomplicated appendicitis is found, the appendix should be removed. If the appendix is normal and no other pathology is detected, the appendix should be removed. If a condition is detected that is likely to cause recurrent pain (e.g., terminal ileitis), the appendix should be removed. The appendix is left only if another surgical problem (e.g., acute cholecystitis) is detected. When unsuspected advanced inflammatory conditions of the appendix (gangrene, phlegmon, abscess) are discovered, conversion to a celiotomy, or rarely antibiotic therapy and interval appendectomy, is indicated. The perforated appendix can be treated laparoscopically if the inflammatory reaction is minimal and the base of the appendix is normal.

Additional indications for a selective approach to laparoscopic appendectomy include obese patients and active patients for whom rapid rehabilitation is paramount. The advantages of a nonselective approach to laparoscopic appendectomy were demonstrated in several prospective, randomized trials. In most of these trials, laparoscopic appendectomy resulted in less pain, a more rapid recovery, and an earlier hospital discharge. A meta-analysis of all prospective randomized trials confirmed these findings.

Laparoscopic appendectomy generally requires three ports, at least one of which is 10 mm in diameter. The surgeon needs an assistant to guide the video endoscope. Trocars are placed in the umbilicus (10 to 12 mm) in the suprapubic position (5 mm), and halfway between the umbilicus and the pubis (5 or 10 mm) for the telescope (Fig. 44-12 A). Working through the umbilical and suprapubic trocars, the appendix is mobilized and the mesoappendix divided with clips, bipolar electrosurgery, or staples. When the mesoappendix is involved with the inflammatory process, it is impossible to mobilize the appendix; it often is best to make a window next to the base of the appendix, divide the appendix with a stapler, then divide the mesoappendix immediately adjacent to the appendix with clips, stapler, or bipolar coagulation. The appendix is removed through a 10-mm trocar if it is small, or in a specimen bag if it is enlarged.

Inguinal Hernia Repair

Laparoscopic inguinal hernia repair is the most controversial of the minimally invasive procedures, for several reasons: it does not shorten hospitalization; the long-term efficacy of the procedures will take 20 years to evaluate; the technique is different from standard hernia repairs; and the hospital costs are higher than those of open hernia repair.

The early laparoscopic hernia repairs (plication of the internal ring, plug of the hernia sac, and prosthetic “onlay” over the hernia defect) have been abandoned. The two repairs performed today are the *transabdominal preperitoneal (TAPP)* repair and the *totally extraperitoneal (TEP)* repair. Although the *TEP* repair is slightly more difficult, many experienced laparoscopic surgeons prefer it for all but incarcerated hernias and for patients who have had previous lower abdominal operations.

The leading indication for the *TEP* repair is recurrent hernia. Anterior repair of recurrent hernias is associated with a 20 percent recurrence rate, division of the ilioinguinal nerve, and the possibility of testicular damage. These are the patients for whom the preperitoneal repairs were developed. The *TEP* repair nearly reproduces these repairs without the need for the large counterincision to place the mesh into the preperitoneal plane. It appears that the *TEP* repair will be able to achieve a recurrence rate similar to those of the open procedures, which is less than 5 percent.

Bilateral hernias are a second indication for *TEP*. The bilateral groin dissection and repair uses the same number and size of incisions as a unilateral repair. The “tension-free” nature of the repair eliminates concerns about a higher recurrence rate associated with simultaneous repair of bilateral hernias. The majority of the expense of laparoscopic herniorrhaphy is related to equipment charges and the time required to establish access. Repair of a contralateral side adds little time or equipment cost. Because of the ease associated with dissection and repair of the contralateral groin after one side has been repaired, a number of surgeons advocate routine dissection of the contralateral groin with simultaneous repair of asymptomatic hernias that are detected.

The preferred *TEP* approach requires three trocars placed in the lower midline (see Fig. 44-12 B). A small infraumbilical incision is opened down to the space between either rectus muscle and the posterior rectus sheath. A large balloon dissector is passed immediately to the pubic symphysis and inflated at this level. The balloon does 90 percent of the dissection of Hesselbach's triangle posteriorly. The balloon is removed and a Hasson cannula and 45-degree telescope are replaced. The only remaining dissection separates the spermatic cord from an indirect hernia sac. Small hernia sacs may be reduced entirely and left behind the mesh in the preperitoneal space. Rather than trying to reduce a large sac from the inguinal canal, sufficient sac is mobilized to ensure that no intraperitoneal structures remain within (sliding hernia). A 30-inch nonabsorbable suture is looped around the sac and tied extracorporeally. Extracorporeal knotting should be performed by sliding square knots down with a “knot pusher.” Most jamming knots (e.g., Roeder's knot) slip before breaking and have been abandoned. The sac is opened widely distal to the knot and left in the inguinal canal. A large piece of mesh, 10 × 15 cm at a minimum, is placed to cover Hesselbach's triangle, the internal inguinal ring, and the femoral canal. The mesh is fixed to Cooper's ligament, the pubic tubercle, the posterior rectus muscle, and the transversalis fascia lateral to the epigastric vessels. The fixation device should be palpable through the abdominal wall before any tacks are placed. This technique eliminates the possibility of entrapping the genitofemoral or lateral femoral cutaneous nerve within a staple.

Fundoplication

In the landmark study from the Veterans Administration Cooperative Study group it was proved that antireflux surgery better eliminated *gastroesophageal reflux (GER)* than did medical therapy in patients with severe reflux. These data arrived at the time that laparoscopic fundoplication was in its infancy. The results achieved with laparoscopic antireflux surgery in the 5 years since its inception have been virtually identical to the results of open fundoplication.

The indications for operation is gastroesophageal reflux that requires daily proton-pump inhibitors, symptomatic reflux associated with the complications of stricture, or Barrett's esophagus. Respiratory symptoms (e.g., asthma, cough, hoarseness) associated with well-documented gastroesophageal reflux constitute another indication. Evaluation of patients should be thorough and generally should include *esophagogastroduodenoscopy (EGD)*, barium swallow, 24-h ambulatory pH, and *esophageal motility study (EMS)*. A gastric emptying study is performed when gastric abnormalities are suspected (e.g., diabetes, vomiting, peptic ulcer disease). Abnormalities in gastric emptying may be addressed with prokinetic agents or concomitant pyloroplasty at the time of laparoscopic fundoplication.

In laparoscopic fundoplication the approach to the field of dissection involves the placement of the primary trocars in a diamond configuration (see Fig. 44-12 C). The surgeon stands at the base of the diamond. The table height is adjusted so that the surgeon's elbows are kept close to the sides. Intracorporeal knot tying reduces tissue trauma and is preferable to extracorporeal tying.

With laparoscopic fundoplication, it is most important to understand which cases are likely to present difficulty. This group includes patients with previous subdiaphragmatic surgery, obese patients, those with large diaphragmatic hernias (>5 cm), and those with significant esophageal shortening.

Assisted Colectomy

Laparoscopic removal of the colon is an area of rapid growth, but its application in patients with colon cancer has engendered heated debate. Metastatic tumor implantation at the site of the laparoscopic trocars occurs about twice as frequently after laparoscopic colectomy as implantation at the site of incision after open colectomy. Current indications for

laparoscopy- assisted colectomy are benign or premalignant disease, colostomy creation, and rectopexy for pelvic floor dysfunction. A creative use of laparoscopy is the mobilization of the splenic and hepatic flexures in procedures being performed primarily in the pelvis. After laparoscopic mobilization of the flexures, almost any operation—e.g., total abdominal colectomy, low anterior resection, or ileoanal pull-through—may be performed through a Pfannenstiel's incision.

The field of dissection usually traverses several quadrants of the abdomen. As the colon is progressively mobilized, it is frequently necessary to change the telescope to other trocar sites, and it may be necessary to add additional trocars. With increasing experience, trocar placement based on triangulation has simplified the approach (see Fig. 44-12 D, E). Additionally, it is necessary to make a counterincision to remove the colon specimen or to create a stoma. The duration of ileus is less than after conventional colectomy.

Splenectomy

The most common indication for laparoscopic splenectomy is immunologic disease (particularly idiopathic thrombocytopenic purpura) in patients with a normal or slightly enlarged spleen. Because of difficulties in getting the spleen into a collection bag, it is wise to restrict laparoscopic splenectomy to spleens that weigh less than 500 g.

The techniques for laparoscopic splenectomy include the supine technique, the left lateral decubitus (“hanging spleen”) technique, and the “leaning spleen” position, which is a compromise between the other two. The patient is positioned on the table at a 45-degree tilt, using a bean bag. Four trocars are placed while the table is rolled so that the patient is nearly supine (see Fig. 44-12 F). It may be necessary to take down the splenic flexure of the colon before placing the last trocar. The splenic artery is identified in the lesser sac by dividing short gastric vessels and rolling the gastric fundus to the right. The splenic artery is ligated (Fig. 44-13). At this point the table is tilted so that the patient moves into the right lateral decubitus position and the spleen allowed to fall anteriorly. The splenophrenic attachments (peritoneal reflection) are divided to allow access to the splenic hilum. In the hilum the splenic vessels are controlled with a linear cutting stapler, or they are individually ligated. Smaller vessels are clipped. The most difficult part of this operation may lie in manipulating the spleen into the retrieval bag. This step is made easier if the bag is opened in the left upper quadrant with the open mouth facing the laparoscope. A strong nylon bag is recommended for retrieving the spleen, as weaker bags may rupture. The neck of the bag is pulled out of a trocar site, and the spleen is morcellated and removed with ring forceps or with mechanized morcellation devices.

Adrenalectomy

Hyperfunctioning adenomas (including pheochromocytoma) are effectively dealt with laparoscopically. When adrenal carcinoma is suspected or the mass is greater than 5 cm, a laparoscopic approach is less appropriate. Similarly, when a search for paragangliomas is indicated or when it is necessary to perform simultaneous bilateral adrenalectomy, a transabdominal or posterior extraperitoneal approach might be warranted. The “no-touch” technique afforded by laparoscopic surgery is ideal for addressing a unilateral pheochromocytoma.

The preferred technique for laparoscopic adrenalectomy is to place the patient in the lateral decubitus position. Four trocars are placed beneath the costal margin, in the epigastrium, the midclavicular line, the anterior axillary line, and the posterior axillary line (see Fig. 44-12 G, H). On the left side, the surgeon begins the dissection by rolling the spleen anteriorly by dividing the splenophrenic ligament. Occasionally this is all that is needed to find the adrenal gland. In an obese patient endoscopic ultrasonography may be helpful in locating the adrenal gland. Sticking close to the adrenal gland but being careful not to grasp the gland, the surgeon teases away the retroperitoneal fat with a monopolar electrocautery hook, blunt dissection, and clips. The adrenal vein is located (inferomedially), and two clips are placed on the renal vein side of the adrenal vein. The adrenal vein is placed immediately into a bag and extracted through one of the trocar sites. The dissection on the right side is nearly identical to that on the left side except that a liver retractor is necessary to hold the right lobe of the liver anteriorly. Dissection of the gland away from the inferior vena cava starts inferiorly, allowing identification of the short right adrenal vein, which is divided between clips.

Common Bile Duct Exploration (CBDE)

Data from several centers have demonstrated that laparoscopic cholecystectomy with laparoscopic retrieval of bile duct stones is substantially more cost-effective than laparoscopic cholecystectomy followed by endoscopic sphincterotomy and stone retrieval.

Laparoscopic bile duct exploration starts with fluoroscopic cholangiography. A cholangiogram should be performed with a ureteral catheter or a modified ERCP catheter that allows the passage of a guide wire. If stones are detected, a hydrophilic guide wire is passed through the cystic duct into the duodenum. If the common bile duct is small, and the stones are small (<3mm), the cholangiogram catheter can easily be advanced over the guide wire to a position just above the stones. Intravenous glucagon is given, and a rapid infusion of saline will push small stones into the duodenum. When stones larger than 3 mm are detected or the common bile duct is dilated, it is best to remove the stones through the cystic duct. A helical stone basket frequently is all that is needed to remove small stones (5 mm or less). Larger common bile duct stones usually require an endoscopic approach. The cystic duct is dilated with an 8-mm balloon that has been passed over a guide wire. The balloon is left in place for 5 min while the choledochoscope is set up and an extra 5-mm trocar is placed (see Fig. 44-12 I). A thin choledochoscope (8 to 10 French) is passed with a rubber shod grasper over the guide wire and into the distal common bile duct. A 2- to 2.5-mm flat wire basket is passed through the operating channel of the choledochoscope and the stone is entrapped. Stones larger than 8 mm in diameter are best dealt with by performing a choledochotomy, stone extraction, and placement of a T tube (choledocholithotomy) (Fig. 44-14). Some use choledocholithotomy as their primary laparoscopic approach to stones of the common bile duct. Most bile duct stones (80 to 90 percent) can be dealt with using these techniques. Laparoscopic CBDE is as effective as endoscopic sphincterotomy and may reduce the risk of postoperative pancreatitis and the long-term and largely unknown sequelae of sphincter destruction.

SPECIAL CONSIDERATIONS

Pediatric Considerations

The advantages of minimally invasive surgery in children may be more significant than in the adult population. Minimally invasive surgery in the adolescent is little different from that in the adult, and standard instrumentation and trocar positions usually can be used. Laparoscopy in the infant and young child requires specialized instrumentation. The instruments are shorter (15 to 20 cm), and many are 3 mm in diameter rather than 5 mm. Because the abdomen of the child is much smaller than that of the adult, a 5-mm telescope provides sufficient illumination for most operations. The development of 5-mm clippers and bipolar devices has obviated the need for 10-mm trocars in pediatric laparoscopy. Because the abdominal wall is much thinner in infants, a pneumoperitoneum pressure of 8 mmHg can provide adequate exposure. Deep venous thrombosis is rare in children, and prophylaxis against thrombosis is probably unnecessary.

Pregnancy

Concerns about the safety of laparoscopic cholecystectomy or appendectomy in the pregnant patient have been eliminated. The pH of the fetus follows the pH of the mother linearly, and therefore fetal acidosis may be prevented by avoiding a respiratory acidosis in the mother. A second concern was that of increased intraabdominal pressure, but it has been proved that midpregnancy uterine contractions provide a much greater pressure in utero than a pneumoperitoneum. Experiences of well over 100 cases of laparoscopic cholecystectomy in pregnancy have been reported with uniformly good results. Operation should be performed during the second trimester if possible. Protection of the fetus against intraoperative x-rays is imperative. Some believe it advisable to track fetal pulse rate with a transvaginal ultrasound probe. Access to the abdomen in the pregnant patient should take into consideration the height of the uterine fundus, which reaches the umbilicus at 20 weeks. In order not to damage the uterus or its blood supply, most surgeons feel that the open (Hasson) approach should be used in favor of direct puncture laparoscopy.

Cancer

Minimally invasive techniques have been used for many decades to provide palliation for the patient with an obstruction cancer. Laser treatment, intracavitary radiation, stenting, and dilation are outpatient techniques that can be used to reestablish the continuity of an obstructed esophagus, bile duct, ureter, or airway. Minimally invasive techniques also have been used in the staging of cancer. Mediastinoscopy is still used occasionally before thoracotomy to assess the status of the mediastinal lymph nodes. Laparoscopy also is used to assess the liver in patients being evaluated for pancreatic, gastric, or hepatic resection. New technology and greater surgical skills allow accurate minimally invasive staging of cancer. Occasionally it is appropriate to perform palliative measures (e.g., laparoscopic gastrojejunostomy to bypass a pancreatic cancer) at the time of diagnostic laparoscopy if diagnostic findings preclude attempts at curative resection.

The most controversial role of minimally invasive techniques is that of providing potentially curative surgery to the patient with cancer. It is possible to perform laparoscopy-assisted colectomy, gastrectomy, pancreatectomy, and hepatectomy in patients with intraabdominal malignant disease, as well as thoracoscopic esophagectomy and pneumonectomy in patients with intrathoracic malignant disease. There are not yet enough data to indicate whether minimally invasive techniques provide survival rates or disease-free intervals comparable to those of conventional techniques. It has been proved that in laparoscopy-assisted colectomy and gastrectomy equivalent numbers of lymph nodes to the open procedure can be removed without any compromise of resection margins. A second concern centers around excessive tumor manipulation and the possibility that cancer cells would be shed during the dissection. Alarming reports of trocar site implantation with viable cancer have appeared in the literature.

Considerations in the Elderly and Infirm

Laparoscopic cholecystectomy has made possible the removal of a symptomatic gallbladder in many patients previously thought to be too elderly or too ill to undergo a laparotomy. Operations on these patients require closely monitored anesthesia. The intraoperative management of these patients may be more difficult with laparoscopic access than with open access. The advantage of minimally invasive surgery lies in what happens after the operation. Much of the morbidity of surgery in the elderly is a result of impaired mobilization. In addition, pulmonary complications, urinary tract sepsis, deep venous thrombosis, pulmonary embolism, congestive heart failure, and myocardial infarction often are the result of improper fluid management and decreased mobility. By allowing rapid and early mobilization, laparoscopic surgery has made possible the safe performance of procedures in the elderly and infirm.

Economics of Minimally Invasive Surgery

Minimally invasive surgical procedures reduce the costs of surgery most when length of hospital stay can be shortened. Shorter hospital stays can be demonstrated in laparoscopic cholecystectomy, fundoplication, splenectomy, and adrenalectomy, for example. Procedures such as inguinal herniorrhaphy that are already performed as outpatient procedures are less likely to provide cost advantage. Procedures that still require a 4- to 7-day hospitalization, such as laparoscopy-assisted colectomy, are even less likely to deliver a lower bottom line than their open-surgery counterparts. Nonetheless, with responsible use of disposable instrumentation and a commitment to the most effective use of the inpatient setting, most laparoscopic procedures can be made less expensive than their conventional equivalents.

(Bibliography omitted in Palm version)